DEMONSTRATIVES, DEICTIC POINTING AND THE CONCEPTUALIZATION OF SPACE

EDITED BY: Holger Diessel, Kenny Coventry, Harmen Gudde and Olga Capirci
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-782-6
DOI 10.3389/978-2-88966-782-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

DEMONSTRATIVES, DEICTIC POINTING AND THE CONCEPTUALIZATION OF SPACE

Topic Editors:
Holger Diessel, Friedrich Schiller University Jena, Germany
Kenny Coventry, University of East Anglia, United Kingdom
Harmen Gudde, University of East Anglia, United Kingdom
Olga Capirci, Italian National Research Council, Italy

Citation: Diessel, H., Coventry, K., Gudde, H., Capirci, O., eds. (2021). Demonstratives, Deictic Pointing and the Conceptualization of Space. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-782-6

Table of Contents

frontiers in Psychology

EDITORIAL
published: 30 March 2021
doi: 10.3389/fpsyg.2021.667667

Editorial: Demonstratives, Deictic Pointing and the Conceptualization of Space

Holger Diessel[1]*, *Kenny R. Coventry*[2], *Harmen B. Gudde*[2] and *Olga Capirci*[3]

[1] Department of English, Friedrich Schiller University Jena, Jena, Germany, [2] School of Psychology, University of East Anglia, Norwich, United Kingdom, [3] Institute of Cognitive Sciences and Technologies, Italian National Research Council, Rome, Italy

Keywords: deixis, demonstrative, pointing, gaze, joint attention

Editorial on the Research Topic

Demonstratives, Deictic Pointing and the Conceptualization of Space

Demonstratives have been studied intensively across several disciplines. There is increasing evidence that demonstratives constitute a unique class of expressions, fundamentally distinct from all other types of linguistic items. In contrast to most other function words, demonstratives seem to exist in all languages and are not derived from content words (Diessel, 2013). Moreover, demonstratives are closely related to gestural communication, notably to pointing (Stukenbrock, 2015), and are among the first and most frequent words in early child language (Clark, 1978). Demonstratives also play an important role in the encoding of reference in space (Coventry et al., 2008) and the organization of linguistic elements in discourse (Himmelmann, 1997). Finally, demonstratives are crucially involved in the diachronic evolution of grammar, as many grammatical markers are historically based on demonstratives (Diessel, 2006). Taken together, these features characterize demonstratives as a very special class of linguistic items that are foundational to communication, spatial orientation, discourse processing, language acquisition, and the emergence of grammar.

This Research Topic includes 14 articles that provide an overview of current research on demonstratives and cast a fresh light on some of the above-mentioned topics. The articles are based on empirical and theoretical research from several disciplines and are concerned with various languages, including several signed languages and a tactile language. Most papers present original research using a variety of data and methods including data from comprehension and production experiments, electronic corpora, video recordings, mobile eye tracking glasses, parental reports of child language, field reports, and a typological database. In what follows, we briefly describe the 14 articles in this collection.

The first paper, by Diessel and Coventry, provides an interdisciplinary review of current research on demonstratives in linguistics and psychology. At center stage is the debate about the nature of demonstrative reference. Older studies analyzed demonstratives as a particular class of spatial terms, but in the recent literature it is often claimed that demonstratives are primarily used for social purposes. Diessel and Coventry argue that demonstratives have both spatial and social functions.

Talmy outlines a new theory of deixis and anaphora, which are traditionally often treated as separate phenomena. Challenging this view, Talmy argues that linguistic reference involves a single cognitive system, i.e., the targeting system, that includes both speech-external and speech-internal uses of demonstratives. The bulk of the paper explains how linguistic triggers for deictic and anaphoric reference are interpreted in light of different types of cues providing information about the referent and its location.

OPEN ACCESS

Edited and reviewed by:
Manuel Carreiras,
Basque Center on Cognition, Brain and Language, Spain

***Correspondence:**
Holger Diessel
holger.diessel@uni-jena.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *14 February 2021*
Accepted: *01 March 2021*
Published: *30 March 2021*

Citation:
Diessel H, Coventry KR, Gudde HB and Capirci O (2021) Editorial: Demonstratives, Deictic Pointing and the Conceptualization of Space. Front. Psychol. 12:667667. doi: 10.3389/fpsyg.2021.667667

Demonstrative pairs (e.g., *this/that)* are commonly used to contrast objects or spaces. In the Demonstrative Choice Task (DCT), developed by Rocca and Wallentin, demonstratives are paired with nouns in the absence of such contextual information. The DCT shows consistent effects of semantic features (i.e., manipulability, motion, time) on demonstrative choice, suggesting that demonstratives mark contrasts along a range of semantic as well as spatial dimensions.

The combination of demonstratives and pointing has been noted in numerous studies, but there is little research on demonstratives and gaze. The study by Stukenbrock uses a mobile eye tracking device to investigate the occurrence of gaze following in deictic communication. The study breaks new methodological ground by using this technology for the first time in order to examine multimodal uses of deictic reference in naturally occurring interactions.

In the acquisition literature, it is commonly assumed that demonstratives are among the earliest and most frequent words children use. Analyzing data from child language corpora and parental reports of English- and Spanish-speaking children, González-Peña et al. challenge this view. While their corpus data are not entirely conclusive, the data from parental reports suggest that demonstratives are not usually among the first 50 child words.

Using the Memory Game (Coventry et al., 2008; Gudde et al., 2018), Vulchanova et al. investigate the potential influence of a second language on the deictic system of speakers' native language. Comparing a group of monolingual Spanish speakers with a group of bilingual Spanish speakers learning Norwegian, they found a significant difference between groups, suggesting that the two-term system of Norwegian changed the way the bilinguals used the three-term system of their native language.

The paper by Forker is a cross-linguistic study on demonstratives encoding elevation (up vs. down). Contrary to what one might expect, the study shows that the occurrence of elevational demonstratives is not predictable from the topographic environment. Moreover, the paper shows that the encoding of elevation is often restricted to distal demonstratives and that elevational demonstratives can be mapped onto non-spatial domains, e.g., time.

Khachaturyan analyzes demonstratives in Mano, a Mande language of Papua New Guinea. Examining field data, the paper challenges the traditional distinction between endophoric and exophoric uses of demonstratives and argues that speakers' choice of a particular demonstrative is primarily determined by common ground rather than by type of reference.

A second field study is the article by Mesh et al., which is concerned with demonstratives in Quiahije Chatino, an indigenous language of Mexico. Specifically, this paper investigates the potential influence of the scale of the search space on the multimodal use of demonstratives. The study shows that speakers are more likely to combine demonstratives with a pointing gesture when the reference space involves a large scale rather than a local activity.

Reile et al. is an experimental study that investigates the influence of distance and salience on the alternation between proximal and distal terms in two different dialects of Estonian. There is good evidence that both factors influence speakers' choice of demonstratives; yet, there is an interesting difference between the two dialects, which the authors explain by the fact that in the northern variety demonstrative pronouns do not encode distance.

The final four papers are concerned with non-spoken languages. The article by Garcia and Sallandre, according to which anaphora is only one type of deixis, develops the theoretical framework known as the 'Semiological Approach' to analyse reference in sign language, recognizing the crucial importance of the role of gaze.

Wilcox and Martínez investigate the conceptualization of space in Argentine Sign Language based on video recordings. The authors show that space is used in various ways for creating reference in Argentine Sign language. Moreover, they argue, in accordance with Talmy, that deixis and anaphora form a continuum in signed languages that involves the same conceptual apparatus.

The paper by Cooperrider et al. examines how pointing is integrated into spoken and signed languages. Using a novel pointing elicitation task, these researchers found that both speakers and signers are especially likely to use a pointing gesture in conjunction with lexical expressions if the latter does not seem to be sufficient to identify the referent. Moreover, they observe that, while speakers tend to use points that span across words, signers' points typically occur in slots between lexical signs.

Finally, the paper by Edwards and Brentari is concerned with demonstratives in a tactile language, i.e., a language used by deafblind people. The analysis draws on video data and considers both synchronic and diachronic aspects of demonstratives. The diachronic data show that demonstratives have acquired new functions in a gradually emerging grammatical system, similar to the grammaticalization of demonstratives in spoken language.

In conclusion, this Research Topic provides a survey of current linguistic and psycholinguistic research on deixis and demonstratives yielding a number of new results that we hope will stimulate future research on this important topic.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This manuscript is an integral part of the research project on deictic communication funded by the European Commission: Deictic Communication–A Multidisciplinary Training (DCOMM), Project Number: 676063. This research was supported by EU H2020 ITN Marie Skłodowska-Curie Action grant agreement no. 676063 (DCOMM) awarded to KC, HD, OC, and colleagues.

REFERENCES

Clark, E. V. (1978). "From gesture to word: on the natural history of deixis in language acquisition," in *Human Growth and Development*, eds J. S. Bruner and A. Garton (Oxford: Oxford University Press), 85?120.

Coventry, K. R., Valdés, B., Castillo, A. and Guijarro-Fuentes, P. (2008). Language within your reach: near-far perceptual space and spatial demonstratives. *Cognition* 108, 889?895. doi: 10.1016/j.cognition.2008.06.010

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463?489. doi: 10.1515/COG.2006.015

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Lang. Cogn.* 5, 239?249. doi: 10.1515/langcog-2013-0017

Gudde, H., Griffiths, D., and Coventry, K. R. (2018). The (spatial) memory game: testing the relationship between spatial language, object knowledge, and spatial cognition. *J. Vis. Exp.* 132:56495. doi: 10.3791/56495

Himmelmann, N. (1997). *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur*. Tübingen: Narr.

Stukenbrock, A. (2015). *Deixis in der Face-to-Face-Interaktion*. Berlin: Mouton de Gruyter.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 07 April 2020
doi: 10.3389/fpsyg.2020.00629

Demonstrative Reference and Semantic Space: A Large-Scale Demonstrative Choice Task Study

Roberta Rocca[1,2,3] *and Mikkel Wallentin*[1,2,4]*

[1] Department of Linguistics, Cognitive Science and Semiotics, School of Communication and Culture, Aarhus University, Aarhus, Denmark, [2] Interacting Minds Centre, Aarhus University, Aarhus, Denmark, [3] Psychoinformatics Lab, Department of Psychology, The University of Texas at Austin, Austin, TX, United States, [4] Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark

Spatial demonstratives (words like *this* and *that*) have been thought to primarily be used for carving up space into a peripersonal and extrapersonal domain. However, when given a noun out of context and asked to couple it with a demonstrative, speakers tend to choose *this* for words denoting manipulable objects (small, harmless, and inanimate), while non-manipulable objects (large, harmful, and animate) are more likely to be coupled with *that*. Here, we extend these findings using the Demonstrative Choice Task (DCT) procedure and map demonstrative use along a wide spectrum of semantic features. We conducted a large-scale (N = 2197) DCT experiment eliciting demonstratives for 506 words, rated across 65 + 11 perceptually and cognitively relevant semantic dimensions. We replicated the finding that demonstrative choice is influenced by object manipulability. Demonstrative choice was furthermore found to be related to a set of additional semantic factors, including valence, arousal, loudness, motion, time and more generally, the self. Importantly, demonstrative choices were highly structured across participants, as shown by a strong correlation detected in a split-sample comparison of by-word demonstrative choices. We argue that the DCT may be used to map a generalized semantic space anchored in the *self* of the speaker, the self being an extension of the body beyond physical space into a multidimensional semantic space.

Keywords: language, semantics, spatial demonstratives, manipulability, the Demonstrative Choice Task

OPEN ACCESS

Edited by:
Harmen Bart Gudde, University of East Anglia, United Kingdom

Reviewed by:
Anna M. Borghi, Sapienza University of Rome, Italy
Alfons Maes, Tilburg University, Netherlands

***Correspondence:**
Mikkel Wallentin
mikkel@cc.au.dk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *25 January 2020*
Accepted: *16 March 2020*
Published: *07 April 2020*

Citation:
Rocca R and Wallentin M (2020) Demonstrative Reference and Semantic Space: A Large-Scale Demonstrative Choice Task Study. Front. Psychol. 11:629. doi: 10.3389/fpsyg.2020.00629

INTRODUCTION

Spatial demonstratives are one of the central ways in which language can be used to coordinate attention and enable social interaction. Words like the pronominal and adnominal forms *this* and *that*, or the adverbs *here* and *there* are among the few undisputed language universals (Diessel, 1999; Kemmerer, 1999). Demonstratives are developmental (Capirci et al., 1996) and evolutionary (Diessel, 2006, 2013; Pagel et al., 2013) cornerstones of language, and are among the most frequent words in the lexicon (Leech et al., 2014; Levinson, 2018).

Demonstratives are *deictic* expressions (from Greek *deixis*, "demonstration and indication"). They can in principle be used to indicate *any* object, and their meaning depends on the context of utterance (Levinson, 1983; Diessel, 1999). Identifying their referent in conversation hinges

on the availability of information in the perceptual context (which objects are available), multimodal cues, such as pointing or gaze cuing (Cooperrider, 2016), expectations, i.e., what the speaker may *intend* to refer to (Levinson, 1983; Clark, 1996) and cues provided by the use of specific demonstrative forms (e.g., a proximal *this* vs a distal *that*).

We primarily use the proximal demonstrative (*this*) to refer to objects within manual reach (Coventry et al., 2008), but demonstratives are also used to establish contrasts in conceptual space, where meaning may be negotiated in the absence of visible objects and interlocutors. Experimentally, the use of specific demonstrative forms has been found to reveal information about the speaker's relationship to the referenced object (e.g., ownership, familiarity; see Coventry et al., 2008, 2014; Rocca et al., 2019b) and about the conversational situation (Peeters and Özyürek, 2016; Rocca et al., 2019c). More generally, demonstratives may signal information about the functional status of the object and its affordances for interaction with respect to the speaker and/or the dyad (Jungbluth, 2003). In line with this, listening to demonstratives embedded in a dialogue has been shown to yield activation in the brain's dorsal parietal cortices, suggesting a link between demonstrative use and where/how processing pathways (Rocca et al., 2019a). These findings show that demonstratives serve a fundamental role in linking language with non-linguistic cognition in order to guide joint attention during communication (Diessel, 2006).

In a recent study (Rocca et al., 2019b), we introduced the *Demonstrative Choice Task* (*DCT*), a new experimental paradigm where participants are asked to match nouns (e.g., *apple* or *tiger*) with a demonstrative (i.e., *this* or *that*) without any further context. Across three languages, we found that participants consistently use the distal demonstrative (*this*) for a word like *apple*, whereas they consistently choose *that* for a word like *tiger*. This effect was interpreted to be related to the inferred manipulability of the object, a compound metric combining perceptual (size), psychological (harmfulness), and semantic dimensions of the object. This is in line with research suggesting that demonstratives are interconnected with kinematic planning (Bonfiglioli et al., 2009; Rocca et al., 2018; Caldano and Coventry, 2019) and interactional affordances (Rocca et al., 2019c), rather than being mere distance indicators.

In this experiment, we further validated the DCT and explored whether semantic dimensions other than manipulability affect how speakers couple demonstratives and content words in the absence of context. First, we attempted to establish whether the distribution of demonstrative choices for particular words (i.e., how often a word is coupled with either *this* or *that*) are reproducible across a large set of words. Secondly, we aimed to replicate our previous finding that word meaning related to manipulability affects demonstrative choice. Thirdly, we tested if additional semantic domains have an influence on demonstrative choice, thus providing a comprehensive characterization of the relationship between semantics and demonstrative use. Lastly, we trained a classifier to investigate the degree to which individual trial level choices of *this* or *that* for particular words can be predicted by word semantics.

Demonstrative use depends on the establishment of an "origio," serving as the frame of reference from which an utterance is constructed (Bühler, 1934/2011; Diessel, 2014). The semantic interpretation of *here* and *this* etc. thus presupposes a coordinate system anchored by some entity, usually the speaker's body. However, we also know that spatial demonstratives can be used to denote non-spatial semantic features, such as time (e.g., *this time*), events (*this event*), emotions (*this emotion*), phenomenology (*this experience*), and abstract notions (*this abstraction*), that have no clear spatial anchoring. More generally, as noted by Bühler (1934/2011), deictic reference can be used in an imagination-oriented fashion ("deixis am Phantasma"), i.e., to refer to non-spatial entities such as discourse elements (anaphoric use), memories, imagined scenes, or other products of "constructive phantasy."

Following this line of reasoning, we speculate that, when demonstratives are used to denote referents not immediately available in perceptual space, the proximal/distal distinction is anchored on a reference frame centered on the speaker's *self*. The notion of self *includes* the speaker's body but extends beyond the body to include multiple semantic dimension such as temporality (i.e., discourse markers such as anaphora), emotions, phenomenology, and social embeddedness (see Hanks, 2009; Stukenbrock, 2014 for similar suggestions). When non-spatial semantic entities are referred to, the interpretation of the proximal/distal distinction may thus be given by the position of the referent in a coordinate system consisting of psychological (e.g., familiarity and affect), semantic, and imaginative dimensions, anchored by the speaker's position within that space.

In this study, we investigated demonstrative use in the latter scenario. We elicited demonstratives by presenting participants with concrete content words. No further context was provided. The words were rated along a comprehensive set of perceptually and psychologically relevant semantic dimensions (Binder et al., 2016). We expected the position of words (and the referred entities) within this semantic hyperspace to influence participants' choices of demonstratives. Not all aspects of an object's semantics might be equally relevant when choosing between proximal and distal demonstrative referencing expressions, and some dimensions are unlikely to contribute at all.

Using the simple behavior elicited by the DCT, we attempted to find out which individual features in the included set of semantic dimensions have an influence on speakers' choices for specific demonstrative forms, and to estimate the extent of such effects.

MATERIALS AND METHODS

Participants

We conducted a large-scale DCT experiment using Qualtrics[1] with participants recruited through the Prolific website[2]. A total

[1]http://qualtrics.com

[2]http://prolific.ac

of 2,197 native English-speakers participated (gender: 1,364 female, 819 male, and 13 other; age: 801 were 18–30 years, 693 were 30–40 years, 347 were 40–50 years, 244 were 50–60 years, and 111 were 60+ years). The study was approved by the Institutional Review Board at Aarhus University.

Procedure

The study took on average 4 min to complete, and participants were rewarded with 0.42 GBP for participation. Participants were presented with 48 or 49 words, selected from a database of 535 words, which have been rated on 65 different semantic dimensions, comprising sensory, motor, spatial, temporal, affective, social, and cognitive experiences (Binder et al., 2016). The 535 words were divided into 11 subsets, and participants were presented with one such subset of words in a pseudorandomized manner. Similar to our previous experiment (Rocca et al., 2019b), participants were asked to couple each word with either the spatial demonstrative *this* or with *that* without further context. They were instructed to simply follow their intuition and choose the combination of demonstrative and word they thought fitted best.

Materials

The 65 semantic dimensions that words are rated along in the Binder dataset are: *Vision*, *Bright*, *Dark*, *Color*, *Pattern*, *Large*, *Small*, *Motion*, *Biomotion*, *Fast*, *Slow*, *Shape*, *Complexity*, *Face*, *Body*, *Touch*, *Temperature*, *Texture*, *Weight*, *Pain*, *Audition*, *Loud*, *Low*, *High*, *Sound*, *Music*, *Speech*, *Taste*, *Smell*, *Head*, *UpperLimb*, *LowerLimb*, *Practice*, *Landmark*, *Path*, *Scene*, *Near*, *Toward*, *Away*, *Number*, *Time*, *Duration*, *Long*, *Short*, *Caused*, *Consequential*, *Social*, *Human*, *Communication*, *Self*, *Cognition*, *Benefit*, *Harm*, *Pleasant*, *Unpleasant*, *Happy*, *Sad*, *Angry*, *Disgusted*, *Fearful*, *Surprised*, *Drive*, *Needs*, *Attention*, and *Arousal* (see **Figures 1**, **2** for illustrations of these features). The database is publicly available at: http://www.neuro.mcw.edu/representations/index.html, and the rationale for the choice of these exact features is that they represent "experiential phenomena for which there are likely to be corresponding distinguishable neural processors" (Binder et al., 2016). The notion that these features should have clearly defined neural underpinnings suggests that they are somehow important and representative for human cognition (see Binder et al., 2016 for further details).

One of the aims of the present work was to test the replicability of results from Rocca et al. (2019b), where manipulability is argued to play a role in demonstrative choice. The Binder et al. (2016) dataset does not provide an explicit manipulability dimension. We initially attempted to extract a proxy for manipulability applying principal component analysis and factor analysis to the Binder dimensions. However, we found no component that could straightforwardly be interpreted as manipulability. We therefore added to our feature set the Lancaster Sensorimotor Norms[3]. This dataset provides ratings along 11 sensorimotor features for a large body of words (Lynott et al., 2019). The 11 dimensions are the following (the affix *Lan* is appended to differentiate from features from the Binder dataset): *Auditory_Lan*, *Gustatory_Lan*, *Haptic_Lan*, *Interoceptive_Lan*, *Olfactory_Lan*, *Visual_Lan*, *Foot_leg_Lan*, *Hand_arm_Lan*, *Head_Lan*, and *Mouth_Lan* (see **Figures 1**, **2** for illustrations of the features ordered according to semantic factors, obtained by factor analysis).

The overlap between the two databases included 506 out of the original 535 word, i.e., all words for which semantic ratings were available in both the Binder et al. (2016) and the Lancaster Sensorimotor Norms dataset. All subsequent analyses are conducted on this subset of the data, using the 65 + 11 semantic feature set. All feature ratings were standardized to make them comparable. Two Binder features contained missing ratings for particular words. These were imputed using the mean of all other words along that feature.

Factor Analysis

We reduced the dimensionality of the semantic space using factor analysis. This was aimed at lowering the number of correlated regressors to be used in statistical analyses while preserving structural factors of the semantic space. To determine the number of latent factors, we used Horn's parallel method (Horn, 1965), implemented in the psych package (Revelle, 2019) in R. This method compares the scree plot from the observed data with one made from random samples (randomized across rows) of the original data, and subtracts out the components that explain less variance than a comparable factor based on non-informative data [see analysis script (text footnote 3) for an illustration]. The estimated number of non-random factors in the semantic features using this procedure was 12. Factor analysis was conducted using ordinary least squares (OLS) to find the minimum residual (minres) solution. Orthogonal rotation (varimax) was applied. The cumulative proportion of variance of the semantic features explained by the 12 factors was 0.75.

Factors were labeled by the authors by inspecting the features yielding the highest factor loadings (see **Figures 1**, **2**). The 12 factors and the proportion of the variance they explained in the original semantic space were: *Vision* (0.14), *Valence* (0.11), *Loudness* (0.09), *Human* (0.06), *Taste/Smell* (0.06), *Motion* (0.06), *Manipulability* (0.06), *Scene* (0.05), *Time* (0.03), *Torso/Legs* (0.03), *Arousal* (0.03), and *Self* (0.03) (see **Figures 1**, **2**). It is important to note that these factors and the relative variance they explain do not necessarily reflect the general distribution in language or semantics, but only in the underlying sample of words and features present in the combined Binder and Lancaster databases. The ordering of the factors is therefore also partly specific to those words.

The 12 factors were used as predictors in two analyses (see below for details): (1) an aggregate-level linear regression analysis investigating the role of semantic dimensions in the distribution of demonstrative choices for words; (2) a logistic regression classifier investigating the degree to which trial by trial choices of *this* or *that* can be predicted by semantic factor scores of the experimental words.

[3] Available at https://osf.io/7emr6/

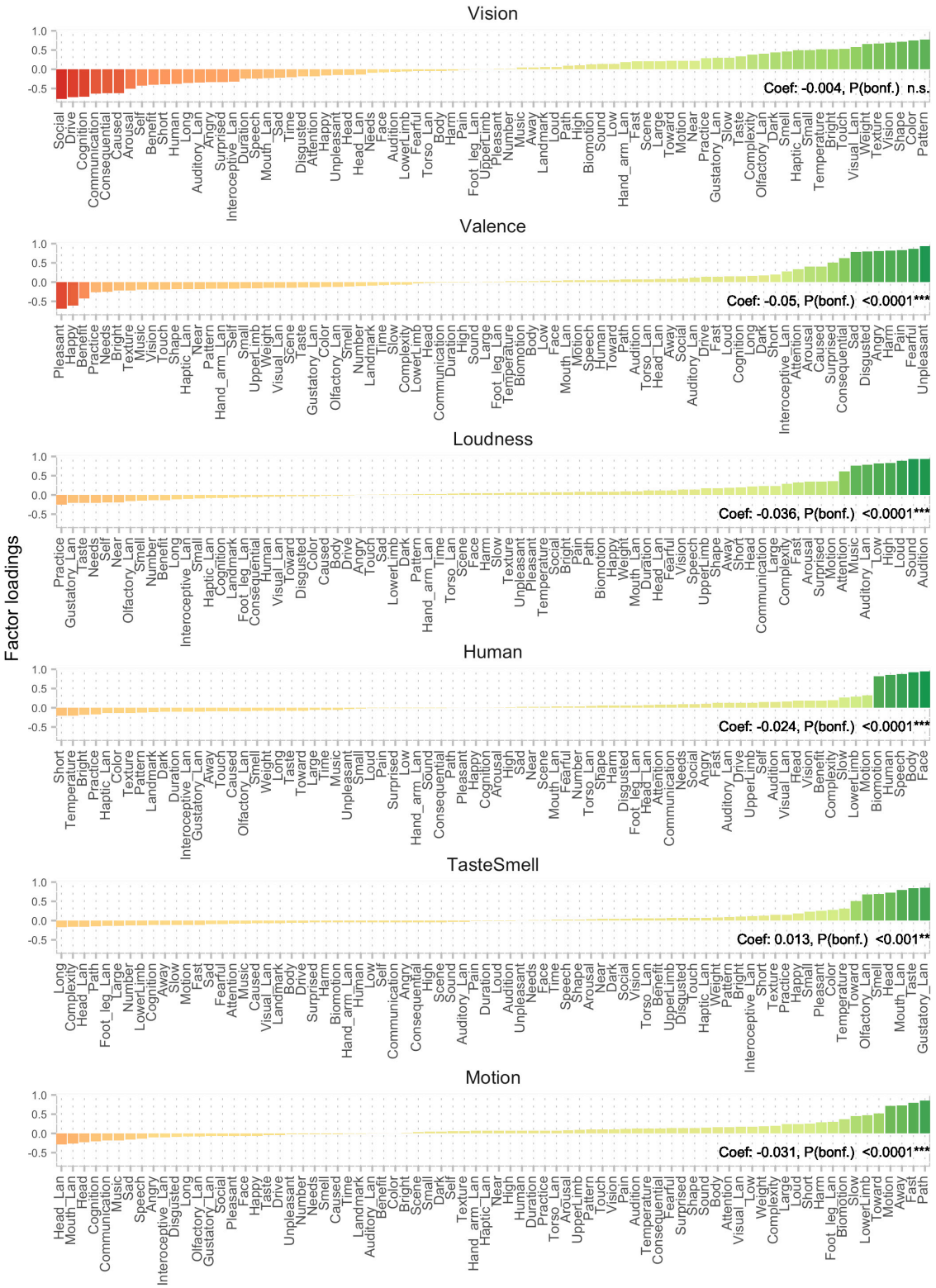


FIGURE 1 | Factor analysis on a combination of Binder and Lancaster features resulted in 12 factors. Here, factors 1–6 are displayed (see **Figure 2** for factors 7–12), with features ordered by loading. Factors are labeled by the authors. Coefficients reflect aggregate level regression results. A significant positive coefficient means that positive (green) sematic features are likely to elicit a proximal demonstrative, whereas features with negative (red) loadings tend to elicit distal demonstratives. When the coefficient is negative, the effect of the factor is reversed in the regression, i.e., features with positive loadings (green) are more likely to elicit distal demonstratives.

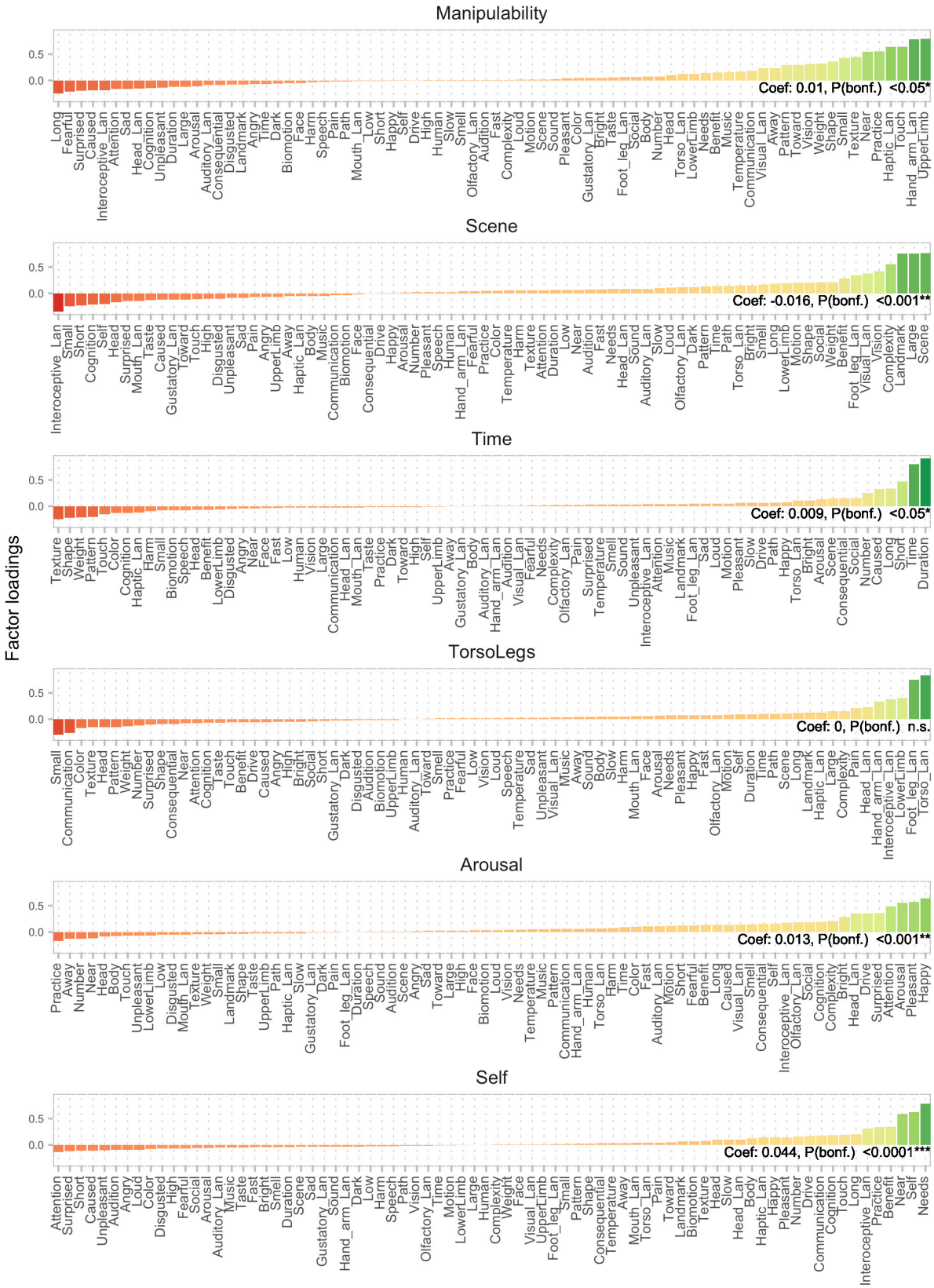


FIGURE 2 | Factor analysis on a combination of Binder and Lancaster features resulted in 12 factors. Here, factors 7–12 are displayed, with features ordered by loading. Factor 7 (top panel) represents manipulability, which was hypothesized and found to explain demonstrative choice together with nine other semantic factors. Coefficients reflect aggregate level regression results. See **Figure 1** for additional details.

Inferential Aggregate Level Analyses

We first analyzed the data at an aggregate level, focusing on the overall proportion of proximal demonstratives chosen for each word as outcome variable.

The first aim of the analysis was to investigate the consistency in demonstrative choices across participants and words. We divided the data into two participant subsamples and calculated the proportion of proximal demonstratives chosen for each word in each sample. This yielded a vector of 506 proportion values (one per word) per participant sample. If participants' choices of demonstrative forms for each word were random or inconsistent, we would expect the two vectors to be uncorrelated or only very weakly correlated. A strong correlation would speak in favor of participants' coupling of demonstratives and words being highly structured and thus, at least to some extent, predictable.

Secondly, we conducted a linear regression analysis with the overall proportion of proximal demonstratives chosen for each word as dependent variable and the 12 factors as independent variables. This allowed to determine which (if any) semantic factors could be used to predict demonstrative choices.

Trial-Level Classification Analysis

To examine the degree to which semantic factors could predict demonstrative choice at the single trial level, i.e., to determine how often word semantics could predict the choice of *this* and *that* on individual trials, we conducted a logistic regression classification analysis using the caret package in R. Individual trial data were initially divided into a training set (80%) and a test set (20%). The test set did not include any data from participants who were part of the training set.

The training set was used to conduct a logistic regression with 10-fold cross-validation. Again, we made sure that each fold in the cross-validation procedure did not contain data from participants that the data had been trained on.

The performance of the model was evaluated using a Monte Carlo permutation test (Ernst, 2004). Here, prediction performance is evaluated in terms of the probability of the observed prediction accuracy given the null. The null distribution of the prediction metric is obtained by randomly permuting the outcome labels, and fitting the model of interest to the permuted labels. To obtain the null distribution, we performed 1,000 permutations on all the data in the training set and obtained a probability value for our prediction score under the resulting distribution.

RESULTS

Descriptive Results

The overall proportion of proximal/distal demonstratives in the data was 0.465/0.535 (standard deviation of proportion of proximal demonstratives across words: 0.114).

Aggregate Level Results

In the split-sample reliability analysis, the proportion of proximal demonstratives was highly correlated across the two samples [$r = 0.82$, $t(503) = 32.7$, $p < 0.0001$; **Figure 3**], which speaks in favor of participants' choices of demonstrative forms not being random.

The linear regression model with semantic factors (**Figures 1**, **2**) as independent variables and overall proportion of proximal demonstratives and dependent variable was highly significant (adjusted $R^2 = 0.6018$), indicating that the semantic factors explained variability in the distribution of proximal/distal demonstratives.

Out of the 12 semantic factors, 10 significantly contributed to the model ($p < 0.05$, Bonferroni corrected): *Valence* [$t(493) = -15.6$, $p < 0.0001$], *Loudness* [$t(493) = -11.3$, $p < 0.0001$], *Human* [$t(493) = -7.4$, $p < 0.0001$], *Taste/Smell* [$t(493) = 4.0$, $p < 0.001$], *Motion* [$t(493) = -9.4$, $p < 0.0001$], *Manipulability* [$t(493) = 3.1$, $p < 0.05$], *Scene* [$t(493) = -4.5$, $p < 0.0001$], *Time* [$t(493) = 2.9$, $p < 0.05$], *Arousal* [$t(493) = -4.0$, $p < 0.001$], and *Self* [$t(493) = 13.0$, $p < 0.0001$]. The factors *Vision* and *Torso/Legs* were non-significant ($p > 0.05$). Positive coefficients (see **Figures 1**, **2**) and t-values indicate that the factor contributes positively to the choice of proximal demonstratives (i.e., elicits *this* more often), whereas negative coefficients and t-values indicate a negative contribution to the choice of proximal demonstratives (i.e., elicits *that* more often).

Single Trial Level Results

The logistic regression model with semantic factors as dependent variables and individual trial choices of proximal/distal demonstratives as predictor led to a prediction accuracy of 57.61% on the training data and an accuracy of 58.40% on the test data (chance level 53.54% in the training set and 53.32% in the test set). All the 1,000 permutations in the null-distribution were lower than these values (range: 53.50–53.58%, based on the training set), indicating that the probability of the model belonging to a random distribution is <0.001.

Out of the 12 semantic factors, 10 significantly contributed to the model ($p < 0.001$, Bonferroni corrected): *Valence* ($z = -28.323$), *Loudness* ($z = -20.152$), *Human* ($z = -12.946$), *Taste/Smell* ($z = 6.657$), *Motion* ($z = -17.105$), *Manipulability* ($z = 5.173$), *Scene* ($z = -9.671$), *Time* ($z = 6.082$), *Arousal* ($z = 7.213$), and *Self* ($z = 22.909$). The factors *Vision* and *Torso/Legs* were non-significant ($p > 0.05$). The results thus closely mirror those from the aggregate level.

A linear combination of factor loadings and regression coefficients for the 10 significant components allows us to project the effects back into feature space (**Figure 4**). This shows how valence is an important driver of demonstrative choice, in combination with self-relatedness, proximity and features relevant for manipulability. Negative valence, motion, and loudness drive choices toward the distal demonstrative.

DISCUSSION

We documented that the DCT (a seemingly meaningless task) yields highly reproducible results. The proportion of proximal demonstratives for specific words in one randomly selected split of the data closely matched the proportion in the complementary

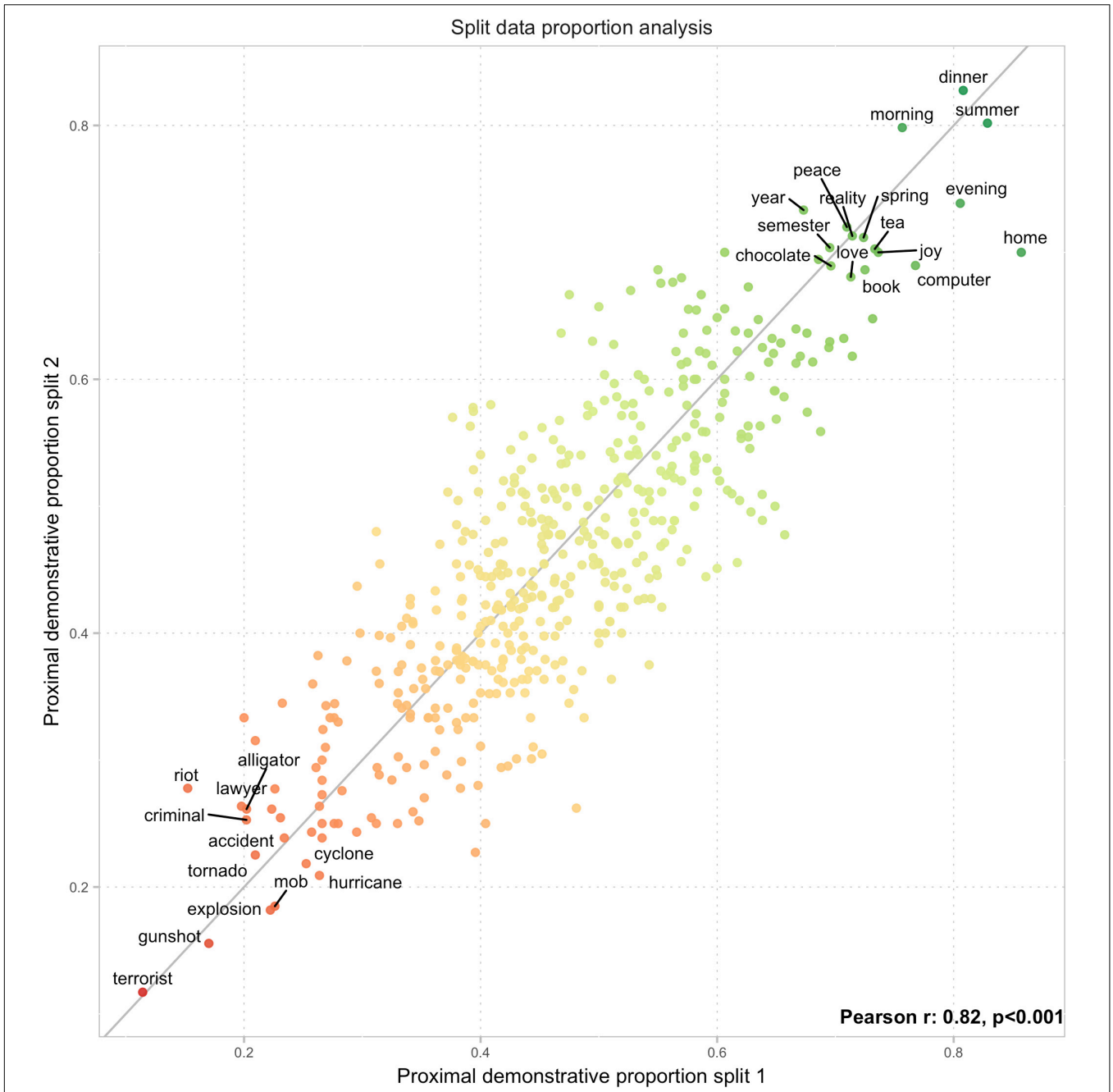


FIGURE 3 | Aggregate proportion of proximal demonstratives in two data splits show a high degree of reproducibility (r = 0.82) in aggregate demonstrative choices for words across semantic categories. Words deviating more than two standard deviations from mean aggregate proportion are reproduced. Valence effect is clearly visible; negative effect of words denoting loudness as well as a positive effect of words relating to temporal events also seem to be visible.

participant sample (see **Figure 3**). This shows that the DCT is a reliable method for characterizing the relationship between demonstrative use and word meaning.

We additionally replicated the result (Rocca et al., 2019b) that affordances for manipulability predict the choice of proximal demonstratives. This effect, however, was found in a combination with nine additional semantic factors that also contributed to demonstrative choice.

On the face of it, the manipulability effect in the current experiment seems less pronounced than the one found in our previous study (Rocca et al., 2019b). The regression coefficient is smaller than several other factors (see **Figures 1**, **2**), suggesting that the manipulability factor is not the main driver of semantic effects in this experiment. This is also clearly visible in **Figure 4**, where semantic features related to manipulability are overshadowed by those related to valence etc. However,

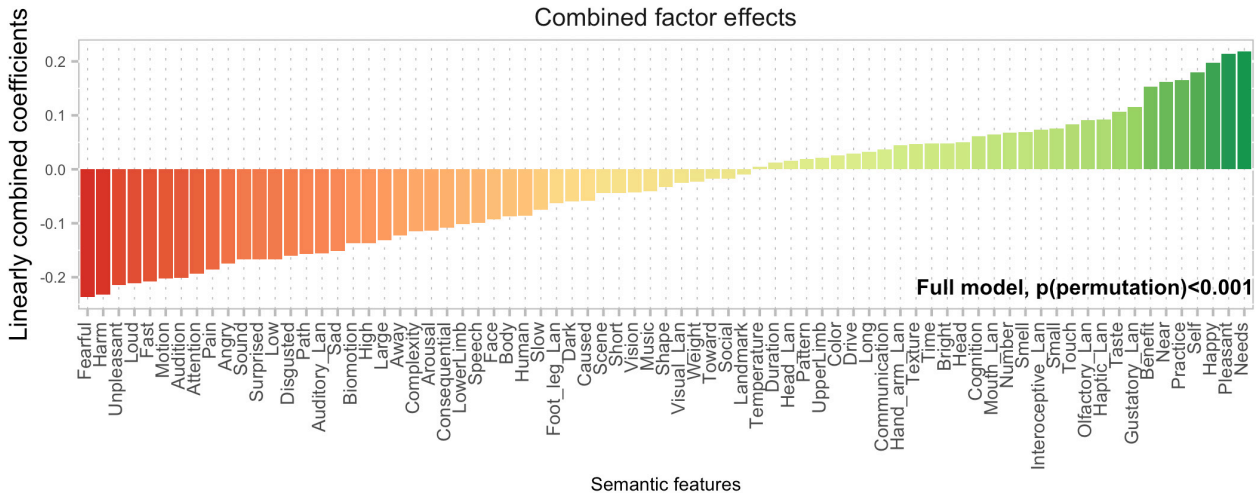


FIGURE 4 | A linear combination of factor loadings and regression coefficients for the 10 significant components shows which semantic features drive demonstrative choice for proximal (green) and distal (red) demonstratives.

manipulability can be more or less narrowly defined. In the previous study (Rocca et al., 2019b) manipulability was defined along three dimensions: "Can you move it with your hands?", "Do you want to move it with your hands?" and "Will it let you move it with your hands?" Together, these dimensions yield a broad definition of manipulability that includes valence ("Do you want to move it. . .") and animacy ("Will it let you move it. . ."). In the present experiment, the manipulability factor spans a narrower semantic space, more akin to "Can you move it. . .," while leaving the other elements of the broader definition to, e.g., the *Valence* and to some degree the *Motion* factors. The present results thus provide evidence that demonstrative choice is affected by manipulability, even with this narrower definition of the term.

When combining the effects of semantic factors obtained via factor analysis and projecting them back into the original feature space, we find that features related to the experiential self (e.g., *Needs*, *Pleasant*, and *Happy*) dominate over features related to proximity and the physical self (e.g., *Near*, *Haptic_lan*). Whether this effect reflects a hierarchy presents outside the experiment or whether it is brought about by the format of the present experimental paradigm (where objects are not spatially available) remains to be investigated. Regardless, this study clearly shows that demonstrative choices signal an entity's status along a wide array of semantic dimensions. Given the importance of communicating epistemic and emotional information, it may not be very surprising that demonstratives, being universal linguistic tools, can also be brought to work in these domains. Taking this line of thought a bit further, we hypothesize that demonstrative choices in the DCT, and perhaps in naturalistic demonstrative use (although this remains untested in this setup), can be taken as indices of the position of a referent relative to the speaker not only in a physical, but also in a *semantic* space. This builds on the idea that the speaker is the *origo* of the coordinate system against which demonstrative choices are evaluated both in physical and conceptual, psychological, and imaginary hyperspaces.

At the single trial level, we were able to predict 58.5% of the DCT trials in a test sample, based on the semantic profile of the word. The overall proportion of demonstratives was 46.5% for proximal and 53.5% for distal demonstratives, leading the null-distribution to be centered narrowly around 53.5%. A score of 58.5% correct classifications has to be measured both against the chance level of 53.5% and against the upper limit of predictability. The proportion of proximal demonstratives chosen for individual words differed from 50%, on average, with 9.5% (either above or below). Demonstrative choice for a single trial instantiation of a word which overall receives 50%/50% proximal/distal demonstratives can never be predicted above 50% based on information about the word itself, e.g., semantic factor scores. Given semantic scores for a certain word, a statistical model will end up always predicting the same demonstrative for this particular word, which can only be correct in 50% of the cases. Words that receive either a very low or a very high proportion of proximal demonstratives, on the other hand (e.g., *terrorist* or *summer* – see **Figure 3**), can theoretically be predicted with high accuracy. A good model predicting "proximal demonstrative" for a word that received, say, 70% proximal demonstratives, can yield 70% prediction accuracy for this word. The upper limit for classification thus becomes how far from 50% proximal demonstratives words are *on average*. If, across all words, the average absolute difference between the observed proportion of proximal demonstratives and the chance value of 50% is 9.5%, then the upper limit for predicting single trial choices based on semantics is 59.5%. With that in mind, a prediction rate of 58.5% is very close to ceiling for trial level predictions.

It is possible, however, that additional variables exist beyond word level semantics that systematically influence demonstrative choice and that these, if included in the models, would enable better predictions. To provide an example, it could be hypothesized that a 50%/50% distribution of proximal/distal demonstratives for a word could result from

a particular subsample of participants always choosing the proximal demonstrative for this word, whereas another would never do so. If valence is a guiding principle, one could imagine that a word like *dog* would be given a proximal demonstrative by all dog lovers, whereas people who dislike dogs will use a distal form. This line of reasoning opens up to the possibility that the DCT could be used to probe individual differences among participants and that taking individual differences into account would raise the predictive power of the model. Demonstrative choice may thus be affected by the way preferences, experiences, and personality traits form our individual semantic landscape. If this is indeed the case, participants' response patterns in the DCT could be used as a tool to predict individual differences in personality, a possibility which needs to be explored in further studies.

As a final note, the DCT provides consistent results on the effect of semantic factors in isolation, but demonstratives are most often used in exophoric contexts with an actual spatial configuration of speaker, referent, and addressee. These factors are key to shaping demonstrative use in naturalistic settings, and they are likely to interact with semantics, e.g., introducing contextual modulations of the effect of specific semantic dimensions. Further studies are needed to clarify to what extent and how semantics influences the choice of demonstratives in more contextualized situations where competing forces govern the selection. Experimentally assessing the relative role of these factors and their interactions may also inform computational models predicting demonstrative choice, a hitherto unexplored avenue in the field.

CONCLUSION

In this study, we validated the use of DCT as a simple method to characterize the relationship between demonstrative use and semantic spaces. We have found that demonstrative choice is influenced by multiple semantic dimensions, including spatial, bodily, and emotional features, and we have showed that demonstratives might be consistently used to signal the relation between speakers and objects not only within physical space, but also in semantic hyperspaces. Further developments to the paradigm may increase the predictive power of the DCT as well as revealing new practical applications.

DATA AVAILABILITY STATEMENT

All data and scripts for analyses are available from Open Science Framework: https://osf.io/tqejb/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Aarhus University's Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RR and MW devised the experiment, conducted the statistical analyses, wrote the manuscript together, and approved the manuscript before submission.

FUNDING

Funding for RR was provided by the DCOMM grant from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Actions grant agreement no. 676063.

ACKNOWLEDGMENTS

The authors wish to thank Camilla M. Andersen for her help with setting up the link between Qualtrics and Prolific.

REFERENCES

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., and Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cogn. Neuropsychol.* 33, 130–174. doi: 10.1080/02643294.2016.1147426

Bonfiglioli, C., Finocchiaro, C., Gesierich, B., Rositani, F., and Vescovi, M. (2009). A kinematic approach to the conceptual representations of this and that. *Cognition* 111, 270–274. doi: 10.1016/j.cognition.2009.01.006

Bühler, K. (1934/2011). *Theory of Language – The Representational Function of Language [Sprachtheorie]*. Amsterdam: John Benjamins.

Caldano, M., and Coventry, K. R. (2019). Spatial demonstratives and perceptual space: to reach or not to reach? *Cognition* 191:103989. doi: 10.1016/j.cognition.2019.06.001

Capirci, O., Iverson, J. M., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two-word speech. *J. Child Lang.* 23, 645–673. doi: 10.1017/s0305000900008989

Clark, H. H. (1996). *Using Language*. New York, NY: Cambridge University Press.

Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53, 632–656. doi: 10.1080/0163853x.2015.1094280

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70 doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near-far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins Publishing.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Lang. Cogn.* 5, 239–249. doi: 10.1515/langcog-2013-0017

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Lang. Linguist. Compass* 8, 116–132. doi: 10.1111/lnc3.12066

Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Statist. Sci.* 19, 676–685. doi: 10.1214/088342304000000396

Hanks, W. F. (2009). Fieldwork on deixis. *J. Pragmat.* 41, 10–24. doi: 10.1016/j.pragma.2008.09.003

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/bf02289447

Jungbluth, K. (2003). "2. Deictics in the conversational dyad: findings in spanish and some cross-linguistic outlines," in *Deictic Conceptualisation of Space, Time and Person*, ed. F. Lenz (Amsterdam: John Benjamins).

Kemmerer, D. (1999). 'Near' and 'far' in language and perception. *Cognition* 73, 35–63. doi: 10.1016/s0010-0277(99)00040-2

Leech, G., Rayson, P., and Wilson, A. (2014). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Routledge.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2018). "Introduction: demonstratives: patterns in diversity," in *Demonstratives in Cross-Linguistic Perspective*, eds M. J. Dunn, N. J. Enfield, S. Cutfield, S. C. Levinson and S. Meira (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781108333818.002

Lynott, D., Connell, L., Brysbaert, M., Brand, J., and Carney, J. (2019). The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behav. Res. Methods* doi: 10.3758/s13428-019-01316-z

Pagel, M., Atkinson, Q. D., Calude, S. A., and Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proc. Natl.Acad. Sci. U.S.A.* 110:8471. doi: 10.1073/pnas.1218726110

Peeters, D., and Özyürek, A. (2016). This and that revisited: a social and multimodal approach to spatial demonstratives. *Front. Psychol.* 7:222. doi: 10.3389/fpsyg.2016.00222

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Rocca, R., Coventry, K. R., Tylén, K., Lund, T. E., and Wallentin, M. (2019a). Language beyond the language system: dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI. *Neuroimage* doi: 10.1016/j.neuroimage.2019.116128

Rocca, R., Tylén, K., and Wallentin, M. (2019b). This shoe, that tiger: semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLoS One* 14:e0210333. doi: 10.1371/journal.pone.0210333

Rocca, R., Wallentin, M., Vesper, C., and Tylén, K. (2018). "This and that back in context: grounding demonstrative reference in manual and social affordances," in *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, Madison, WI.

Rocca, R., Wallentin, M., Vesper, C., and Tylén, K. (2019c). This is for you: social modulations of proximal vs. distal space in collaborative interaction. *Sci. Rep.* 9:14967. doi: 10.1038/s41598-019-51134-8

Stukenbrock, A. (2014). Pointing to an 'empty' space: deixis am phantasma in face-to-face interaction. *J. Pragmat.* 74, 70–93. doi: 10.1016/j.pragma.2014.08.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 July 2020
doi: 10.3389/fpsyg.2020.01406

The Conceptualization of Space: *Places* in Signed Language Discourse

Sherman Wilcox[1] and Rocío Martínez[2]*

[1] Department of Linguistics, University of New Mexico, Albuquerque, NM, United States, [2] Instituto de Lingüística, Facultad de Filosofía y Letras, Universidad de Buenos Aires and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

We examine the conceptualization of space in signed language discourse within the theory of cognitive grammar. Adopting a Places view, we define Place as a symbolic structure that associates a schematic semantic pole and a schematic phonological pole. Places acquire full contextual meaning and a specific spatial location in the context of a usage event. In the present article, we analyze the referential function of Places in different grammatical constructions throughout a selection of videos produced by deaf Argentine Sign Language signers. Our analysis examines Places, which are associated with entities in the surrounding spatial environment as well as Places that are created or recruited in discourse without reference to surrounding physical entities. We observe that Places are used in pointing, placing, and other grammatical constructions in order to introduce and track referents in ongoing discourse. We also examine the use of conceptual reference points, by which Places afford mental access to new related concepts that are the intended focus of attention. These results allow us to discuss three related issues. First, for signed language discourse, space is both semantically and phonologically loaded. Signers' semantic and phonological choices for Place symbolic structures are motivated by embodied experience and the abstraction of usage events. Second, Places occur along a continuum from deixis to anaphor, united by the same conceptual system and differing in extent of phonological subjectification. Third, we suggest developmental implications of our Place analysis.

Keywords: signed language, cognitive grammar, pointing, reference points, deixis

OPEN ACCESS

Edited by:
Olga Capirci,
Italian National Research Council, Italy

Reviewed by:
Terry Janzen,
University of Manitoba, Canada
Virginia Volterra,
Italian National Research Council, Italy

****Correspondence:***
Sherman Wilcox
wilcox@unm.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 February 2020*
Accepted: *26 May 2020*
Published: *08 July 2020*

Citation:
Wilcox S and Martínez R (2020) The Conceptualization of Space: Places in Signed Language Discourse. Front. Psychol. 11:1406. doi: 10.3389/fpsyg.2020.01406

INTRODUCTION

Signed languages are uniquely suited for studying the conceptualization of space. Signs are produced by moving hands in three-dimensional space. As Stokoe (1980) observed, "In producing a sign language utterance, some part (or parts) of the signer's body acts. If the active part is mobile enough, there are various places in which the action may occur, i.e., begin, take place, or end." These three aspects of a signed utterance have been recognized as three basic phonological parameters: handshape (active part), location (place), and movement (action). Of the three, the most significant for revealing how signers conceptualize space is location. Signs were originally described as incorporating locations on or near the body and those that are produced in an unmarked three-dimensional signing space in front of the signer's head and torso, extending from

a little above the head to a little below the waist. It is signs that are produced in this signing space that we examine, and our primary focus is their referential function.

Research on reference in signed languages has been closely connected to the use of space. Based on the theoretical framework used and the similarities of their proposals, these studies can be grouped into four main views on the use of space for achieving reference in signed languages: (1) the referential locus (R-locus) view, (2) the mental spaces view, (3) the locus with semantic-pragmatic conventions view, and (4) the symbolic Places view.

Researchers working within formalist theory adopt a *referential locus (R-locus) view*, claiming that spatial locations are used for identifying referents previously associated with that location. These are called R-loci. R-loci are distinguished from referential indexes (R-indexes): The former are the physical spatial locations toward which a signer points, whereas the latter are abstract formal devices indicating reference within and across sentences (Lillo-Martin and Klima, 1990). Within this view, the location in space for achieving reference (R-locus) is randomly chosen by the signer. More recent research claims that, whereas abstract indices are part of the grammar, loci are determined outside of grammar. This leads some proponents of this view to a provocative conclusion: "On our view, the grammar doesn't care which point in space is used for a particular referent. Abstract indices are part of the grammar, but loci are determined outside of grammar. Therefore, the connection between referents and loci requires language to interface with gesture" (Lillo-Martin and Meier, 2011, p. 121)[1].

The *mental spaces view* is based on mental space theory (Fauconnier, 1985, 1997). As applied to signed languages, its main proponent is Liddell (1995). In his first approach, Liddell proposed that three mental spaces are recruited for creating and maintaining reference in American Sign Language (ASL) discourse: real space, surrogate space, and token space. Real space is a person's current conceptualization of the immediate environment based on sensory input. Real space is used when the signer refers to entities that are conceptualized as being physically present, such as using a pronoun toward the addressee or toward objects that are present in the physical situation. Surrogate space describes a type of full-sized, invisible entity. Pronouns and indicating verbs make reference to a surrogate by being directed toward it. Tokens are entities that, like surrogates, are given manifestation in physical space. The difference is that, unlike surrogates, tokens use a limited size of the signing space in front of the signer and only assume third person roles in discourse. Liddell (2003) later revised this theory, following blending theory (Fauconnier and Turner, 1996), showing how real, surrogate, and token space become part of different blended mental spaces[2].

[1]We should point out that not all linguists working within the formalist tradition agree with this claim. Two prominent opponents are Quer (2011) and Wilbur (2013).

[2]Blending is a process that operates over two mental spaces as inputs. Structures from each of the two input spaces is projected to a third space, referred to as the blend. The blend inherits partial features from each of the input spaces and also includes features that belong only to the blend.

Many sign linguists who adopt the mental spaces view claim that the number of locations in space is unlistable and, therefore, cannot be an element of the grammar. According to this claim, any specific instance of a pronoun directed toward an entity will be a combination of lexically fixed features encoding the symbolic pronoun and a non-symbolic pointing direction selected for the specific context in which it is being used (Liddell, 2003). In addition to pronouns, other structures receive similar treatment. For instance, it is claimed that directional verbs, which are called indicating verbs by Liddell, are composed of both lexically fixed features and gestural elements. The actual placement of the hand during the initial or final hold is said to be "gradient" because it depends on the locations of the entities toward which it is directed. Comparable analyses can also be found in research discussing language–gesture fusions (Fenlon et al., 2018).

The *locus with semantic-pragmatic conventions view* (Engberg-Pedersen, 1993) defines locus as an abstract category whose members are specific spatial loci in paradigmatic contrast. Engberg-Pedersen (1993, p. 69) asserts that conventions influence the signer's choice of loci. The space around the signer is semantically "loaded": The choice of a locus for a given referent is not arbitrary, but influenced by semantic and pragmatic conventions. For instance, the convention of semantic affinity states that referents with semantic affinity to each other (for example, a person and the place where she works or a person and his possessions) are usually represented by the same locus unless they need to be distinguished for discourse reasons; the convention of comparison occurs when a signer chooses the locus forward-sideward-left for one referent and the locus forward-sideward-right for another referent when she wants to compare or contrast the two referents. These conventions are neither exhaustive, nor do they have the character of obligatory rules.

We adopt a *Places view* (Wilcox and Occhino, 2016; Martínez and Wilcox, 2019), a usage-based approach developed within the model of cognitive grammar (CG; Langacker, 1987, 1991, 2008; Wilcox, 2014). Our view is grounded in sensory and physical experience and, thus, is an embodied approach in which embodied cognition and experiential conceptual archetypes are fundamental (Langacker, 2006; Barsalou, 2008). Within this approach, the unlistability and the gradience of the locations in signing space are not matters of concern given that we assume a non-structuralist conception of language and its units (Wilcox, 2014). The locations that signers use meaningfully within signing space, as well as any other unit, cannot be conceptualized *a priori* as discrete and categorical, but as elements that arise in a bottom-up fashion. In previous studies, we have called these meaningful locations in signed languages Places (Wilcox and Occhino, 2016; Martínez and Wilcox, 2019). A Place is a symbolic structure, a pairing of a meaning and a location in space that plays a major role in reference in signed languages. Places are, thus, semantically and phonologically substantive, derived from embodied experience and abstraction from actual usage events. Places are components of more complex symbolic structures, such as pointing and placing constructions.

The present article analyzes further dimensions of the Place symbolic structures, using data from Argentine Sign Language

(LSA). Particularly, we focus on Places that create or track different kinds of reference from perceptually accessible entities in the ground to anaphoric referents in discourse. We suggest that deictic and anaphoric constructions, which incorporate Places, are aspects of the same conceptual system and that there is a continuum of Place symbolic structures in signed languages that varies in terms of subjectification. We also explore the way these different Places may function as reference points within larger constructions with the goal of providing mental access to related referents, which are the intended focus of attention.

In section "Cognitive Grammar," we offer a brief background in the basic concepts of CG that are used in our analysis. Section "Pointing and Places" describes our account of pointing and of Places and introduces our proposal of the continuum of Places. In section "Places and Reference Points in Discourse," we examine the use of Places and reference points in discourse. In section "Discussion," we discuss Places in terms of subjectification, examine the implication of Places for infant pointing, and explore Place in relation to the development of demonstratives into grammatical markers. In the conclusion, we offer a summary of our findings and suggest areas deserving further research.

COGNITIVE GRAMMAR

We adopt CG as our theoretical framework for examining the conceptualization of space. The central claim of CG is that only three structures are posited (Langacker, 1987): semantic, phonological, and symbolic. *Semantic structures* are conceptualizations exploited for linguistic purposes. *Phonological structures* include sounds, gestures, and orthographic representations. *Symbolic structures* are the association of phonological and semantic structures such that one is able to evoke the other. The structuralist category of morpheme is viewed as a structure with zero symbolic complexity that has undergone progressive entrenchment and become established as a more or less conventional unit within a language community.

CG claims that lexicon, morphology, and syntax form a continuum of symbolic assemblies comprised of phonological structures, semantic structures, and the symbolic links between the two (Langacker, 1987). Symbolic assemblies vary along two dimensions: schematicity and complexity. *Schematicity* pertains to level of detail or precision. Schematic elements are elaborated or instantiated by more specific elements. Schematicity is, therefore, relational: An element is schematic to a more specific elaboration, and schemas are immanent in these more detailed instantiations. Schematic elements emerge as the result of the cognitive ability to extract and reinforce commonalities across multiple experiences. Symbolic structures also vary along a dimension of complexity. Symbolic structures combine with other symbolic structures to form complex symbolic assemblies. Constructions are symbolic assemblies, composed of component symbolic structures integrated to form a composite structure (Langacker, 2008).

In CG, conceptualization is seen as being "both physically grounded and pervasively imaginative" (Langacker, 2008, p. 539). Thus, grammar incorporates the full scope of our conceptual world and of the physical and spatial world within which we interact with other entities. CG adopts a conceptual semantics based on embodied cognition. Meaning is conceptualization that is grounded in our sensory and physical interactions with the world.

The experiential and embodied nature of cognition is reflected in conceptual archetypes and idealized cognitive models that feature prominently in the organization of grammar. *Conceptual archetypes* are experientially grounded concepts, such as a physical object, an object in a location, an object moving through space, seeing something, holding something, exerting force to effect a desired change, a face-to-face social encounter (Langacker, 2008). One conceptual archetype important to the analyses being offered here is "the common everyday occurrence of physically pointing to something" (Langacker, 2006, p. 34), which is arguably the baseline conception for nominal grounding (Langacker, 2016b). Another conceptual archetype consists in the organization of a scene into a global *setting* and mobile *participants*. "At a given instant, each participant is found at some location. A location is part of the setting (any point or area within it)." (Langacker, 2008, p. 355). This conceptual archetype is manifest in the *stage model*. The term evokes viewers watching action on a stage. We cannot observe the entire auditorium and its audience, the entire stage, and all the actors and action. Therefore, viewers must focus and direct their attention: From the maximal scope of their visual field, they attend only to certain elements, and within that more narrow scope, they focus on specific actors and their actions. This visual *perceptual* description is more than merely a metaphor. The embodied view of cognition claims that our *conceptual* organization also reflects a maximal and *immediate scope* of conception within which certain elements are profiled. A linguistic expression's *profile* is the focus of attention within its immediate scope.

Reference point phenomena are ubiquitous in our experience of the world (Yamanashi, 2015). Reference points rely on our ability to direct attention to a perceptually salient entity as a point of reference to find some other entity, the target (Langacker, 1993). Each reference point provides access to a set of potential targets, called the reference point's dominion. Reference points form the conceptual basis of many constructions, including possessives, topic constructions, metonymy, and pronominal anaphora. Reference point constructions have been shown to play a significant role in the grammars of signed languages (Wilcox and Occhino, 2016; Martínez and Wilcox, 2019).

The *ground* plays a pervasive and essential role in grammar. The ground consists of the speech or sign event, the participants, their interaction, their knowledge, and the time and place of the communicative usage event. The ground features in grammar through *grounding elements*: symbolic structures that specify the status of a nominal or a clause in relation to the ground. For nominals, the primary epistemic concern is identification (Langacker, 2017). Nominal grounding, such as demonstratives, articles, and certain quantifiers, directs the interlocutor's attention to an intended discourse referent (Langacker, 2008; Martínez and Wilcox, 2019). Clausal grounding indicates whether a profiled occurrence has been realized by locating it in relation to the speaker's or signer's conception of reality,

for example, by marking tense and modality (Langacker, 2017). Grounding is, thus, a deictic referential strategy: A deictic expression is one that includes some reference to the ground (Langacker, 1987, p. 126).

All of these principles and models are integrated in discourse. Discourse consists of *usage events*, specific instances of actual language use. Usage events consist of both poles of a symbolic structure: semantic (conceptualization) and phonological (expression). The conceptual pole of usage events "includes the expression's full contextual understanding – not only what is said explicitly but also what is inferred, as well as everything evoked as the basis for its apprehension" (Langacker, 2008, p. 458). The expressive side consists of the full phonetic detail of an utterance. We do not limit usage events to only a single modality: A usage event includes *all perceptible detail*. Usage events have no particular size; depending on level of analysis, a usage event may be a word or sign, clause, conversational turn, or an extended discourse.

Discourse takes place in a discourse space comprising "everything intersubjectively accessible to the interlocutors as the basis for communicating at a given moment in the flow of discourse" (Langacker, 2016b, p. 108). Intersubjective accessibility here means both conceptual accessibility, that which is in the immediate scope of each interlocutor's conceptual space, and perceptual accessibility. Perceptual accessibility includes the ground and the immediate physical context: that which is visibly accessible to the interlocutors.

For signed language discourse, the critical point to recognize is that space plays a role both conceptually and expressively. Signers point to and use physical locations in space to achieve intersubjective reference in discourse. All language users conceptualize space, and thus, space is meaningful in spoken and signed languages, but only signed languages incorporate physical space into their form. The significance of this dual role of space is revealed throughout this paper.

POINTING AND PLACES

In the sections that follow, we present a variety of discourse excerpts from LSA, which incorporate pointing. We analyze pointing as a construction (Wilcox and Occhino, 2016; Martínez and Wilcox, 2019). Pointing constructions consist of two component symbolic structures: a *pointing device* and a *Place*[3]. Both component structures of the pointing construction are symbolic structures consisting of a form, the phonological pole, and a meaning, the semantic pole. One type of pointing device is an index finger, but others may include hand(s), eye gaze, mouth or nose pointing, and body orientation. The pointing device functions to direct attention; this is its schematic meaning. The schematic semantic pole, thus, is dependent, making reference to some autonomous element that is the focus of attention. This is the function of the Place symbolic structure; its semantic pole is the thing referred to, and its phonological pole is the spatial location in the current ground of that referent.

Place structures play a role in a variety of grammatical constructions in LSA and other signed languages. These Place structures are typically quite schematic semantically and phonologically. They acquire full contextual meaning and a specific spatial location in the context of a usage event. One example [A] of the use of the Place symbolic structure is in proxy-antecedent constructions (Wilcox and Occhino, 2016). A proxy-antecedent construction from LSA is shown in **Figure 1** (Martínez and Wilcox, 2019)[4].

The full nominal clause is formed by a possessive (POSS1), an adjective (NEW), a noun (TEACHER), a point to a location in the right of the signing space [POINT(right)], and a relative clause starting with SAME in which there are two more pointing signs. The ungrounded noun TEACHER provides the type specification. The first pointing sign occurs in a proxy-antecedent construction. The antecedent nominal (TEACHER) is grounded deictically by the possessive "my." The proxy-antecedent construction associates the nominal antecedent with a Place, whose schematic meaning is elaborated by the nominal "my new teacher." The proxy-antecedent construction

[3]We capitalize Place to indicate that the term applies to the entire symbolic structure. We use "location" as the term for the phonological pole of Place.

[4]Information about data sources and methodology of coding is provided in **Supplementary Appendix 1**. Glossing conventions are listed in **Supplementary Appendix 2**. Full glosses of all examples are given in **Supplementary Appendix 3**. Citations to data glosses are given in brackets in the main text, e.g., [B] refers to data gloss B in **Supplementary Appendix 3**.

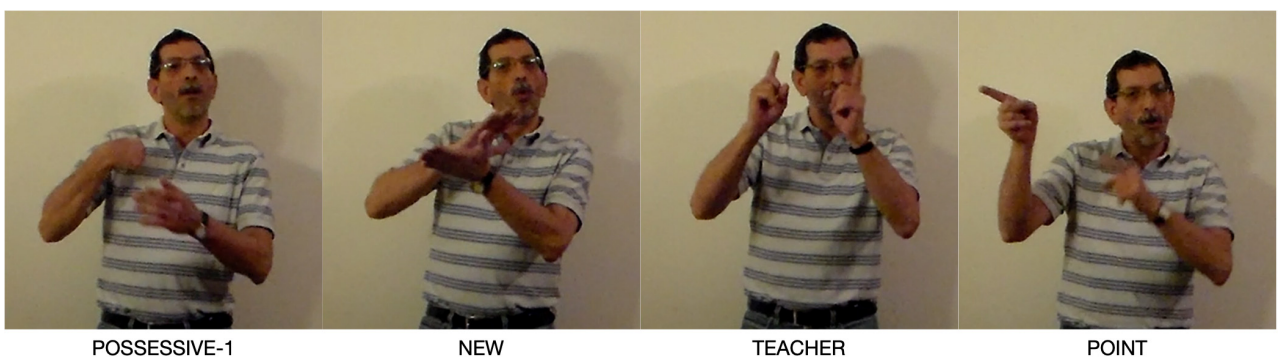


FIGURE 1 | Proxy-antecedent construction.

also carries as part of its conceptual base the meaning that the antecedent will be used later in discourse. This occurs in the relative construction: two pointing signs are used to refer back to the antecedent "my new teacher." Both use the same pointing device to direct attention to the Place on the right of the signer, referring anaphorically to the same antecedent. **Figure 2** depicts this construction. A dotted correspondence line shows that the phonological pole (P) of the two Places structures share the same location in space. Correspondences lines also indicate that the antecedent (TEACHER), the proxy-antecedent (Place), and the anaphor (Place) refer to the same entity; they are coreferential. The dashed rectangle indicates overlap of the three in conceptual space.

The Continuum of Places

The Places examined so far have been grammatical structures functioning to mark proxy-antecedent and anaphor relations. The phonological location of the proxy-antecedent and anaphor Place has nothing to do with actual entities in the spatial environment; the phonological location of Place is entirely specified by the grammar of LSA.

This is not always the case. Signers, like speakers, point to objects in the current discourse environment. It is important to understand the relationship between pointing to physical objects in the spatial environment and pointing to more abstract entities, such as proxy-antecedents and anaphors, whose location is determined by the grammar. Talmy (2018, p. 1) describes the distinction we wish to make as two domains of linguistic reference, "those traditionally termed anaphora and deixis." He goes on to describe these domains: "Broadly, an anaphoric referent is an element of the current discourse, whereas a deictic referent is outside the discourse in the spatiotemporal surroundings. This is a distinction made between the lexical and the physical, one that has traditionally led to distinct theoretical treatments of the corresponding referents." Talmy proposes that these two domains of reference engage the same conceptual system.

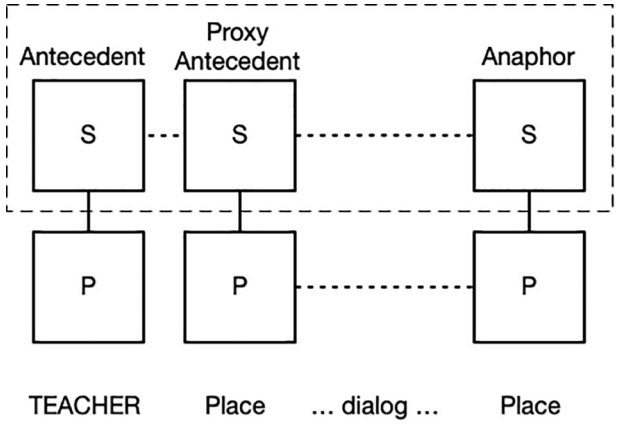


FIGURE 2 | Proxy-antecedent construction.

Engberg-Pedersen (1993) describes deictic and anaphoric frames of reference in signed languages. In deictic reference, the signer points to entities or locations in the context of the utterance. The frame of reference is dependent on the actual locations of those entities or locations. Consequently, if the signer or the entities change their location, the deictic frame of reference changes. Anaphoric frames of reference are independent of the utterance context and, thus, do not change.

Within our approach, we would say that these two domains of reference are not distinct categories; rather, they form a continuum. In order to understand the role of Place along this continuum, we must introduce the concept of "immanence." Immanence has been a central concept in CG since its inception (Langacker, 1979). Immanent means "contained within" or "lies within." Immanence plays a ubiquitous role in grammar, both semantically and phonologically. As we have seen, some units of language are schematic relative to others. Schematic meanings "are immanent in (i.e., they "lie within") those of instantiating expressions, which elaborate them ("flesh them out") in their own individual ways" (Langacker, 2009b, p. 14). The abstracted commonality of a type, such as "dog," is immanent in the conception of any instance of dog. Immanence forms the basis for analyzing a host of expressions, including possessives, epistemic vs. root modality, and grammatical categories.

In all of these cases, the relationship is between the degree of attenuation of semantic units. The classical case is the relation between "going" to mean spatial movement and "go" marking future time. As described by Langacker (2008, p. 538), "In the former case, the conceptualizer scans through time by way of tracking the subject's movement through space. On the future interpretation, this subjective temporal scanning occurs independently of any conception of spatial motion. It is merely a way of mentally accessing an event's location in time." This dynamic semantic relationship between more and less attenuated units is reflected in the concept of subjectification, a semantic shift in which an entity originally construed objectively comes to receive a more subjective construal, which we discuss in more detail in section 5.

We apply these concepts not only to the semantic pole of symbolic structures, but also to the phonological pole. In our usage-based approach, all linguistic units, including phonological units such as location, are abstracted by language users from actual usage events. The units abstracted are immanent in these usage events and motivate new expressions. We also assume that language is grounded in sensory and physical experience in which embodied cognition and experiential conceptual archetypes are fundamental. One conceptual archetype that is central to our proposal of Place is *an object in a spatial location*: this is, in fact, the archetypal source of Place. In pointing to physical entities in a usage event, the conceptualizer produces a pointing construction. Setting aside the pointing device for now, the entity, the thing referred to, is the semantic pole of the Place component of that construction, and the entity's spatial location is its phonological pole. Of course, we point to or otherwise direct attention to any number, in fact an unlimited number, of entities in the environment. As a signer perceives and produces more of these usage events, she abstracts away from the specifics of

any particular entity and its location, developing an ever-more schematic concept of directing attention to an *entity* in a *location*. This is the Place symbolic structure, in which the "entity referred to" is the schematic semantic pole and "some spatial location" is the schematic phonological pole.

Thus, Place symbolic structures are abstracted from actual usage events – in this case, the archetypal usage event being pointing to a physical object in a spatial location. Conceptualizers schematicize these usage events, arriving at a conception representing a higher level of abstraction. This higher level of conception is a schematic Place, which has neither a specific meaning nor a specific spatial location; rather, it associates a schematic meaning with a schematic phonological location. In use, the schematic meaning and the schematic location are instantiated, resulting in a fully contextualized Place symbolic structure.

Schematicity is not an all or none affair; it is a matter of degree, and the path involves attenuation. If pointing to a physical object in a location is the conceptual *baseline*, there are various ways in which this baseline can be *elaborated* (Langacker, 2016a, 2019). One elaboration involves the temporal stability of the object and its location. Suppose that you are sitting in a coffee shop with a friend who is drinking a cappuccino. She points to the cup and says, "This is the best cappuccino I have ever had." She then leaves for a moment, and when she returns her cappuccino has disappeared (probably the waiter thought she was finished and took the cup away). She can point to the location of the now missing cup and say, "Where's my cappuccino?" Now, she is pointing to the Place that was immanent in the cup's spatial location. Even though the cup is no longer physically present in this location, your friend and you remember that it was. In another elaboration, one might return after many years to the house where she grew up and say, "My father's desk was here, my sister's here" and point to their former locations. Here, the elaboration is even more attenuated both because it involves a longer expanse of memory, and for the interlocutor, it requires imagination. Imagination can be used by both the speaker/signer and the interlocutor in further elaborations, such as pointing to purely hypothetical or virtual entities.

To summarize, Places are symbolic structures consisting of a phonological pole (a location in space) and a semantic pole (the most schematic meaning of Place is "thing"). Places fall along a continuum starting from a baseline of real objects in the spatiotemporal surroundings, the conceptual archetype of Place. Various cognitive processes operate to yield elaborations of this baseline situation. The entity with which a Place is associated may disappear, requiring memory. The entity may be present but not within the signer's or addressee's perceptual field, such as a Place associated with the spatial location of a distant house. Entities and the Places associated with them may be real but imagined, as in the teacher example, or they may be abstract, such as two theories located in signing space for purposes of comparison. All of these elaborations beyond the baseline of a real, physically present object require additional conceptual resources.

The entity with which a Place is associated may attenuate completely. All that is left is the Place (which was always immanently present). In this case, the meaning has become almost entirely schematic because the Place is not associated with any actual entity until it is used in an utterance. Its phonological location is largely schematic as well. Both the meaning and the location – the semantic pole and the phonological pole – of the Place are specified by the grammar of the language (although certainly contextual and pragmatic influences still may remain, e.g., focused referents may appear on the signer's dominant side); the Place is fully instantiated semantically and phonologically in a usage event. Our claim is that this continuum captures both deictic and anaphor systems of reference, and that Place symbolic structures span the entire continuum.

Finally, we note that in our usage-based view, phonology is not a static list of *a priori* elements (in this case specified locations), but is instead dynamic, developmental, and emergent. As users visually track, point to, and direct conceptual attention to some physical entity, they build up a symbolic structure that becomes increasingly schematic the more they direct attention to different entities: The specific entity generalizes and attenuates to "thing," and the location of the entity attenuates to "location." That symbolic structure is a Place. The Place symbolic structure can now be recruited in more abstract uses, such as marking remembered or imagined entities, person reference, demonstratives, proxy-antecedents, and nominal components of directional verbs.

Last, our cappuccino-drinking friend reveals another significant aspect of Place structures. Although she directed attention to the Place of the cappuccino cup, the cup was not the ultimate target of attention. She had a motive for directing attention to the cup: in the first instance, to make a comment on its taste. Pointing to the Place established a topic, and her spoken utterance constituted a comment. This is a reference point construction. We see this function of Places in many of the following examples.

An example [B] of pointing to a Place associated with a physical object in the environment occurs in a video produced to introduce children to animals in the zoo. The signer, Eliana, is on location in the zoo explaining about the Tortuga Gigante "Giant Turtle." Typically, a signer would select an area on the dominant signing side, in this case, the signer's right side, to introduce the main topic of a discourse. Here, the signer is standing in front of the area in which the turtle is located, but it is on her left. She orients her body to the left, points to the left, and signs TURTLE GIANT TO-BE-CALLED and then again points to the left (**Figure 3**). Thus, the high perceptual accessibility of the actual spatial location of the turtle's Place on the left of her signing space motivates her choice to establish the main topic of her discourse (the giant turtle) on her non-dominant side. This Place is maintained throughout her discourse.

In another example [C], Pablo and Alejandro, members of the Movimiento Argentino de Sordos (MAS), have organized a demonstration to support a bill on the national recognition of Argentine Sign Language. They also prepared a video to describe strategies for explaining the linguistic problems of the deaf community in Argentina to hearing people unacquainted with the issues. Alejandro says that demonstrators should not talk to the press; instead, they should let the leaders communicate with the press, not because they don't want demonstrators to

FIGURE 3 | Place associated with physically present animate entity.

express their ideas, but because they have strategies that will make an impact on the people. Then, he says that Pablo will give an example. In the fragment we analyze, Pablo introduces the problem (hearing people have a medical, not a cultural, view of deaf people). He then points (thumb-point) to the building of the National Congress, directing attention to it as a Place (**Figure 4**).

Places and Reference Points

When signers point to turtles or buildings to direct the interlocutor's attention to an entity's Place, they do so for the purpose of establishing mental contact with another entity.

FIGURE 4 | Place associated with physically present non-animate entity.

They are pointing to Places in order to create conceptual reference points.

In the giant turtle example, pointing to the turtle serves as a deictic strategy, along with the descriptive strategy of naming the entity, to produce a grounded nominal (Martínez and Wilcox, 2019). The signer then goes on to describe various characteristics of giant turtles: They can live more than 100 years, they have hard shells and scales on their legs, etc. The semantic pole of the Place, the turtle, serves as a reference point, and its dominion is the conceptual region to which it provides access: in this case, to the set of characteristics of giant turtles.

The discourse that continues in the Alejandro and Pablo example also reveals the use of Place to create reference point constructions. Pablo points to the building not for the purpose of directing attention to the building itself, but to establish a reference point, which he then uses to continue the discourse by talking about hearing legislators and their views of deaf people. Pablo directs attention to the building, establishing its Place as a reference point for the purpose of affording mental access to a target. In this case, Pablo ultimately intends to direct the interlocutor's conceptual attention to the legislators, their views, and legislative activity that takes place in the building. Another way of describing this is that the legislators and their activity is the reference point target. The choice of spatial location (i.e., a phonological location) for establishing a new discourse referent in these examples is not randomly selected. The signers use perceptually accessible entities in the current physical environment (i.e., the ground) – the Places of the turtle and the building of the National Congress.

Places and Placing

In addition to using Places as simple reference points, signers also incorporate Places as components in complex placing constructions. Continuing his narrative [C], Pablo explains that, because of their ideology, hearing people regard deaf people as mentally challenged, not equal to hearing people, mute, incapable.

To express this Pablo produces the sign PERSON, but rather than articulating it at an unmarked location in signing space, he signs it in the spatial location of Alejandro. We analyze such constructions as *placing* (Martínez and Wilcox, 2019). In placing, a sign is produced at a specific meaningful location in space. We identify two types of placing: create-placing, in which a new meaningful location, a Place, is created, and recruit-placing, in which the signer produces a sign in an existing Place. In this case, Pablo recruits a Place associated with Alejandro[5].

This placing construction is a component in a larger, simultaneous construction (**Figure 5**). While Pablo continues to hold the placed sign PERSON with his non-dominant hand, he signs DEAF TO-SEE DEAF with his dominant hand. Because DEAF is a body anchored sign, unlike PERSON and TO-SEE, it cannot be placed. The verb TO-SEE is produced with a path

[5]Martínez and Wilcox (2019) analyzed this as a create-placing construction. Because Pablo uses a Place, our analysis is that it is recruit-placing.

FIGURE 5 | Placing construction as a component of a simultaneous construction.

movement moving from Pablo toward Alejandro. Pablo then lists the negative characteristics (mentally challenged, incapable, etc.).

For this entire discourse segment, Pablo is the conceptualizer. TO-SEE in the sense used here is not a perceptual verb; it is, rather, a verb of cognitive activity. TO-SEE means "to regard as" or "to think of," and in this use, it expresses the cognitive activity of categorizing: hearing people categorize deaf people as those who are incapable, etc.

In the previous discourse segment, Pablo placed PERSON on his own body as a rhetorical device so as to frame hearing people, such as the legislators in the legislative building to which he has just pointed, as neutral addressees, lessening the tensions between hearing and deaf people – in effect saying "people in general" (Martínez and Wilcox, 2019). Because of the previous placing construction, Pablo is a type – hearing people. The conceptualizer of TO-SEE is "hearing people" – a virtual conceptualizer. Although TO-SEE is directed at Alejandro, he is not the object of categorization. In order to understand who is the object of categorization, we have to unpack two further constructions.

Looking at only the semantic poles, we see that, in the first construction, the semantic pole of the lexical sign PERSON is a type specification. When it is placed, it integrates with the semantic pole of Alejandro's Place. However, the semantic pole of this Place is not Alejandro as an undifferentiated whole. Rather, Alejandro's Place serves as a reference point, affording mental access to a dominion of targets, in this case of characteristics associated with Alejandro. We do not yet know which of those target characteristics are relevant. With Alejandro as a referent point, the targets could be Alejandro's gender, his hair color, his clothing, or any number of other characteristics. Which is the selected target? In the discourse scene we are describing, the most salient target is Alejandro's deafness. We can confirm this because it is also the characteristic explicitly mentioned when Pablo signs DEAF TO-SEE DEAF. The first construction, thus, integrates the component type specification PERSON with Alejandro's Place, specifically the target "deaf" of the Place reference point, to create the composite construction "deaf people."

This composite construction is then a component in the higher-level construction that integrates "deaf people" with TO-SEE. TO-SEE is a cognitive activity verb with two schematic semantic elements: the categorizer and the object of categorization. These two elements are the semantic poles of two Place structures[6]. The first schematic Place, the categorizer, is elaborated in the prior discourse frame when Pablo uses the placing construction to present himself as a hearing person; it is hearing people who are doing the categorization. The second schematic Place, the object of categorization, is elaborated by the composite construction "deaf people," producing the complex construction "hearing people see deaf people" as incapable, etc. **Figure 6** depicts the semantic side of these constructions. Dotted lines indicate correspondence or conceptual overlap; filled lines with arrows indicate elaboration.

As a result, in this composite construction, we have virtual hearing people categorizing virtual deaf people – both of which are represented by real people (Pablo and Alejandro) in the discourse ground. The virtual deaf people are evoked by Alejandro's Place when integrated with the type specification PERSON. The conceptualizer of TO-SEE is also a Place (Pablo's), which has been semantically extended through the placing construction in the previous discourse frame to create virtual hearing people. Thus, Places associated with physical entities in the ground (Pablo as signer, Alejandro as one of the interlocutors) play essential roles in the component structures that go into forming this complex construction.

Looking at the phonological side of these constructions, we see comparable complexity. The relevant issue is the phonological poles of the various Place structures, which is their locations. The phonological location of PERSON is schematic, which is what permits it to be placed. When PERSON is placed, the phonological location of Alejandro's Place elaborates its schematic phonological location. The two schematic elements of TO-SEE are Places. Pablo's prior placing construction elaborates the phonological location of the first schematic Place, the

[6]In a fuller treatment. we would analyze TO-SEE as an agreement verb. There is widespread disagreement among sign linguists about agreement. We adopt a cognitive-functional analysis of agreement (Barlow, 1999; Croft, 2013). Specifically, we treat agreement as multiple symbolization, a special case of conceptual overlap characteristic of all grammatical constructions (Langacker, 2009a). In signed languages, the conceptual overlap marking agreement is achieved via Places.

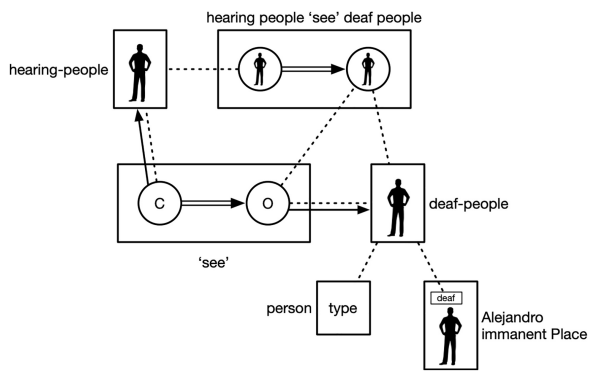


FIGURE 6 | Complex simultaneous construction.

categorizer, with the phonological location of his Place. The schematic phonological location of the second Place, the object of categorization, is elaborated by the phonological location of Alejandro's Place.

Another way to view this complex construction is in terms of conceptual overlap. The semantic pole of Alejandro's Place conceptually overlaps with the semantic pole of the placed sign PERSON as does the semantic pole of the object of TO-SEE: all three map to the same entity in conceptual space, deaf people. This *conceptual* mapping or overlap is achieved by *phonological* overlap: The phonological pole of all three structures are articulated at the same location in space in the discourse ground, the phonological locations of the Place structures.

Finally, this analysis reveals a complex level of grammatical iconicity grounded in conceptual archetypes. Participants (hearing people, deaf people) in an interactional setting are phonologically represented by the locations they occupy in Place symbolic structures. The subjective cognizing activity on the part of the categorizer (depicted by the double-line arrow in **Figure 6**) is phonologically represented as a path movement from the categorizer (hearing people) to the object of categorization (deaf people).

PLACES AND REFERENCE POINTS IN DISCOURSE

We have shown that pointing constructions can incorporate Place referents in the physical environment. These deictic pointing constructions integrate with grammatical constructions and reference point constructions to create extended, cohesive discourse.

The Life of Quinquela

The next examples are taken from a video describing the life of Benito Quinquela Martín (1890–1977), an Argentine painter born in La Boca, Buenos Aires. The signer, Mercedes, is standing in front of a photo of the orphanage where Quinquela spent his early years. In this discourse excerpt [D], the photo of the orphanage, which is behind and on the left of the signer, is a recruitable Place. The signer uses a placing construction with PERSON in proximity of the photo and its Place and several pointing constructions using this Place as a component (**Figure 7**).

She then points to this Place, directing attention to it as a reference point. The target of the reference point, the reason why she points to the photo, is the situation of Quinquela's life during this period, including the fact that his parents abandoned him at a young age. The signs expressing the target occur in this left Place. Placing a sign in the same Place as a previous reference point is the grammatical mechanism by which targets are identified and associated with their reference point (Martínez and Wilcox, 2019).

The signer continues her narrative, explaining that, at the age of six, Quinquela was adopted, still using the left Place for this phase of his life. She then introduces a new phase. For this new period, the signer reorients her body to the right and uses the right signing space for this portion of her narrative[7].

She then signs OTHER, produced with an index finger and an arc movement toward the right. Although OTHER is a lexical sign, it also functions as a pointing device in a pointing construction, which creates and directs attention to a new Place with a phonological location on the signer's right. She then places the sign THEY-TWO in this right Place. This Place is also recruited as a reference point to introduce a series of targets, aspects of Quinquela's life with his adopted family. An initial pointing construction and all of the non-body-anchored signs used in this portion of the narrative are placed in the right Place: The phonological pole is the right signing space, and the semantic pole is this phase of Quinquela's life.

The signer introduces two Places in this excerpt. The first, on her left, is recruited from the Place of the photo, and the second, on her right, is a new discourse Place created by pointing and placing constructions. These Places are two components in a sequential-events construction. This construction is based on timelines commonly observed in signed languages in which time is metaphorically represented as movements among locations in space (Engberg-Pedersen, 1993; Winston, 1995; Nilsson, 2016).

[7]Winston (1993) suggests that orienting the body is a type of pointing; this is compatible with our analysis, in which the body is a pointing device used to direct attention to a Place (Wilcox and Occhino, 2016).

FIGURE 7 | Placing and pointing constructions in discourse construction.

In this case, the construction is used to describe a sequence of events comprising the two phases of Quinquela's early life. Thus, in this example, a pointing construction that incorporates an entity in the physical environment via a point to its Place integrates with a conventionalized grammatical construction to create a coherent discourse structure.

The Order of the Screw

Our last example [E] from the life of Quinquela comes from a portion of the narrative in which the signer describes Quinquela's *Orden del Tornillo* (Order of the Screw). In 1948, Quinquela created this Order with a playful name for men and women (mostly artists) who, in the eye of common people, live in a state of madness. All the people who were to become members of the Order received a screw with a warning: "This screw will not make you sane. On the contrary, it will prevent you from losing this luminous madness of which you feel so proud."

The setting has the signer standing near a poster describing the history of the Order of the Screw and showing Quinquela in his Order regalia, consisting of robes and a hat. The signer explains that Quinquela created this group and gave each member in the group a screw, which was the symbol indicating that they were now members of the Order of the Screw. She signs GROUP, a two-handed sign (**Figure 8**). While she holds her non-dominant hand in the GROUP sign, she then signs a circular point with her index finger. She then signs GIVE, a distributive verb indicating that Quinquela gave each member individually a screw.

The signer uses a simultaneous construction as we saw in the example of Pablo and Alejandro. The sign GROUP is produced near the poster, placing and recruiting the poster's Place. The poster evokes the semantic frame of the Order of Screw, which serves as a reference point. The reference point could be used to evoke any number of targets, such as Quinquela, the regalia worn by Quinquela and the members, or the physical screw. By placing GROUP, the signer evokes one aspect of the Place's semantic pole, the members of the Order construed collectively as a group. This collective construal is reinforced by the circular index pointing construction directing attention to this collective plural. In other words, GROUP directs attention to or profiles a collective entity; any substructure of that entity may be conceptually present in the immediate scope but unprofiled.

GROUP now serves as a new reference point, providing mental access to the giving event. By signing GIVE as a distributive verb, the signer changes the profile from the collective construal to one profiling the internal substructure of the group, its individual members. This profile shift has the effect of conceptual zooming from focusing attention on the collective to focusing attention on the individuals (**Figure 9**; Langacker, 1997).

As we saw for the sign TO-SEE in the Pablo and Alejandro example, GIVE is an agreement verb in which the final location of the verb's path movement is a Place. The semantic pole of this Place conceptually overlaps with the individual members of the group[8].

San Martín and Argentine Independence

We have seen that Places associated with entities in the physical environment can be used as components in grammatical constructions. The methods of directing attention to Places include pointing, placing of signs, and body orientation. Often, the signer directs the interlocutor's attention to a Place in order to use it as referent point, a conceptual stepping stone so to speak, affording access to a target, which is the intended focus of attention.

In this section, we examine Places that are created and used in discourse. We see that Places are created by grammatical constructions, such as pointing, placing, body orientation, and agreement verbs. Places serve as reference points for introducing and tracking referents in ongoing discourse. We also see that discourse Places are often "repurposed" by a series of reference point chains in which a reference point target serves as a new reference point in subsequent discourse.

[8]In this construction we see what would be, according to a formal approach, an agreement discord: GROUP is singular but GIVE is distributive, expressing the event of giving screws to multiple individuals within the group and not to the group. Following Barlow's (1999) discourse-based treatment of agreement, we would say such "feature mismatches" suggest that agreement is providing information about the nature of the discourse referents rather than information about the morphosyntax of the controller. Thus, this example could be explained in terms of an association of two compatible properties ("collective" and "composed of individuals") between the singular noun GROUP (the primary discourse referent) and the inflection of the verb GIVE (the secondary discourse referent). A fuller treatment of how a cognitive-discourse approach to agreement integrates with our Place analysis is beyond the scope of this paper.

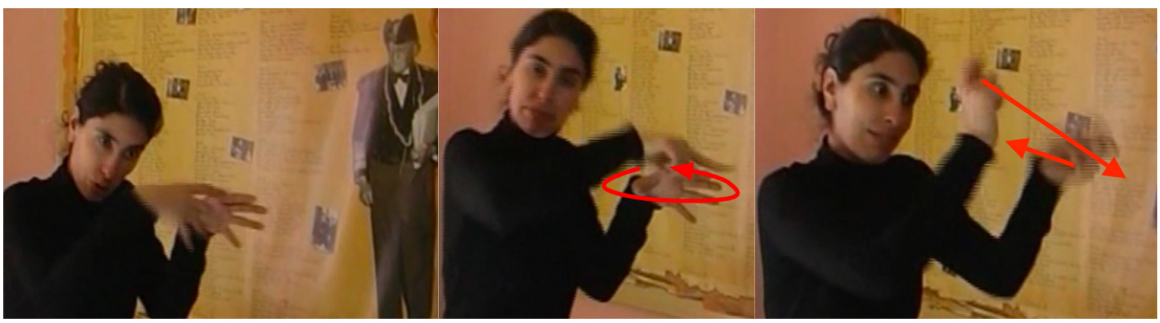


FIGURE 8 | Placing and pointing constructions in simultaneous construction.

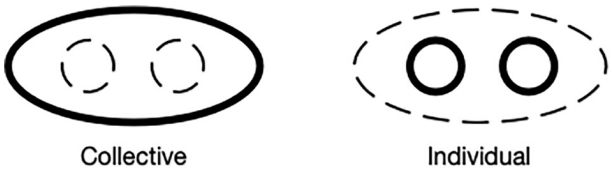


FIGURE 9 | Collective vs. individual profile.

The next examples come from a narrative signed by Diego Morales about the famous hero of the Argentine independence, José de San Martín (1778–1850). In the first [F], the narrator introduces Argentina with a proximal (downward) pointing construction (Martínez and Wilcox, 2019). He explains that San Martín lived in a small town called Yapeyú in the province of Corrientes, Argentina. San Martín had no opportunity to study and progress there, only to harvest the land or serve in the military. Two signs, PROGRESS (**Figure 10**) and HARVEST, are Placed in the left, creating an Argentina Place. In signing PROGRESS, the narrator also orients his body toward the Argentina Place.

He then comments that, a long time ago, Argentina was conquered by Spain. Spain is introduced with SPAIN, a body-anchored sign, and a pointing construction to the upper right signing space (**Figure 11**), creating a Spain Place: in this construction, the schematic semantic pole of this Place conceptually overlaps with Spain. He then signs the verb COME and CONQUER starting from the Spain Place and ending in the Argentina Place: again we see conceptual overlap with the semantic poles of the Spain place and the agent of COME and CONQUER and with the Argentina Place and the goal and patient, respectively, of these two agreement verbs.

FIGURE 10 | Argentina Place created by placing constructions.

We learn that San Martín had been living in Spain, where he became a successful military leader [G]. He returned to South America where he joined forces with another revolutionary, Manuel Belgrano. Two groups in Argentina had been battling, one opposed to the King of Spain, led by Belgrano, and another in support of the King. The narrator now "repurposes" the Spain Place and the Argentina Place. He does this by using the Spain Place as a reference point to refer to the King of Spain (the target); the Argentina Place is used to refer to two opposing groups in Argentina. Both of these reference point constructions express metonymic relations: the entity being referred to (the reference point Spain and Argentina Places) affords mental access to the intended target (King of Spain and two opposing groups).

In addition to identifying the two opposing groups in Argentina as a unitary discourse element, the narrator creates two new subordinate Places, beginning with the signs BATTLE (**Figure 12**) and AREA: "In Argentina, two sides had been battling." BATTLE is a two-handed sign. Each hand creates a new Place, dividing the previous left-hand Argentina Place into two new Places, one on the far left and the other center left. The narrative effect is to create two new discourse Places, both located in Argentina and each associated with a group engaged in battle.

San Martín discovers that one group, indicated by pointing to the left center Place, is the monarchists. The monarchists respect (give allegiance to) the King: RESPECT is signed toward the upper right Place formerly associated with the Spanish King.

The other is a group of revolutionaries opposed to the King (far left Place). They want to remove (KICK-OUT) the King. KICK-OUT is an agreement verb: the semantic agent, those who want to remove the King, is instantiated by the semantic pole of the revolutionaries Place (far left), and the semantic patient is the King, instantiated in the King Place (right and upper right). KICK-OUT moves from far left to upper right (**Figure 13**).

FIGURE 11 | Spain Place created by recruiting the Spain Place.

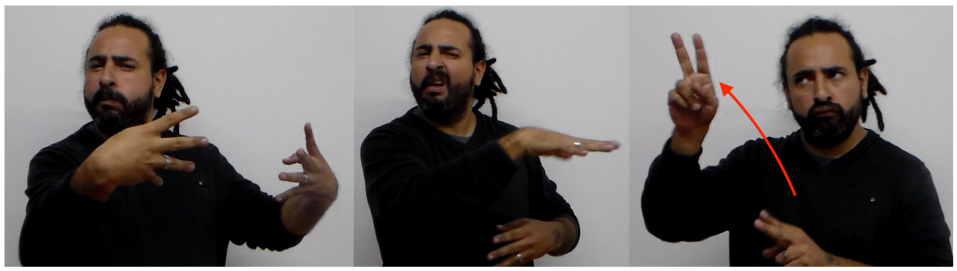


FIGURE 12 | Argentina Place and Spain Place repurposed.

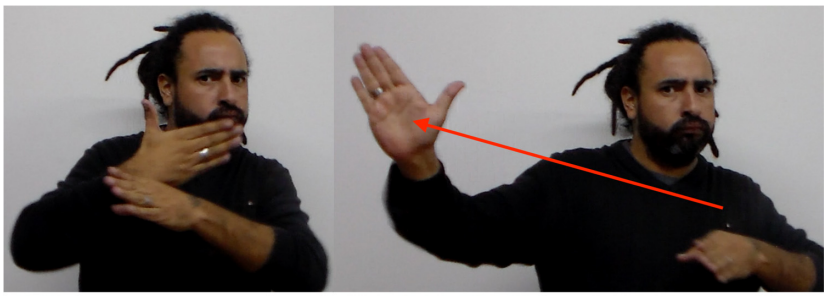


FIGURE 13 | Agreement verb incorporating two Places.

The narrator then explains how San Martín came to meet Belgrano. First, he points to the left, reactivating the revolutionary Place, and indicates San Martín's surprise at what he is about to learn with a facial display. He then signs BATTLE (**Figure 14**) again, but only with the left hand; simultaneously, his right index finger points to the hand signing BATTLE. It is among this group that San Martín finds Belgrano and realizes that they share revolutionary views. This discovery is expressed with a complex pointing and reference point construction. The right index point is the pointing device directing attention to the left hand, which has been recruit-placed to correspond to the left-hand Place: the left hand of BATTLE now conceptually overlaps with the revolutionary group. The narrator then uses this Place as a reference point, evoking as the target a new referent, Belgrano.

The reference point constructions in this small section of discourse are used creatively to repurpose the left Place with new referents, each conceptually connected with the previous one. As we saw in the Order of the Screw example, the effect is a zoom-in strategy by metonymic association:

> Argentina > two battling sides in Argentina (revolutionaries vs. monarchists) > the revolutionaries > Belgrano.

FIGURE 14 | Reactivation of revolutionary Place.

DISCUSSION

As Engberg-Pedersen (1993) has so cogently observed, space is semantically loaded. We wish to point out that space is also phonologically loaded. Conceptual archetypes are the experiential basis of the semantic pole of linguistic items and constructions. For signed languages, conceptual archetypes are also the experiential basis of the phonological pole of linguistic items and constructions. In other words, the conceptualization of space is manifest semantically and phonologically in signed languages. This is especially important for Places because location in space is the phonological pole of Places.

In this section, we examine the semantic and phonological implications of subjectification on Places. Our analysis of Places has implications for the development of infant pointing. We also suggest that the various grammatical functions of Places mirror a pattern in the development of demonstratives described by Diessel (2006).

One aspect of the conceptualization of Places is the nature of their construal – from more objectively to more subjectively construed. These different construals form the basis of the continuum of Places described in section "The Continuum of Places." We suggest that Places fall along a continuum of *subjectification*. As used in CG, subjectification concerns the asymmetry between objective and subjective construal. An expression or scene is objectively construed to the extent that it goes "on stage" as an explicit, focused object of conception. An element is subjectively construed to the extent that it remains "off stage" as an implicit, subject of conception. According to Langacker (2006, p. 21), subjectification can be thought of as a kind of semantic attenuation or "fading away": in subjectification, a subjectively construed entity remains as a vestige of an objectively construed counterpart that was actually there all along, immanent in the latter.

An example of subjectification in adjective use is given in Athanasiadou (2006, p. 217):

(1) a. *The complete works of Shakespeare.*
b. *He is a complete stranger to me.*

As Athanasiadou explains, in (1a), *complete* describes a spatial configuration rather than a property. It expresses an objective configuration. In (1b), the meaning shifts to a different type of quantification with a subjective construal. Here, the meaning of *complete* resides in the conceptualizing activity of the speaker. This conceptualizing activity is present in (1a), immanent in the conceptualization of the objectively construed spatial scene, but in (1b) it is used in abstraction from any objective spatial configuration (Athanasiadou, 2006).

In these examples, objective vs. subjective construal has pertained only to the semantic pole. Because signed languages are produced in visible space, the phonological structure of signed languages affords a second "layer" of construal and subjectification. It is this second layer that is central to the different construal of Places. In section "The Continuum of Places," we identified attenuation as a factor that underlies the continuum of Places. One dimension of attenuation is the degree to which elements are objectively or subjectively construed with elaborate conceptual content inviting more objective construal.

Consider the following example (from Langacker, 2008), in which the arrow indicates a pointing construction:

(2) *I want this [→] one*

As Langacker (2008, p. 468) observes: "In addition to its signaling role, this gesture is part of the situation being described. The sentence describes a relationship in which the speaker goes onstage as a focused participant. Part of this onstage situation is the very fact that the speaker is pointing at something, and the object is specifically identified as what the speaker is pointing at." This is the baseline Place associated with a physical object in the environment. Because of the elaborate conceptual content of the actual physical object, this Place is objectively construed.

An example in our data of more objective construal occurs when Pablo points to the legislative building. He directs mental attention but also (potentially) visual attention to the physical building in the current spatial environment – to the Place of the building. While he uses that Place as a reference point to direct mental scanning along a path leading to the target, the physical presence of the building invites an objective construal.

The functional difference between pointing and placing also figures in objective vs. subjective construal. Pointing directs attention, both perceptual and conceptual. When the signer points to the photograph of Quinquela in the Order of the Screw example, she directs attention to the photograph and its objectively construed Place. Placing, however, attracts attention rather than directing it (Martínez and Wilcox, 2019). This lowered focus of attention to the physical scene also lowers the objective construal. When she signs GROUP, the physical location of the photograph is no longer salient; the placing construction, which serves as a reference point, leaves behind only the mental operation of scanning to locate the intended target.

Places such as those used by Diego in the San Martín narrative show further subjectification. When Diego uses an upward location for Spain and a download location for Argentina, he puts onstage elements of the ground and general knowledge about the world (he is signing in Argentina, and recruits knowledge about maps and where the countries are), thus retaining a vestige of objective construal. The use of Places with directional verbs, for example, in the Pablo TO-SEE construction ("hearing people see deaf people as incapable") and the distributive "giving" construction in the Order of the Screw example are more highly subjectified Places, used for the mental operation of locating referents in conceptual space. Finally, when Places are used in constructions such as the proxy-antecedent, as in the example of Pablo describing his new teacher, the physical location of the Place has no significance in terms of the physical environment other than to allow the signer focus the interlocutor's mental attention (through the use of the proxy-antecedent Place) on the antecedent referent.

We have now arrived at the conceptual doorstep of creating and using Places in grammatical constructions: point to a location in space and map that location conceptually to some discourse element. Point or otherwise direct attention to that location in space (the Place) later in discourse and it becomes a component in a grammatical construction: anaphor, third person pronoun, agreement marker. In other words, when different instances of Places are used, an abstracted schema, a Place, emerges.

When a signer uses a pointing or placing construction to create a Place and then points to or uses that Place in subsequent discourse, the interlocutor is not instructed to direct visual attention to the Place, and in fact, nothing is visible in the Place. The interlocutor is only instructed to direct mental attention to the semantic structure of the Place. This then is one manifestation of the "complete disappearance of an objectively construed entity with retention of mental operations immanent in its conception" (Langacker, 2006, p. 29).

The same cognitive processes of subjectification and reference points are also implicated in the development of infant pointing. The pioneering work on deictic pointing in infants proposes two functions: protodeclarative and protoimperative (Bates et al., 1975). Summarizing and expanding on this research, Tomasello

et al. (2007) observe that, in protoimperatives, infants point to objects they want or to request an action involving that object. Protodeclaratives not only are used for directing attention to something, but also for many different reasons, including remembering non-present events. In all cases, the infant invites the recipient to attend to the referent for a reason. To understand why the infant is pointing, the adult must understand both *what* the infant is directing attention to and *why* the infant is directing attention to it (the motive) (Tomasello et al., 2007)[9]. We claim that, in these pointing constructions, the *what* is a Place serving as a reference point, and the *why*, the motive, is one of the many potential targets in the reference point's dominion.

Reference points serve as the basis for the analysis of topic constructions (Langacker, 2008). Infant pointing exhibits the same structure: "Pointing serves to establish a new topic, about which further things may then be communicated" (Tomasello et al., 2007, p. 719). In fact, we would claim that, in these cases, the reference point structure of pointing manifests a related type of conceptual archetype, searching and finding (Langacker, 2006). In order to understand a pointing construction, the interlocutor must search and find the motive for the point. In spatial searching and finding, a *search domain* is the spatial region in which the searched for entity is located. Prepositions, for example, reflect the search and find conceptual archetype. In locative expressions, such as *under the table*, the search domain is the spatial region to which a locative expression confines its subject.

Our claim is that, in pointing constructions, the search domain is the reference point's dominion, the region in which the interlocutor must search for the motive for directing attention to the reference point. The person pointing is trying to do something. Why is the infant directing the adult's attention to a Place associated with some object? In signed language Place constructions, why is the signer directing attention to an anaphor Place? In some instances, the answer is provided in accompanying language. In others, the interlocutor must discover the answer: the infant wants the object, or the signer is directing the addressee to search for the antecedent (by way of the proxy-antecedent).

Finally, we note that the function of Places mirrors the pattern described by Diessel (2006) in the development of demonstratives. Diessel observed that, in exophoric use, demonstratives focus the interlocutor's attention on concrete entities in the physical world. When used in discourse, demonstratives focus attention on linguistic elements in the surrounding discourse context. These represent the two ends of our Place continuum. Further, Diessel notes that the communicative function of demonstratives extends from the physical world to discourse: "Demonstratives are not only used with reference to concrete entities in the surrounding situation, they may also refer to linguistic elements in the ongoing discourse" (Diessel, 2006, p. 481). Diessel claims that in both cases, the same psychological mechanisms are at work. We see this as the same conceptual underpinnings that unite the continuum of Places.

Diessel suggests a developmental path of demonstratives into grammatical markers:

(3) deictic DEM > anaphoric DEM > 3.PRO > pronominal clitic > agreement marker > Ø

We note that the functions Diessel has documented are much the same as those we have described for Places. Symbolic Place structures function as demonstratives (both exophoric and discourse), as anaphoric pronouns, as non-first person (but also first person) pronouns, and as agreement markers.

CONCLUSION

We have examined the conceptualization of space in signed language discourse within the theory of CG. Symbolic structures are basic explanatory concepts in CG; lexicon and grammar form a gradation consisting solely in assemblies of symbolic structures varying in degree of complexity and schematicity. We have proposed that Places are basic elements of signed language structure, defining Place as a symbolic structure that associates a semantic pole ("thing") and a phonological pole (location). Places acquire full contextual meaning and a specific spatial location in a usage event.

We suggest that our account of Places reveals new aspects of how space is semantically and phonologically conceptualized. Places provide a unified and natural account of signed language data that is often compartmentalized into separate cognitive systems. As we have seen, some sign linguists argue that pointing to or incorporating locations in the physical environment lies outside of language altogether and must be treated as part of a gesture system. We see no need to segregate the conceptualization of locations in space into distinct cognitive domains. Our primary claim is that Places unify deixis and anaphor. Rather than representing two distinct domains of reference, we suggest that they are ends of a symbolic continuum that varies in terms of subjectification.

The various functions of Places are accounted for with nothing more than core concepts of CG such as conceptual archetypes, schematicity, subjectivity, reference point constructions, conceptual overlap, and conceptual elaboration. We have, however, extended the use of these core CG concepts beyond the semantic pole of symbolic structures to the analysis of the phonological pole. For example, we have claimed that an object in a location is the conceptual archetype for Place. We would also suggest that conceptual archetypes are the experiential basis of basic phonological categories: a physical object (hand shape), an object in a location (location), and an object moving through space (movement). Further elaboration of these CG concepts could prove fruitful for the development of a cognitive phonology of signed languages.

[9]Tomasello et al. make the important observation, with which we agree, that we cannot assume infants comprehend and produce their early pointing gestures with full adult-like meaning. They suggest we can assume, however, that infants may operate with some kind of "primordial, undifferentiated communicative intention that contains the basic structure, but not all of the adult details" (Tomasello et al., 2007, p. 715).

CG has shown that reference point phenomena are manifest across a broad range of grammatical and discourse functions, including possessives, topic-comment, metonymy, and pronoun-antecedent relationships. We have shown that Places serve as both perceptual and conceptual reference points with many of the same grammatical functions.

Our analysis makes certain predictions for future research. The conceptual archetype for Places is a physical object in a location. This is also the archetype for noun: a physical object composed of material substance residing primarily in space (Langacker, 1987, 2008). Thus, the schematic meaning of Place is compatible with the schematic meaning of noun. This suggests that Places play a role in linguistic expressions of signed languages that incorporate the schematic conception of thing, such as nominals and verbal constructions. Concerning the latter, further research should be carried out to better understand the role of Places within so-called directional or agreement verbs, which incorporate nominal referents to the verb. CG treats agreement as multiple symbolization, a special case of conceptual overlap characteristic of all grammatical constructions (Langacker, 2009a). Place symbolic structures are, we suggest, the site of conceptual overlap in these directional verb constructions.

We have suggested that the different functions of Places result from increased and subjectification of the phonological and the semantic poles. One way in which subjectification is manifest is the diachronic process of grammaticalization. As we have shown, the patterns of semantic function of Place show notable similarities with the function and grammaticalization pattern report by Diessel (2006) for demonstratives. Although long-term patterns of diachronic change are difficult to study in unwritten languages such as signed languages, these patterns of change are attested at much shorter time-scale such as the verbalization of experience in narratives (Croft, 2010), suggesting a possible method for confirming these patterns.

Finally, we have suggested that conceptual elaborations, such as those that account for the semantic change from more deictic to more anaphoric reference, require increasing conceptual resources such as memory and imagination. This suggests that these more elaborated meanings are acquired later in development.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

RM selected, transcribed, organized the data in LSA, and wrote the **Supplementary Appendix**. SW wrote the first drafts of each section of the manuscript, designed, and edited all the figures. Both authors contributed to the conception and design of the study, thoroughly analyzed, and discussed the data.

FUNDING

The University of New Mexico College of Arts and Sciences supported the publication of this manuscript.

ACKNOWLEDGMENTS

We thank Eliana González, Pablo Lemmo, Alejandro Makotrinsky, Diego Morales, and Mercedes Pandullo and the NGO CEA (Centro de Estudios y Acción para la Comunidad Sorda) for permission to use the source material in this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01406/full#supplementary-material

REFERENCES

Athanasiadou, A. (2006). "Adjectives and subjectivity," in *Subjectification: Various Paths to Subjectivity*, eds B. Cornillie, C. Canakis, and A. Athanasiadou (Berlin: Walter de Gruyter), 209–239.

Barlow, M. (1999). Agreement as a discourse phenomenon. *Folia Linguist.* 33, 187–210.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill Palmer Q. Behav. Dev.* 21, 205–226.

Croft, W. (2010). The origins of grammaticalization in the verbalization of experience. *Linguist. Interdiscipl. J. Lang. Sci.* 48, 1–48. doi: 10.1515/ling.2010.001

Croft, W. (2013). "Agreement as anaphora, anaphora as coreference," in *Languages Across Boundaries: Studies in Memory of Anna Siewierska*, eds D. Bakker and M. Haspelmath (Berlin: De Gruyter), 107–129.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489. doi: 10.1515/COG.2006.015

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and Morphosyntax of the Use of Space in a Visual Language*. Hamburg: SIGNUM-Verlag.

Fauconnier, G. (1985). *Mental Spaces*. Cambridge, MA: MIT Press.

Fauconnier, G. (1997). *Mappings in Thought and Language*. New York, NY: Cambridge University Press.

Fauconnier, G., and Turner, M. (1996). "Blending as a central process of grammar," in *Conceptual Structure, Discourse, and Language*, ed. A. E. Goldberg (Stanford: Center for the Study of Language and Information Publications).

Fenlon, J., Schembri, A., and Cormier, K. (2018). Modification of indicating verbs in British Sign language: a corpus-based study. *Language* 94, 84–118. doi: 10.1353/lan.2018.0002

Langacker, R. W. (1979). Grammar as image. *Linguist. Notes Jolla Jolla Cal.* 6, 87–126.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Volume I, Theoretical Foundations*. Stanford: Stanford University Press.
Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Volume II, Descriptive Application*. Stanford: Stanford University Press.
Langacker, R. W. (1993). Reference-point constructions. *Cogn. Linguist.* 4, 1–38. doi: 10.1515/cogl.1993.4.1.1
Langacker, R. W. (1997). "Generics and habituals," in *On Conditionals Again*, eds A. Athanasiadou and R. Dirven (Amsterdam: John Benjamins), 191–222.
Langacker, R. W. (2006). "Subjectification, grammaticization, and conceptual archetypes," in *Subjectification: Various Paths to Subjectivity*, eds B. Cornillie, C. Canakis, and A. Athanasiadou (Berlin: Mouton de Gruyter), 17–40.
Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
Langacker, R. W. (2009a). Cognitive (Construction) Grammar. *Cogn. Linguist.* 20, 167–176.
Langacker, R. W. (2009b). *Investigations in Cognitive Grammar*, Vol. 42. Berlin: Mouton de Gruyter.
Langacker, R. W. (2016a). Baseline and elaboration. *Cogn. Linguist.* 3, 405–439. doi: 10.1515/cog-2015-0126
Langacker, R. W. (2016b). *Nominal Structure in Cognitive Grammar*. Lubin: Marie-Curie Skłodowska University Press.
Langacker, R. W. (2017). "Evidentiality in cognitive grammar," in *Evidentiality Revisisted*, eds J. I. Marín-Arrese, G. Haßler, and M. Carretero (Amsterdam: John Benjamins), 13–55. doi: 10.1075/pbns.271.02lan
Langacker, R. W. (2019). Levels of Reality. *Languages* 4, 1–20.
Liddell, S. K. (1995). "Real, surrogate, and token space: grammatical consequences in ASL," in *Language, Gesture, and Space*, eds K. Emmorey and J. Riley (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc), 19–41.
Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. New York, NY: Cambridge University Press.
Lillo-Martin, D., and Klima, E. (1990). "Pointing out differences: ASL pronouns in syntactic theory," in *Theoretical Issues in Sign Language Research*, eds S. Fischer and P. Siple (Chicago, IL: University of Chicago Press), 191–210.
Lillo-Martin, D., and Meier, R. P. (2011). On the linguistic status of 'agreement' in sign languages. *Theoret. Linguist.* 37, 95–141.
Martínez, R., and Wilcox, S. (2019). Pointing and placing: nominal grounding in argentine sign language. *Cogn. Linguist.* 30, 85–121. doi: 10.1515/cog-2018-0010
Nilsson, A.-L. (2016). Embodying metaphors: signed language interpreters at work. *Cogn. Linguist.* 27, 35–65. doi: 10.1515/cog-2015-0029
Quer, J. (2011). When agreeing to disagree is not enough: further arguments for the linguistic status of sign language agreement. *Theoret. Linguist.* 37, 189–196.
Stokoe, W. C. (1980). Sign language structure. *Annu. Rev. Anthropol.* 9, 365–470.
Talmy, L. (2018). *The Targeting System of Language*. Boston: MIT Press.
Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x
Wilbur, R. B. (2013). The point of agreement: changing how we think about sign language, gesture, and agreement. *Sign Lang. Linguist.* 16, 221–258. doi: 10.1075/sll.16.2.05wil
Wilcox, S. (2014). Moving beyond structuralism: usage-based signed language linguistics. *Ling. Señ. Int.* 5, 97–127.
Wilcox, S., and Occhino, C. (2016). Constructing signs: place as a symbolic structure in signed languages. *Cogn. Linguist.* 27, 371–404. doi: 10.1515/cog-2016-0003
Winston, B. (1995). "Spatial mapping in comparative discourse frames," in *Language, Gesture, and Space*, eds K. Emmorey and J. Reilly (Hillsdale, NJ: Lawrence Erlbaum), 87–114.
Winston, E. (1993). *Spatial Mapping in Comparative Discourse Frames in an American Sign Language Lecture*. Washington, DC: Georgetown University.
Yamanashi, M.-A. (2015). Aspects of reference point phenomena in natural language. *J. Cogn. Linguist.* 1, 22–43.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 July 2020
doi: 10.3389/fpsyg.2020.01778

Acquisition of Demonstratives in English and Spanish

Patricia González-Peña[1], Martin J. Doherty[1] and Pedro Guijarro-Fuentes[2]*

[1] School of Psychology, Faculty of Social Sciences, University of East Anglia, Norwich, United Kingdom, [2] Departament de Filologia Espanyola, Moderna i Clàssica, Universidad de las Islas Baleares, Palma, Spain

The present work re-evaluates the long-standing claim that demonstratives are among infants' earliest and most common words. Although demonstratives are deictic words important for joint attention, deictic gestures and non-word vocalizations could serve this function in early language development; the role of demonstratives may have been overestimated. Using extensive data from the CHILDES corpora (Study 1, *N* = 66, 265 transcripts) and McArthur-Bates CDI database (Study 2, *N* = 950), the language production of 18- to 24-month-old Spanish- and English-speaking children was analyzed to determine the age and order of acquisition, and frequency of demonstratives. Results indicate that demonstratives do not typically appear before the 50th word and only become frequent from the two-word utterance stage. Corpus data show few differences between Spanish and English, whereas parental report data suggest much later acquisition for demonstratives in English. These findings expand our knowledge of the foundations of deictic communication, and of the methodological challenges of assessing early production of function words.

Keywords: English, Spanish, deixis, spatial demonstratives, language acquisition, corpus linguistics, CDI

OPEN ACCESS

Edited by:
Holger Diessel, Friedrich Schiller University Jena, Germany

Reviewed by:
Marianne Gullberg, Lund University, Sweden
Maria Reile, University of Tartu, Estonia

****Correspondence:***
Martin J. Doherty
Martin.Doherty@uea.ac.uk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 February 2020*
Accepted: *29 June 2020*
Published: *24 July 2020*

Citation:
González-Peña P, Doherty MJ and Guijarro-Fuentes P (2020) Acquisition of Demonstratives in English and Spanish. Front. Psychol. 11:1778. doi: 10.3389/fpsyg.2020.01778

ACQUISITION OF DEMONSTRATIVES IN ENGLISH AND SPANISH

Infants communicate about objects and locations in space early in development. By interacting with their caregivers in relation to an object, they are engaging in *deictic communication.* This happens by 12 months, before children have learnt their first words, with the onset of pointing (Tomasello et al., 2007). Pointing is a deictic gesture, and is crucial in language acquisition as it supports word learning and facilitates the transition to two-word utterances (Iverson and Goldin-Meadow, 2005). Demonstrative words (*here*, *there*, *this,* and *that*) are deictic terms. They function to establish joint attention, and often appear in conjunction with pointing (Diessel, 2006; Todisco et al., in press). Given the importance of deictic pointing in language acquisition, it is plausible that demonstratives also have a central role, and therefore would be some of the first and most frequent words of infants - this assumption has been conventional in the literature (Clark, 1978; Clark and Sengul, 1978). Clark claimed that demonstratives are typically acquired among the first 10 words, and always among the first 50. Her claim was based on observational studies with English speaking American children (Nelson, 1973; Braine and Bowerman, 1976) and single-case diaries of other languages. However, no systematic empirical work has addressed this issue.

Given the recent growth of child language databases and the emergence of tools to process them, it now seems appropriated to re-evaluate the claim that demonstratives appear at the start of language development, and are thus foundational to deictic communication and word learning.

Several works on child early speech challenge the claim of an early acquisition of demonstratives. Caselli et al. (1995) described the language acquisition of English and Italian speakers based on parental report with the MacArthur-Bates Communicative Development Inventory (CDI) on over 800 children, and did not find any demonstratives among the 50 words first produced in either language. These data are striking but inconclusive, since the sensitivity of parental report to detect function words in child vocabulary is as yet unclear (Salerni et al., 2007). Rodrigo et al. (2004) observed deictic communication in child-mother dyads. They found deictic words to be rare before the age of two and more frequent afterward, whereas younger infants established joint attention often by using a non-word vocalizations in combination with pointing. In line with this, Capirci et al. (1996) found a small proportion of deictic words in 16- and 20-month-old Italian infants, and a greater proportion of deictic gestures (in combination or not with a content word).

This evidence challenges the idea that demonstratives are essential words in early child speech. It instead suggests that deixis in early stages of language acquisition could rely on gestures, or verbal expressions other than demonstratives.

The aim of this work is to test the claim of an early acquisition of demonstratives to assess the role of these words in language development and deictic communication in infancy. To that aim, we look at child productive speech between 18 and 24 months, which encompasses the typical onset of expressive language and development toward two- or multi-word utterances. We compare demonstrative acquisition in two languages, English and Spanish, chosen because of the differential characteristics of their demonstrative systems (greater syllabic and morphological complexity in Spanish) and because both languages have a large amount of data available as open source for study. Data are obtained from two large repositories of child language acquisition: the CHILDES corpus, comprising transcripts of child spontaneous speech, and the MacArthur-Bates CDI Wordbank, comprising data from parental surveys. A secondary aim is to describe the use of demonstratives in English and Spanish in infant speech and parent-directed speech.

Demonstratives in English are the words *this* and *that* (and their plural forms *these* and *those*) and the locative adverbs *here* and *there*. *This* and *that* can function as pronouns (e.g., "what is *that*?") or determiners (e.g., "*that* book on the right"). Most authors include locative adverbs in the category of demonstratives (Diessel, 2006), although their functions differ slightly; locative adverbs specify a place, whereas determiners and pronouns refer to an object, and are often not used with the aim of disambiguating object position. Spanish demonstratives have three terms instead of two, for proximal, medial and far distance, and vary not only in number but in grammatical gender. See **Table 1** for a full list of Spanish demonstratives. We will compare data from determiners/pronouns with data from the locative adverbs, and ask whether they might have different roles in child speech and be acquired at different times. To preview the results, locatives appear to be acquired earlier, particularly in English, and unlike determiners/pronouns, they do not correlate with language development, measured by mean length of utterance (MLU). Thus, determiner/pronouns and locatives may have different roles.

Sources of Child Speech Data

The CHILDES project is a collection of corpora that feature transcripts of first language acquisition (MacWhinney, 2000). The earliest transcripts date back to the 1973, and it has grown greatly since. The *childesr* package for the statistical software R now allows extracting data from all selected transcripts simultaneously. The MacArthur-Bates CDI (Fenson et al., 2007) is a family of parent inventories that collect data of child expressive and receptive vocabulary and gestures in multiple languages. It has been extensively used as a measure of language development for over 20 years. Since 2017, data are available to use in a structured database called Wordbank, that features data from more than 75000 children[1] (Frank et al., 2016).

As methods for the study of child language acquisition, the analysis of spontaneous speech and parent report have different strengths and potential biases. The advantage of CHILDES data is that they feature naturalistic language production, including parent child-directed speech. However, they do not contain the child's total vocabulary size, and the words in a transcript might be task biased, and not fully representative of child speech in other contexts. The CDI's main strengths are very large sample sizes and that it applies the same items to all children. Abundant studies support the CDI as a reliable and valid measure of child language development (Dale et al., 1989; Feldman et al., 2005) with high predictive validity even several years later (Can et al., 2013). However, CDI data could underestimate function words in children's vocabulary, as opposed to child corpora, where they might be overrepresented (Salerni et al., 2007). Demonstratives are generally studied within the category of function words in the literature in language acquisition, together with words such as articles, prepositions, and conjunctions (Caselli et al., 1995; Salerni et al., 2007). Moreover, it has been suggested that parents from low socioeconomic status background (SES) could be less accurate at reporting their child's vocabulary in inventories. Higher CDI scores have been reported for low SES children relative to high SES children, whereas the literature has consistently reported a disadvantage in language acquisition for children from low SES backgrounds (Reznick, 1990; Fenson et al., 1994). In the case of function words, the demographic differences in parental report might be higher, because these words might be harder to detect (Fenson et al., 1994). Thus, it has been suggested that neither corpus data nor parent report are ideal methods on their own to estimate the frequency of a particular word type in child speech, and using both in combination has been recommended (Pine et al., 1996; Salerni et al., 2007).

To sum up, the principal aim of this work is to study the emergence and frequency of demonstratives in early child speech in order to re-evaluate our knowledge about the function of demonstrative words in early stages of language acquisition. An early acquisition of demonstratives (among the first 10 or 50 words as suggested by Clark) and high frequency would indicate an essential role of this word class for language acquisition

[1] http://wordbank.stanford.edu

TABLE 1 | Demonstrative words in Spanish.

		Proximal		Medial		Distal	
		Det/pronoun	**Locative**	**Det/pronoun**	**Locative**	**Det/pronoun**	**Locative**
Singular	Male	este		ese		aquel	
	Female	esta		esa		aquella	
	Neutral	esto	aquí	eso	ahí	aquello	allí
Plural	Male	estos	acá	esos		aquellos	allá
	Female	estas		esas		aquellas	

Spanish locative adverbs aquí and acá, and allí and allá will be treated as synonymous in our work.

TABLE 2 | Mean length utterance (MLU) and number of word types (number of different words) of the transcripts used, displayed by age and language.

	Spanish				English			
Age (months)	**N of transcripts**	**N of children**	**MLU Mean (*SD*)**	**Word types Mean (*SD*)**	**N of transcripts**	**N of children**	**MLU Mean (*SD*)**	**Word types Mean (*SD*)**
18	2	1	0.97	23	20	19	1.13	40.75
			(0.07)	(15.56)			(0.2)	(27.29)
19	18	5	1.65	43.5	18	18	1.19	78.11
			(0.5)	(17.47)			(0.21)	(47.52)
20	8	3	1.41	132	13	11	1.48	74.08
			(0.18)	(45.68)			(0.36)	(56.24)
21	22	6	1.65	79.77	31	23	1.58	73.39
			(0.41)	(44.43)			(0.38)	(50.47)
22	20	5	1.63	121.5	16	6	1.65	115.13
			(0.36)	(63.63)			(0.26)	(23.06)
23	22	5	1.79	142.45	75	24	1.66	110.71
			(0.39)	(65.91)			(0.41)	(41.53)
Total	92	7	1.64	100.04	173	59	1.54	90.2
			(0.41)	(63.24)			(0.39)	(48.69)

MLU was calculated on the number of words instead of morphemes, because the number of morphemes was not available for all transcripts. Therefore, unintelligible vocalizations (in the transcripts, xxx) were computed as words, and contracted forms (I'm, what's) were computed as one word.

and communication. Contrarily, a later acquisition or marked differences between-languages would support the hypothesis that demonstratives are just one of the possible forms of deixis, and not essential to language acquisition. Specifically, the acquisition of the first demonstrative words will be examined in relation to chronological age, mean length of utterance (MLU, in corpus data) and estimated vocabulary size (CDI data). Study 1 will examine the data from spontaneous speech and Study 2 from parent report.

Additionally, we compare the use of determiners/pronouns with that of locatives. Subtle differences between the two types of term may affect their developmental trajectory. We also compare parent and child use of demonstratives in the same conversation to examine whether parents tend to adopt the demonstratives used by the child regardless of their own perspective.

To preview the results, we find that demonstrative words do not typically appear among the first 50 words, and are more frequent in child's speech toward the age of two years and in two- and multi-word utterances than in the earliest stages of language acquisition. We find cross-linguistic differences, namely late acquisition of demonstratives in English with respect to Spanish. However, these differences are evident only in parental report data. The discussion will cover the implications for deictic communication and methodological considerations regarding the study of function words in child speech.

STUDY 1: CHILDES CORPORA

Study 1 investigates the acquisition and use of demonstrative words using data from spontaneous speech.

Method

Origin of the Data

Data come from monolingual children aged 18 to 24 months from the European Spanish and British English corpora in CHILDES (MacWhinney, 2000). All transcripts that fit these criteria and included an interaction with the mother or father were selected. Seven Spanish corpora (Linaza, Vila, SerraSole, Aguirre, OreaPine, Nieva, and Ornat) and six British English corpora (Forrester, Wells, Manchester, Lara, Howe, and Cruttenden) were included. The British sample comprised 173 transcripts from 59 children, and the Spanish sample 92 transcripts from seven children (see descriptives in **Table 2**).

The number of transcripts per child ranged from one to 39, and they will be analyzed as independent data. Transcripts contained between 9 and 840 target-child utterances (M = 240, SD = 156); t-tests confirmed that there are no significant differences between languages in the number of child utterances by transcript for each of the age groups 18–20 months, 21–22 months, and 23–24 months (all p's > 0.3).

Parent data were obtained in most cases from maternal transcripts, because they were much more frequent than paternal transcripts and generally had more utterances. Paternal transcripts were used when maternal transcripts were not available. In the case of one child of the Spanish corpus (12 transcripts), the father was selected for all instances, because the mother had few utterances and was absent in three of them.

Data Processing and Analysis

Data were extracted and processed in R (R Core Team, 2018) in December 2019 using the R package *childesr* (Braginsky et al., 2019). The number of occurrences of each demonstrative word for parent and child was computed. In Spanish we extracted proximal, medial and distal pronouns/determiners and locative adverbs (*este, ese, aquel*[2] including gender and number inflections and *aquí, ahí* and *allí*, see **Table 1**) and English proximal and distal terms (*this, that, these, those, here* and *there*). In English, demonstratives also have non-deictic uses, such as *there is/are* to indicate existence or in fixed expressions such as *there you go*, and the conjunction *that* (as in *the lady that we met today*). This is not the case for Spanish. We were concerned about the possibility of children using these words non-deictically prior to the acquisition of proper demonstrative use in English. Thus, we checked manually the transcripts of the 10 children from the English corpus who produced only *that* or *there*, which could indicate this non-deictic usage (e.g., in the fixed expression *there you go)*. In all cases we found they apparently functioned as demonstrative words[3].

All statistical analyses were performed on the raw frequencies. Due to differences in sample size between languages and the violation of the normality and homoscedasticity assumptions, non-parametric tests were used: Chi-squared tests (χ^2) were used for dichotomous variables and Mann-Whitney U Tests for continuous variables with Bonferroni adjusted alpha levels for multiple comparisons. The correlational analysis was performed with bootstrap.

[2]An alternative spelling of demonstratives in Spanish, now obsolete, features a written accent on the demonstrative pronouns (éste, ése. . .) to differentiate them from the determiners. *Childesr* word retrieval is sensitive to written accents, and we included both spelling forms in our search. The sensitivity to written accents allowed distinction of the verb form *está* (is) from the proximal, female demonstrative *esta*/*ésta*.

[3]We considered filtering out the non-demonstrative uses of these words using the MOR line of the transcripts, that specifies the word class of each word. However, after analyzing several transcripts, we found this categorization to be unreliable for demonstratives. Nevertheless, although the results might overestimate demonstrative use in English for parents, we do not consider this a serious concern for child data.

Results

First, we describe children's acquisition of demonstrative words with respect to age and MLU, and which demonstrative terms appear in infancy. We then examine whether demonstratives are among children's most frequent words in our sample. Next, we look at the frequency of use of demonstratives per thousand words through development and in comparison with adult use. Finally, we test whether parents and children tend to use the same or opposite demonstrative terms within a conversation. The acquisition of the correct gender and number demonstrative forms as well as the distance contrast conveyed with demonstratives are not within the scope of this work.

Emergence of Demonstratives in Child Speech

We first looked at the percentage of children who used at least one demonstrative word by age and by MLU (see **Figure 1**). A minimum of 60% of children used at least one demonstrative word at any age and MLU point for either language. Over 80% of children used demonstratives from the single word stage (MLU = 1 to 1.5), rising to ceiling at MLU 1.5 to 2.

There were no between-languages differences in the percentage of children who produced at least one demonstrative word: determiners/pronouns, χ^2 (1) = 0.32, p = 0.6; locatives, χ^2 (1) = 1.7, p = 0.2; or any demonstrative, χ^2 (1) = 0.59, p = 0.4. Locatives featured more often in children's vocabulary than determiners/pronouns: in Spanish, χ^2 (1) = 3.96, p = 0.047; and English, χ^2 (1) = 42.76, p < 0.001. In Spanish, this difference was only significant for the youngest age group, 18 to 20 months [χ^2 (1) = 12.40, p < 0.001, Bonferroni adjusted alpha level of 0.017] and at none of the MLU bins. In English it was significant in the two youngest groups [18 to 20 months, χ^2 (1) = 14.25, p < 0.001; 20 to 22 months, χ^2 (1) = 13.85, p < 0.001], and the two lower MLU bins [MLU 1 to 1.5, χ^2 (1) = 20.42, p < 0.001; MLU 1.5 to 2, χ^2 (1) = 9.27, p = 0.002], Bonferroni adjusted alpha levels of 0.017.

Most Common Demonstrative Terms in Child Lexicon

After finding that demonstratives featured in a similar proportion of Spanish and English transcripts, we tested which demonstrative words occurred in each language, irrespective of how frequently they were used. The percentages of children who used each demonstrative term at least once are displayed in **Figure 2**. A greater proportion of Spanish children than British children used proximal terms [*este/aquí, this/here*, χ^2 (1) = 9.5, p = 0.002]. Contrarily, English distal terms *that* and *there* appeared in more transcripts than Spanish medial terms *ese* and *ahí* [χ^2 (1) = 9.78, p = 0.002]. Spanish distal terms *aquel* and *allí* were rare:1% of Spanish transcripts featured the demonstrative *aquel* and 28% the locative *allí*.

Demonstrative Frequency in Child Speech in Relation to Other Words

Corpora transcripts were processed with the *tidytext* R package (Silge and Robinson, 2016) to extract the most frequent words in both languages. For this descriptive analysis, the *stem* transcript line was used. Some transcripts feature only the *gloss* transcript

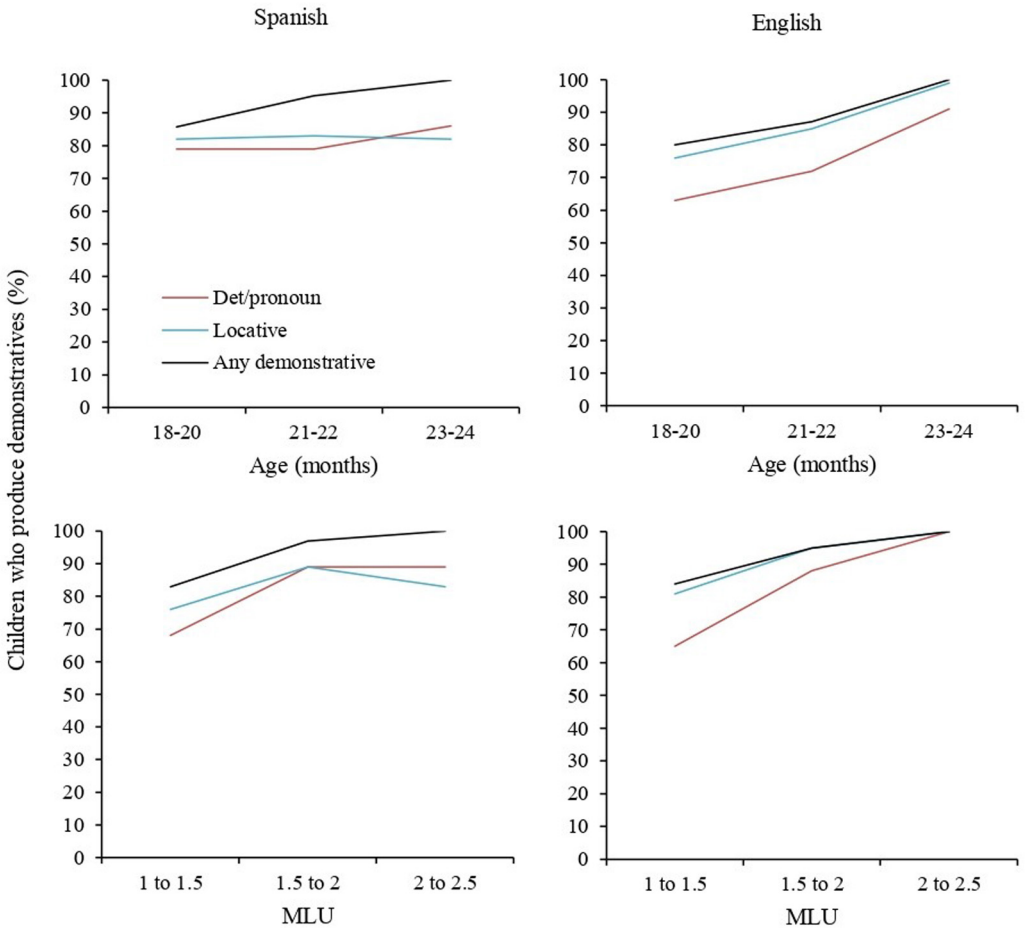


FIGURE 1 | Children who produce at least one demonstrative word in CHILDES corpora, by language, above by Age and below by MLU (%).

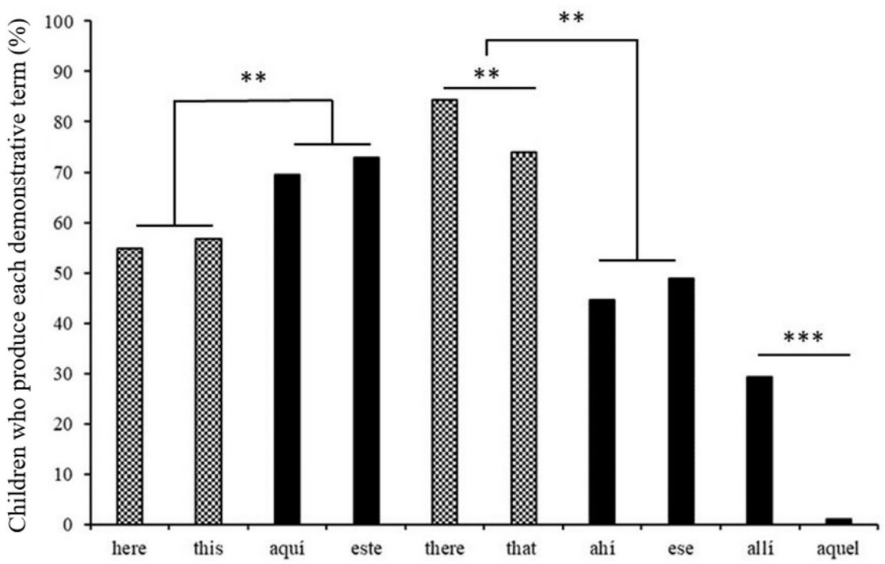


FIGURE 2 | Children who use any demonstrative word in CHILDES corpora, by word (%). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

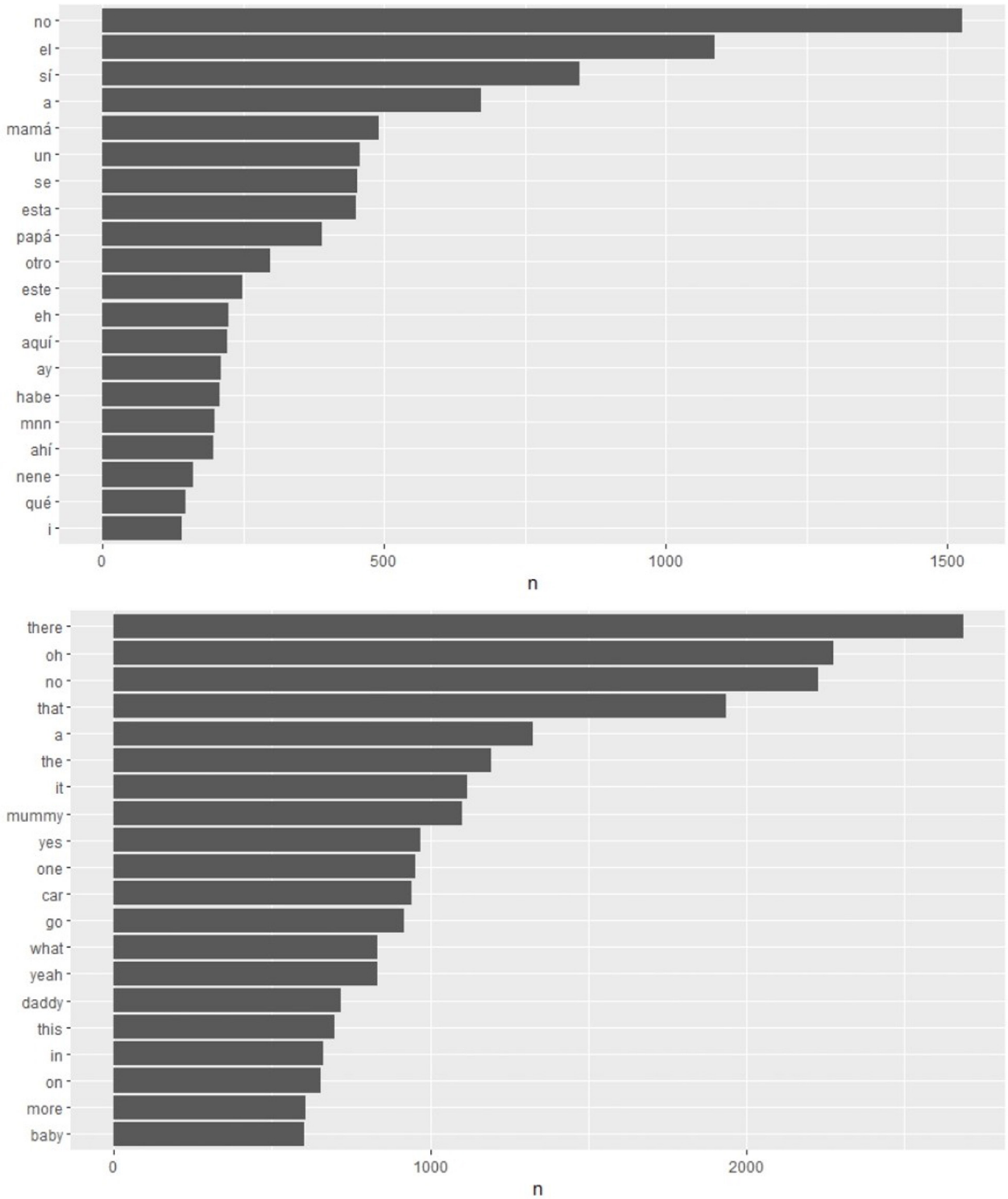


FIGURE 3 | Word frequency of the 20 most frequent words in CHILDES corpora in Spanish, above, and British children, below. Notice in the Spanish plot the 8th word *esta* does not refer to the demonstrative word, but to the root of the verb *estar* (to be).

line. This contains the actual vocalizations of the child, and thus is unsuitable to count frequencies if one wishes to disregard phonetic errors. The *stem* line has the corrected word and the word root in case of verbs. There were 174 transcripts with *stem* line from English children (mean Age = 20 months) and 65 from Spanish children (mean Age = 21 months).

Word frequencies were computed for all words in all scripts for each language. **Figure 3** displays the number of occurrences of the 20 most frequent words for each language. In Spanish, *este* (this), *aquí* (here) and *ahí* (there) were among the 20 most frequent words, in 11th, 13th, and 17th position, respectively. In English, *there, that,* and *this* were among the 20 most frequent words. *There* was the single most frequent word in the corpus, and *that* and *this* occupied 4th and 16th positions, respectively.

Demonstrative Frequency in Child and Parent Speech

The number of demonstratives per thousand words was computed for determiners/pronouns and locatives in both languages and is displayed in **Figure 4**. In child speech, determiners/pronouns were equally frequent in Spanish and English (28 vs. 31 occurrences per thousand words, Mann-Whitney *U* Test, $Z = 1.0$, $p = 0.32$). However, locatives were much more frequent in English than in Spanish in child speech (45 vs. 22 occurrences per thousand words, $Z = 3.7$, $p < 0.001$). In parent speech, both determiners/pronouns and locatives were slightly more frequent in English than in Spanish (determiners/pronouns, 26 vs 25 occurrences per thousand words, $Z = 3.6$, $p < 0.001$; locatives, 15 vs. 14 occurrences, $Z = 2.1$, $p = 0.03$).

Next, we examined demonstrative frequency across the age and MLU range using correlational analysis[4]. There were positive correlations between MLU and determiner/pronoun frequency in Spanish ($r = 0.25$, $p = 0.02$) and English ($r = 0.20$, $p = 0.009$): determiners/pronouns were more frequent in children with longer MLU. Locative adverbs did not significantly correlate with MLU in Spanish ($r = 0.17$, $p = 0.11$) or English ($r = -0.10$, $p = 0.19$). Age correlated with MLU in English, $r = 0.40$, $p < 0.001$, but not in Spanish, $r = 0.14$, $p = 0.2$. Correlations between demonstrative frequency and age did not approach significance (r's < 0.15, p's > 0.14).

We also examined possible differences in child-directed speech across development. Parent demonstrative frequency correlated negatively with child MLU: parents used more demonstratives at the early stages of language acquisition and parent usage decreased with child language development: in English, $r = -0.17$, $p = 0.031$, and Spanish, $r = -0.22$, $p = 0.037$. Nevertheless, parents' and children's demonstrative frequency correlated positively in English, $r = 0.41$, $p < 0.001$, and Spanish, $r = 0.277$, $p = 0.008$. Changes in frequency of demonstrative words by MLU for children and parents are displayed in **Figure 5**.

Demonstrative Types in Child and Parent Speech

This analysis examined the relationship between the demonstrative words used by each parent-child dyad, particularly whether they tend to use the same demonstrative words during an interaction. A correlational analysis was performed on the frequency of each demonstrative word per thousand words between speakers (parent and child) within transcripts. Results are displayed in **Table 3**. Parents tended to use the same determiners/pronouns as the children, and rarely used others. This was also the case for distal locatives, but when children used proximal locatives parents were equally likely to use distal or proximal (English), or distal or medial (Spanish).

Conclusions of Study 1 (CHILDES Data)

Analysis of the spontaneous speech of 18 to 24 month old English and Spanish speaking children revealed that demonstratives are used by more than half of children from age 18 months, and at the single-word utterance stage. However, it is not until children are starting to produce two-word utterances that we see demonstratives in nearly all children. There were no significant between-language differences. What CHILDES data do not reveal is the order of acquisition of demonstratives, nor whether they appear among the first 50 words. That will be examined using parental report (CDI) data in Study 2. Findings from the descriptive analysis of CHILDES data on demonstrative use and parental input will be discussed in the General Discussion.

[4]Due to the number of outliers in the sample, bootstrap based on 1000 bootstrap samples was calculated. In none of the significant correlations did the 95% bootstrap confidence interval contain zero; therefore, we can be confident of the correlations' significance.

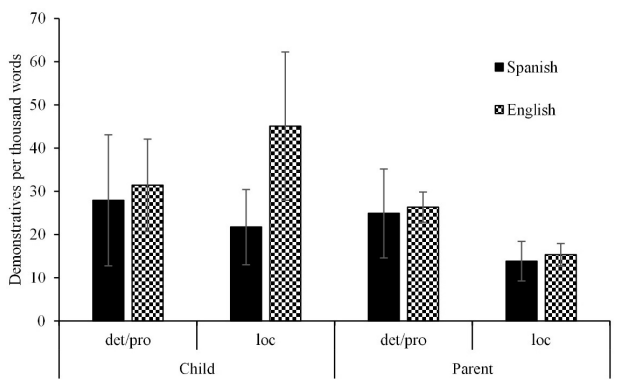


FIGURE 4 | Mean frequency of determiner/pronouns and locatives per thousand words in CHILDES corpora, by language and speaker. Error bars correspond with the 95% confidence interval for mean. Demonstratives were present in all Spanish parents' transcripts and in 98% of British parents' transcripts.

STUDY 2 (BASED ON CDI-WORDBANK DATA)

Study 2 investigates the acquisition of demonstrative words in English and Spanish using data from parental report. Specifically, we look at when the majority of children use demonstratives with respect to their vocabulary size and age in both languages.

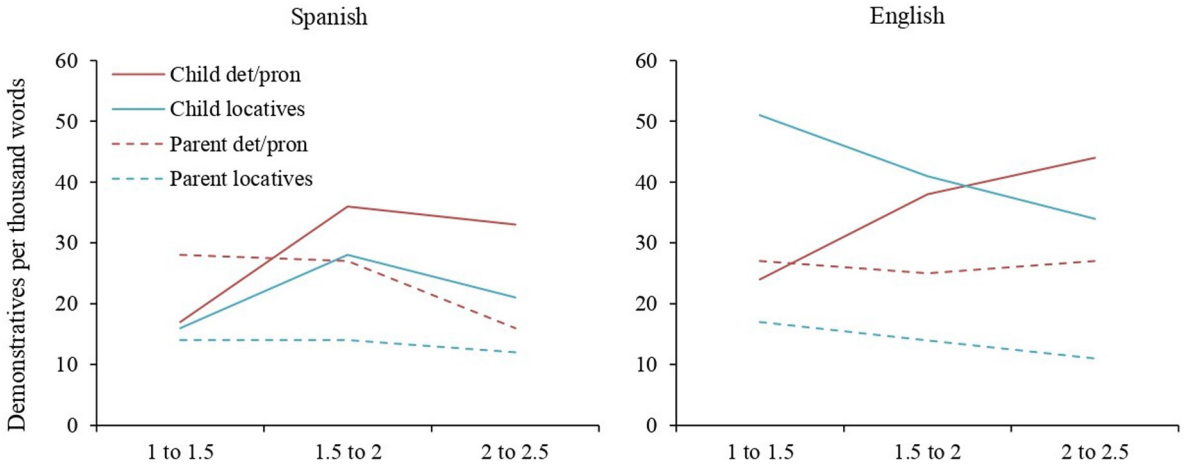


FIGURE 5 | Demonstrative frequency per thousand words for children and parents at each level of MLU in CHILDES corpora.

TABLE 3 | Within-transcripts correlations between parent and child demonstratives' frequency per thousand words.

			Child					
			Det/pronoun			Locative		
	Parent		Proximal	Medial	Distal	Proximal	Medial	Distal
Demonstrative	Spanish	Proximal	**0.21***	−0.00	−0.04	**0.26***	0.03	−0.03
		Medial	0.17	**0.42****	0.05	0.24*	**0.19**	0.03
		Distal	−0.07	−0.06	**−0.02**	−0.14	0.04	**0.40****
	English	Proximal	**0.17***	–	0.13	**0.22****	–	0.00
		Distal	0.13	–	**0.27****	0.23**	–	**0.28****

*Data from CHILDES. *p < 0.05; **p < 0.01; ***p < 0.001. Bold values indicate the correlation between parent's and child's frequency of use of the same word.*

Method

Origin of the Data

Data come from 277 monolingual speakers of European Spanish and 673 of British English, between the age of 18 and 24 months. Sample distribution by age is displayed in **Table 4**. Data sources: López et al. (2005), Floccia (2017).

Instrument

The instruments used were the Oxford CDI for British English and the Words and Sentences for European Spanish (Hamilton et al., 2000; López et al., 2005). These questionnaires are not a direct translation of each other, but an adaptation to fit linguistic and cultural differences. Therefore, although they include the same word categories, the Spanish version features more items (588) than the British one (418). The average vocabulary size for each age and language group is displayed in **Table 4**.

Demonstrative words in the English instrument include *this, that* and *there*, but not *here*, nor the plural forms *these* and *those*. The Spanish questionnaire features all demonstrative words, including gender and number variations (13 items, see **Table 1**).

Data Processing and Analysis

Data were extracted and processed using the *wordbankr* R package (Braginsky, 2018) on 25/11/2019. To make the two languages comparable, in Spanish we worked only with the singular forms of demonstratives[5]. A dummy variable was computed to indicate whether a child produced any demonstrative word, irrespective of the frequency. The percentage of children that produced demonstratives was compared at each Age and MLU level. Age levels were each month from 18 to 24 months. Minimum vocabulary size (CDI score) was binned in groups of 50 words (CDI score of 0 to 50 words, 51 to 100 words, and up to 400). Chi-squared tests on the raw data were used throughout. Two separate analyses were made, one for determiners/pronouns only, and one for all demonstratives including locatives.

Results

Acquisition of Demonstratives by Age in CDI Data

Figure 6 displays the percentage of children who used at least one demonstrative word by age and language group. From 21 months onward, more than half the Spanish children used at least one determiner/pronoun (*este, ese* and/or *aquel*). Including locatives, 68% of Spanish children produced at least

[5]None of the children produced only plural forms of demonstratives; plural forms in Spanish were always acquired after the singular forms. Therefore, this selection had no effect on the findings.

TABLE 4 | Sample size and mean productive vocabulary size and *SD* for each age and language group.

	Spanish		English	
Age (months)	**Sample size**	**Vocabulary Mean (*SD*)**	**Sample size**	**Vocabulary Mean (*SD*)**
18	50	70 (79)	118	51 (60)
19	27	84 (64)	109	82 (82)
20	36	117 (105)	144	110 (93)
21	41	144 (105)	75	130 (92)
22	38	184 (125)	28	151 (118)
23	30	230 (122)	112	187 (121)
24	55	257 (161)	87	220 (113)
Total sample size	277	No. items: 588	673	No. items: 418

Data from CDI Wordbank.

one demonstrative word from 18 months, and approached 100% at 22 months. In contrast, only 9% of British children produced at least one determiner/pronoun word by 18 months, 17% when including locatives. At 24 months, less than 50% of English speakers produced determiner/pronouns, and 55% when including locatives. At any age point, a greater number of Spanish children compared to British children produced at least one demonstrative, whether or not locatives were included in the analysis [all χ^2s (1) > 10, p's < 0.001, Bonferroni adjusted alpha level of 0.007].

Acquisition of Demonstratives by Vocabulary Size in CDI Data

Figure 7 displays the percentage of children who used demonstratives by minimum vocabulary size (CDI score) for each language. Less than half of the English speakers produced determiners/pronouns below a vocabulary of 300 words. Including locatives, more than half of the children produced at least one demonstrative from 200 words on, and reached ceiling after 350 words. For the Spanish sample, more than half of children produced determiners/pronouns from a vocabulary of 50 words on, and when including locatives, from 0 to 50 words, reaching ceiling at a vocabulary of 150-200 words. More Spanish children than British children produced demonstratives up until a vocabulary of 250 words, either considering determiners/pronouns alone or with locatives [all χ^2s (1) > 10, p's < 0.001, Bonferroni adjusted alpha level of 0.006]. There were no significant between-language differences thereafter [all χ^2s (1) > 3, p's > 0.1].

Conclusions of Study 2 (CDI-Wordbank Data)

Data from parental report reveal important crosslinguistic differences. The majority of Spanish speakers use at least one demonstrative from 18 months and among their first 50 words if locatives are included, whereas English speakers do not use demonstratives up until age two and a vocabulary size of 200 words, and even later if considering determiners/pronouns only. It was expected that fewer children would use demonstratives in CDI data compared to CHILDES data. However, the striking crosslinguistic differences solely in CDI data suggest possible sampling differences.

Demonstrative Production and Parental Education in the Spanish Sample

In the Spanish CDI sample, high education families were over-represented, with 77% of parents having college and graduate education. Maternal education is not reported in the British data, although it is presumably lower, since authors state that their sample SES was representative of the British population (sample composite or SES measurement were not reported in detail; Hamilton et al., 2000). Thus, our hypothesis is that the lower report of demonstrative use in British sample is due to the higher proportion of parents with low education, and the associated bias of underestimating children's knowledge of function words (Fenson et al., 1994). This was tested by analyzing the differences in report of demonstrative words between high education level (college and University, n = 222) and low education level parents (primary and secondary school, n = 52) in the Spanish sample (missing cases, n = 3). The mean age of children of both groups did not differ significantly (Mann-Whitney U, $Z = -0.38$, $p = 0.7$), nor the total CDI score (Mann-Whitney U, $Z = -0.65$, $p = 0.5$). More parents with higher education reported that their children used demonstratives, 88% vs 77%, $\chi^2(1) = 4.56$, $p = 0.03$. This supports the hypothesis that parental education might play a role in their accuracy in reporting demonstrative production. However, only 34% of British parents from our data reported demonstrative use, thus sampling issues cannot fully account for the cross-linguistic differences in Study 2.

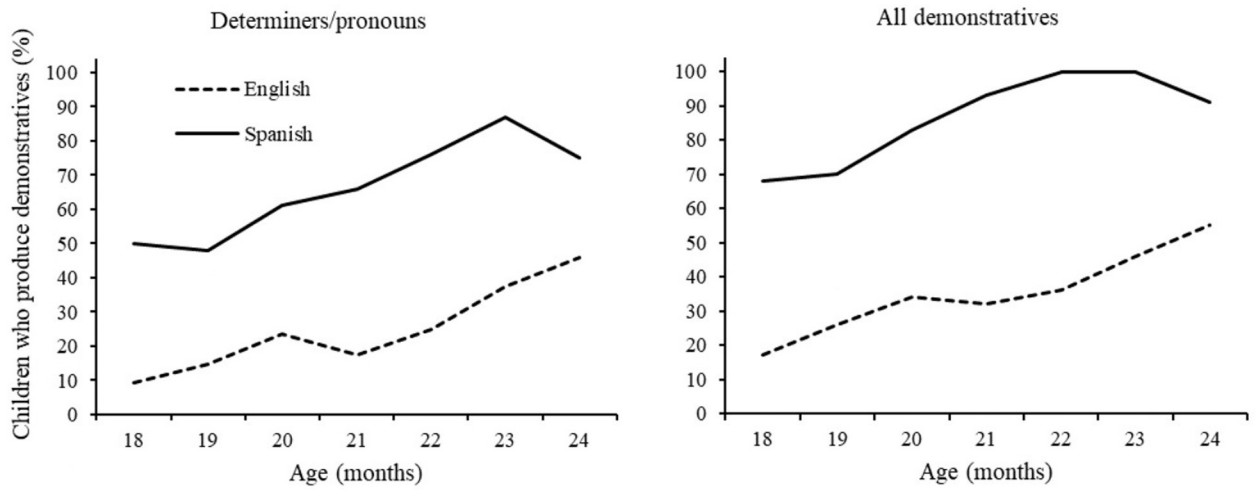


FIGURE 6 | Children who produce any demonstrative word by age and language (%). Data from CDI Wordbank.

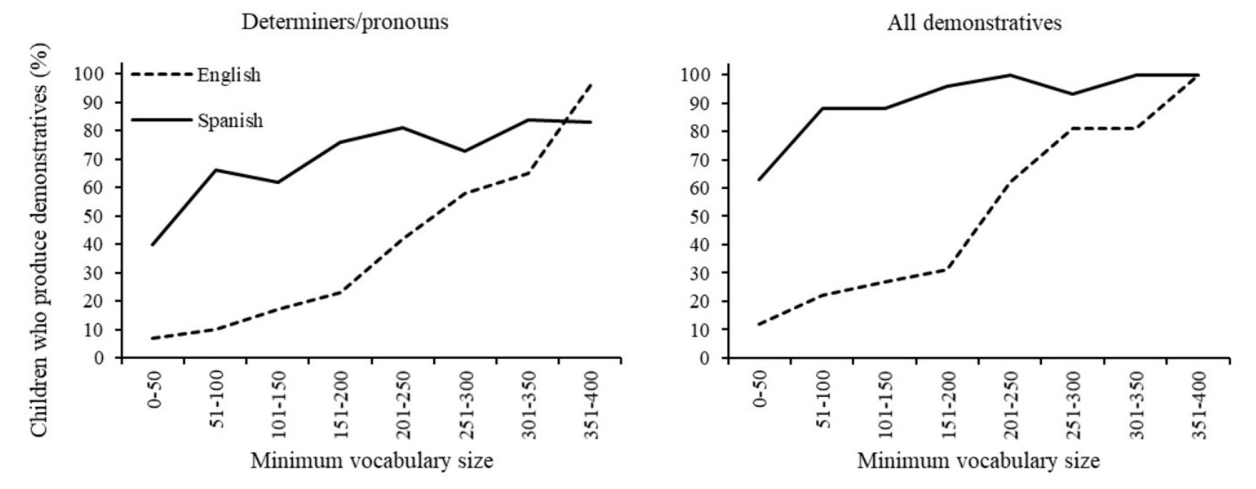


FIGURE 7 | Children who produce demonstrative words by vocabulary size (%). Data from CDI Wordbank.

GENERAL DISCUSSION

This work aimed to describe the acquisition and use of demonstrative words in infants and possible cross-linguistic differences. In Study 1, we analyzed corpus data, that allow measurement of mean length of utterance (MLU), word frequency and parent input. In Study 2, we looked at data from parental report, that feature a measure of vocabulary size and a large sample size. Results will help understand the role of demonstrative words in deictic communication and language acquisition in infancy. They are also interesting from a methodological point of view, contributing to assessing the suitability and validity of parental report and corpus analysis in the study of function words.

First, we asked whether demonstratives appear among children's first 50 words and at the earliest stages of language development (18 months). Results on age of acquisition differ between measures: according to the CDI results (Study 2), only around half of the English speakers use demonstratives by 24 months, whereas nearly all Spanish speakers used at least one demonstrative by the age of 22 months. In contrast, corpus data (Study 1) indicated that the majority of children of both languages produced at least one demonstrative word from 18 months and all of them did at 24 months. Data from CHILDES indicates that the majority of children from both languages use demonstratives from MLU 1 to 1.5, and reach ceiling with an MLU of 1.5 to 2. Data from the CDI showed at what point in vocabulary acquisition demonstratives appear. The majority of Spanish speakers have a demonstrative among their first 50 words (after the 50th word if considering determiners/pronouns only), reaching ceiling after the 150th word. In contrast, the majority of English speakers do not use demonstratives before their 200th word, reaching ceiling

only after their 350th word. This reflects a great discrepancy between CDI and CHILDES data, and it is unclear which one of these sources reflects a more accurate estimation. Nevertheless, we can confidently say that demonstratives do not typically appear before the 50th word, and they are more frequent in two-word utterances. We cannot make any firm statement about possible cross-linguistic differences because the results we obtained were very different between the two sources. We will discuss the possible methodological and sampling sources of discrepancies.

It was expected that the CDI data would underestimate demonstrative production with respect to corpus data (Salerni et al., 2007); however, CDI data also show striking differences between languages, while the corpus data do not. We suggested that differences might be due to sample SES disparity between languages and measures. This bias could have affected the results at two levels: first, because children of parents with higher education levels have an advantage for language development (Hoff, 2006); and second, because parents of low educational level may underestimate children's knowledge of function words in language inventories (Fenson et al., 1994). In contrast to the CDI data, the CHILDES sample for English may have an overrepresentation of higher SES families: one of the two largest corpora that compose the English corpus (Manchester corpus) is formed of middle-class families, while the other (Wells) has a representative sample extracted from the birth censuses. Thus, the average SES level in the British sample might be higher in CHILDES than in CDI data. Comparisons between high and low education parents in the Spanish sample support the hypothesis that low educational level parents might underestimate their children's use of demonstratives, but it is unlikely that it can fully explain the magnitude of the differences between languages in CDI data. One possibility is that language-specific factors, such as phonetics, might pose a disadvantage for the identification of demonstratives in English. Having listened to several CHILDES transcripts, our subjective impression is that young infant's verbalizations of *there* and *that* were often hard to distinguish from babbling, whereas the Spanish words *esto* or *aquí* were easier to recognize, perhaps because they are disyllabic words.

As argued in the introduction, neither checklist nor observational methods alone are ideal for estimating the proportion of particular word types in children's early vocabulary (Pine et al., 1996). However, combining both methods did not offer conclusive results either, because it is unclear whether the disparity between the two studies is due to methodological or sampling differences. We encourage researchers to take into consideration demographic variables in studies of this kind, while further research that will apply both methods to the same participants is needed to evaluate its impact in the results.

The second aim of this work was to describe the use of demonstratives in child spontaneous speech (Study 1). The analysis of CHILDES data revealed no significant differences between languages in the acquisition of demonstratives with respect to age and MLU. However, it did show that proximal demonstratives appear more often in Spanish and distal demonstratives in English, both in terms of frequency of use and of percentage of children using them at least once. Thus, whereas the use of demonstratives by infants is not a language-specific communicative tool, the preferred demonstrative term varies across languages.

One striking finding is that locatives and determiners/pronouns do not seem to have the same function in language development. Locatives appear earlier and are more frequent, particularly in English and in earlier stages. They are less complex than determiners/pronouns, which are more frequent in children with higher MLU. The most salient difference between languages in children transcripts is in the locative *there/ahí*. In English, it was the most frequent word in children's lexicons, and its frequency was particularly high in the youngest children. In contrast, the Spanish equivalent *ahí* (and the proximal *aquí*) was no more frequent than the determiner/pronouns. Our hypothesis is that *there* in English (unlike locative adverbs in Spanish) functions as a fixed expression instead of a deictic term, or as a verbalization linked to a particular action. This was the case for the children studied by Harris et al. (1988) and Barrett et al. (1991), who found that children acquired *there* among the first 10 words, but they used it in a very specific context: for example, one participant would only use it with the action of handing a toy. This use might be a precursor of the acquisition of deictic words (i.e., of generalizing *there* to indicate location). However, the analysis of transcripts provides limited context, particularly those of infants in the single-word stage, and thus makes it difficult to assess when children use demonstratives in a ritualistic way or as a deictic communication tool. Future research in the development of deictic communication might take this into consideration, and perhaps analyze separately determiners/pronouns and locatives.

Another interesting difference between the two languages is in the frequency of demonstratives: in English, two demonstratives, *there* and *that,* were among the five most frequent words of child's lexicon, whereas in Spanish the most frequent demonstratives, the proximal terms *este* and *aquí,* are the 11th and 13th most frequent words. Demonstrative words were also very common in parent speech, although parents used fewer demonstratives than children per thousand words, presumably due to their larger vocabulary.

The analysis of spontaneous speech also allowed description of parent use of demonstratives. Data revealed that parents use more demonstratives in children's earlier stages of language development, as indicated by a negative correlation between parents' frequency of demonstratives and children's MLU. This might indicate that parents move on to use words that are more complex than demonstratives at the moment in their child's language development when they are acquiring new words at a fast rate.

Interestingly, the frequency of use of each demonstrative term correlated between parent and child. This has potentially interesting implications for later development of spatial demonstratives to convey distance and semantic information. That parent and child are using the same demonstrative word in a given speech suggests that children are not switching the demonstrative term, as happens in adult speech: frequently in an interaction with objects, the speakers view them from opposite sides and therefore use opposed demonstratives (the speaker may use *this* for an object closer to them, whereas the

conversational partner refers to the same object with *that*). Our hypothesis is that parents repeat the demonstrative that the child uses in order to reinforce their word learning, while the spatial content of demonstratives (close or far) is not relevant at this stage. Taumoepeau and Ruffman (2008) have demonstrated that mothers are sensitive to what their child can and cannot understand in this age range; when talking about mental states, the speech parents use is only slightly more complex than their child's current level and within their zone of proximal development (Vygotsky, 1980), plausibly in order to aid their learning. This would predict that parents use demonstratives without considering their spatial dimension or deliberately adopt their child's perspective when the distance contrasts are too complex for the child's current level. One example of such behavior might be in the following script (Anne, 1;11, free play with mother).

Child: What [is] baby doing?
Mother: Which baby?
Child: This baby.
Mother: This one?
Child: Yeah
Mother: Oh dear that baby's fallen out of the pram.

In this example, the child uses the proximal demonstrative, then the mother repeats it, but her next sentence features the distal demonstrative for the same referent. The child, mother, and the referent (the baby doll) do not apparently change location during the exchange, so the mother's appropriate demonstrative would have been *that*. However, the mother first repeated the child's demonstrative as a reinforcement. Here is another example, in Spanish (Mendía, 1;08, free play with mother, includes video):

Child and mother are playing on the floor. Child turns around and refers to a game that is located slightly further, indicating that he would like to play with it some more. The child uses the proximal demonstrative and the mother uses it too.
Child: éte [: éste]. - This.
Child: má [: más]. - More.
Mother: muy bien (.) ¿más? - Very well. More?
Child: má [: más]. - More
Mother: ¿éste? - This one?
Mother: ¿hacemos éste otra vez? - Do we do this one again?
Child: títo [/] [?].
Mother: ¿éste otra vez? - This one again?

This hypothesis, however, should be taken with caution, since there are frequent examples where it does not occur. There are also numerous events in which it cannot be assessed because only parent or child use demonstratives. Parents' use of demonstratives according to the child's perspective might be limited to a specific developmental stage. Further research could investigate parent-child synchrony of demonstratives in video-recorded interactions, to see at what stage in development parents take their children's perspective with demonstrative words and how it influences their subsequent acquisition of the spatial contrast.

Results from the CHILDES corpora are to be interpreted with caution because of the small sample size in Spanish (seven children). Individual differences and preferences might have been overrepresented in our results. The CHILDES database would benefit from more contributions of early speech in languages other than English. Particularly, parent-child interactions in video format would be a valuable addition to the study of deictic communication in infancy.

CONCLUSION

We studied the acquisition and frequency of demonstrative words in English and Spanish using transcripts of spontaneous speech and parental report data. Results indicate that demonstratives do not typically appear before the 50th word and are more frequent at the two-word-utterance stage than at the onset of productive language. This work challenges previous claims about the acquisition of demonstratives (Clark, 1978). In line with other studies that have looked at deictic communication in infants (Capirci et al., 1996; Rodrigo et al., 2004), we conclude that demonstratives may not be the most frequent means of early verbal deixis; other words or verbalizations may take that function earlier in development, whereas demonstratives become more frequent in more elaborate utterances later on. Our work is limited to two languages and shows important discrepancies between measures; nevertheless, it might encourage researchers to pay closer attention to other word types or vocalizations when studying verbal deixis in early language development.

From a methodological point of view, comparing parental report and spontaneous speech data in the study of function words has highlighted the potential limitations of both measures. Further research needs to examine the suitability, limitations, or improvement of both methods for the study of function words in child speech.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

PG-P contributed with the conceptualization, methodology, investigation, data analysis, visualizations, theoretical framework, results interpretation, and writing. PG-F contributed with the conceptualization, supervision, theoretical framework, results interpretation, review, and editing. MD contributed with the theoretical framework, results interpretation, review, and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by EU H2020 ITN Marie Skłodowska-Curie Actions grant agreement no. 676063 (DCOMM).

ACKNOWLEDGMENTS

We would like to thank Olga Capirci for the remarks and discussion that motivated this study, and Samuel Forbes and two reviewers for their feedback, that substantially contributed to the improvement of the manuscript. Any remaining errors are solely ours.

REFERENCES

Barrett, M., Harris, M., and Chasin, J. (1991). Early lexical development and maternal speech: a comparison of children's initial and subsequent uses of words. *J. Child Lang.* 18, 21–40. doi: 10.1017/s0305000900013271

Braginsky, A., Sanchez, A., and Yurovsky, D. (2019). *childesr: Accessing the 'CHILDES' Database. R Package Version 0.1.2*. Available online at: https://CRAN.R-project.org/package=childesr (accessed October 17, 2019).

Braginsky, M. (2018). *wordbankr: Accessing the Wordbank Database. R Package Version 0.3.0*. Available online at: https://CRAN.R-project.org/package=wordbankr (accessed March 14, 2018).

Braine, M. D. S., and Bowerman, M. (1976). Children's first word combinations. *Monogr. Soc. Res. Child Dev.* 41:1.

Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., and Hirsh-Pasek, K. (2013). A long-term predictive validity study: can the CDI short form be used to predict language and early literacy skills four years later? *J. Child Lang.* 40, 821–835. doi: 10.1017/s030500091200030x

Capirci, O., Iverson, J., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two-word speech. *J. Child Lang.* 23, 645–673. doi: 10.1017/s0305000900008989

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A cross-linguistic study of early lexical development. *Cogn. Dev.* 10, 159–199. doi: 10.1016/0885-2014(95)90008-x

Clark, E. V. (1978). "From gesture to word: on the natural history of deixis in language acquisition," in *Human Growth and Development*, eds J. S. Bruner and A. Garton (Oxford: Oxford University Press), 85–120.

Clark, E. V., and Sengul, C. J. (1978). Strategies in the acquisition of deixis. *J. Child Lang.* 5, 457–475. doi: 10.1017/s0305000900002099

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Dale, P. S., Bates, E., Reznick, J. S., and Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *J. Child Lang.* 16, 239–249. doi: 10.1017/s0305000900010394

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs−Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Dev.* 76, 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–173.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*, 2nd Edn. Baltimore, MD: Brookes.

Floccia, C. (2017). Data collected with the Oxford CDI over a course of 5 years in Plymouth Babylab, UK. With the permission of Plunkett, K. and the Oxford CDI from Hamilton, A., Plunkett, K., & Schafer, G., (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: lower scores in the UK than the USA. *J. Child Lang.* 27, 689–705.

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* 44, 677–694. doi: 10.1017/s0305000916000209

Hamilton, A., Plunkett, K., and Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative development inventory: lower scores in the UK than the USA. *J. Child Lang.* 27, 689–705. doi: 10.1017/s0305000900004414

Harris, M., Barrett, M., Jones, D., and Brookes, S. (1988). Linguistic input and early word meaning. *J. Child Lang.* 15, 77–94. doi: 10.1017/s030500090001206x

Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002

Iverson, J., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–373.

López, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., and Martínez, M. (2005). *Inventarios de Desarrollo Comunicativo MacArthur: Manual Técnico*. Madrid: TEA Ediciones.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* 38, 1–135.

Pine, J. M., Lieven, E. V. M., and Rowland, C. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *J. Child Lang.* 23, 573–590. doi: 10.1017/s0305000900008953

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Appl. Psycholinguist.* 11, 145–166. doi: 10.1017/s0142716400008742

Rodrigo, M. J., González, A., de Vega, M., Muñetón-Ayala, M., and Rodríguez, G. (2004). From gestural to verbal deixis: a longitudinal study with Spanish infants and toddlers. *First Lang.* 24, 71–90. doi: 10.1177/0142723704041040

Salerni, N., Assanelli, A., D'Odorico, L., and Rossi, G. (2007). Qualitative aspects of productive vocabulary at the 200-and 500-word stages: a comparison between spontaneous speech and parental report data. *First Lang.* 27, 75–87. doi: 10.1177/0142723707067545

Silge, J., and Robinson, D. (2016). tidytext: text mining and analysis using tidy data principles in R. *J. Open Source Softw.* 1:37. doi: 10.21105/joss.00037

Taumoepeau, M., and Ruffman, T. (2008). Stepping stones to others' minds: maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Dev.* 79, 284–302. doi: 10.1111/j.1467-8624.2007.01126.x

Todisco, E., Guijarro-Fuentes, P., Collier, J., and Coventry, K. (in press). The temporal dynamics of deictic communication. *First Lang.*

Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard university press.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 July 2020
doi: 10.3389/fpsyg.2020.01712

Elevation as a Grammatical and Semantic Category of Demonstratives

*Diana Forker**

Department of Caucasus Studies, Friedrich-Schiller University of Jena, Jena, Germany

In this paper I study semantic and pragmatic properties of elevational demonstratives by means of a typological investigation of 50 languages with elevational demonstratives from all across the globe. The four basic verticality values expressed by elevational demonstratives are UP, DOWN, LEVEL, and ACROSS. They can be ordered along the elevational hierarchy (UP/DOWN > LEVEL/ACROSS), which reflects cross-linguistic tendencies in the expression of these values by demonstratives. Elevational values are frequently co-expressed with distance-based meanings of demonstratives, and it is almost always distal demonstratives that express elevation, whereas medial or proximal demonstratives can lack elevational distinctions. This means that elevational demonstratives largely refer to areas outside the peripersonal sphere in a similar way as simple distal demonstratives. In the proximal domain, fine grained semantic distinctions such as those encoded by elevational demonstratives are superfluous since this domain is accessible to the interlocutors who in the default case of a normal conversation are located in close proximity to each other. I then discuss metaphorical extensions of elevational demonstratives to non-spatial uses such as temporal and social deixis. There are a few languages in which elevational demonstratives with the meaning UP express the temporal meaning future, whereas the DOWN demonstratives encode past. This finding is particularly interesting in view of the widely-debated use of Mandarin Chinese spatial terms 'up' for past events and 'down' for future events, which show the opposite metaphorical extension. I finally examine areal tendencies and potential correlations between elevational demonstratives and the geographical location of speech communities in mountainous areas such as the Himalayas, the Papuan Highlands and the Caucasus. I tentatively conclude that languages spoken in similar topographic environments do not tend to have similar systems of elevational demonstratives if they belong to different language families.

Keywords: elevation, vertical axis, space, deixis, time, demonstrative pronouns

OPEN ACCESS

Edited by:
Holger Diessel, Friedrich-Schiller University of Jena, Germany

Reviewed by:
Niclas Burenhult, Lund University, Sweden
Eva Schultze-Berndt, The University of Manchester, United Kingdom

***Correspondence:**
Diana Forker
diana.forker@uni-jena.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *03 March 2020*
Accepted: *22 June 2020*
Published: *30 July 2020*

Citation:
Forker D (2020) Elevation as a Grammatical and Semantic Category of Demonstratives. Front. Psychol. 11:1712. doi: 10.3389/fpsyg.2020.01712

INTRODUCTION

The expression of space in grammars of natural languages is ubiquitous and 'spatial language' has been investigated for decades within many different linguistic subdisciplines and by means of various approaches and frameworks. However, research on the spatial category of elevation is just at the beginning and typological studies are lacking so far. Elevation refers to the

expression of a location of a figure with respect to the ground on the vertical axis. Many languages have words for 'up' and 'down' or 'higher' and 'lower,' but not all languages have this semantic distinction grammaticalized as part of certain closed class items, most notably demonstratives, which are the topic of this paper, but also spatial preverbs and case systems. Elevational meanings have repeatedly been grouped together with grammatical items that refer to salient landmarks (e.g., 'seawards'/'landwards,' 'upriver'/'downriver'). Such systems have been called 'environmental space deixis' (Bickel, 1997), 'spatial coordinate systems' (Burenhult, 2008) or 'topographical deixis' (Post, 2011). From a number of surveys, we can conclude that demonstratives expressing elevational distinctions are cross-linguistically not extremely common but also not extremely rare, but we lack detailed comparative studies (e.g., Hyslop, 1993; Ebert, 1994; Diessel, 1999; Dixon, 2003; Post, 2011, 2017; Schapper, 2014; Aikhenvald, 2015; Breunesse, 2019).

In this paper, I concentrate on the semantic and pragmatic properties of elevational demonstratives, more specifically, adnominal, spatial adverbial, and pronominal demonstratives. This study therefore goes beyond general cross-linguistic studies of demonstratives, which devote only a few sentences to demonstratives with elevational meaning. It also goes beyond more specific surveys such as Post (2011, 2017) and Schapper (2014), which devote considerable space to elevationals, but focus on particular linguistic areas/languages families. I first lay out the conceptional and notional background for verticality and its relation to deixis, and describe morphological, syntactic and semantic properties of elevational demonstratives. I then propose the elevational hierarchy along which the basic elevational meaning categories can be ordered. Subsequently, I examine the relationship between elevational meaning and distance contrasts of demonstratives and further semantic extensions of elevationals to indicate cardinal directions, social hierarchies, and temporal meanings. The data for this paper mainly come from grammatical descriptions of some 50 languages with elevational demonstratives from a range of different language families across the globe.

Abbreviations: 1, first person; 2, second person; 3, third person; A, most agent-like argument of a transitive verb; ABL, ablative; ABS, absolutive; ACC, accusative; ACT.FOC, action focus; ADD, additive focus; ART, article; AZR, adjectivalizer; CERT, certainty; CLF, nominal class; CMPL, completive aspect; COMP, comparative; COOR, coordinator; COP, copula; CQ, content question; CTR, contrastive; CURR.REL, current relevance; D, d-classifier; DAT, dative; DEM, demonstrative; DERIV, derivational affix; DOWN, down(ward); DST, distal; DU dual, number; DUR, durative; DXVB, deictic verb; EMPH, emphasis; EXIS, existential; F, feminine; FUT, future; GEN, genitive; H, hearer; IMP, imperative; INCL, inclusive; INST, instrumental; IPFV, imperfective; IRR, irrealis; ITER, iterative; LOC, locative; M, masculine; MAN, manner; MIR, mirative; N, neuter; NMLZ, nominalizer; NON.FUT, non-future; NPST, non-past; NSG, non-singular; PFV, perfective; PL, plural; PN, proper name; POL, polite; PROG, progressive; PROX, proximal; PROXH, hearer-proximal; PROXS, speaker-proximal; PRS, present; PRT, particle; PST, past tense; PURP, purposive; REL, marker of relative clause; REMPST, remote past tense; REP, reported; RN.TOP, relator noun with the meaning 'top'; S, speaker; SG, singular; SR, subordinator; SUB, subject; SUBJ, subject cross-referencing; TAG, tag particle; TOPIC, topic; TSR, temporal subordinator; UP, up(ward); VIS, visible; VOC, vocative.

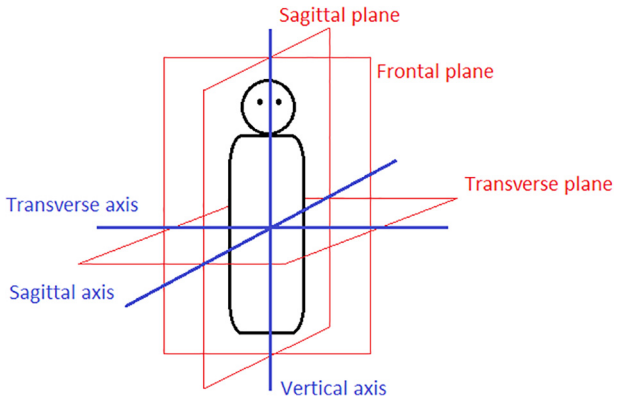


FIGURE 1 | The three body planes and axes.

CONCEPTIONAL AND NOTIONAL BACKGROUND

Verticality Within the Domain of Spatial Language

As said in the introduction, elevation refers to the expression of a location of a figure with respect to the ground on the vertical axis.[1] The three axes and planes through the human body provide the ground for three pairs of (linguistic) concepts, namely UP/DOWN, BACK/FRONT, and LEFT/RIGHT (**Figure 1**). Languages have a plethora of linguistic means to express locations of objects along the vertical axis, e.g., A is *above/over* B, A is *higher* than B, A can be *upward*, *uphill*, *up the road* with respect to B. This exemplification of English prepositions, adjectives and adverbs is far from being exhaustive. However, many languages do not have words referring to the sagittal (back/front) or transverse (left/right) axes or do not employ them regularly and in the same fashion as familiar European languages (Levinson, 2003, p. 46). And what is more relevant for the topic of this paper, the vertical axis is the only of the three axes that is encoded by demonstratives. No language has been reported so far to have demonstratives for the other two axes.[2]

The vertical axis is special in comparison to the other two axes of the body (frontal and sagittal) (**Figure 1**) for one major reason: gravity normally determines what counts as *up* and *down*. The position of a figure *above* or *over* a ground object is usually defined by gravity and thus in most instances in practice absolute (see also Clark, 1973). Positions along the vertical axis cannot easily be rotated or reflected in contrast to positions on the back/front and the left/right axes (i.e., *front* becomes *back* or *left* becomes *right* through rotation or reflection). Locations *in*

[1]The term 'figure' as used in this paper refers to the object or referent whose location is expressed and 'ground' is the 'relatum' in terms of Levinson (2003) or ground object or reference point, which can be ego or not. The third technical term employed here is 'anchor' or 'anchoring point,' which is the origin of the coordinate system by means of which the position of the figure is defined.

[2]Tamil may represent an exception to this claim. According to Levinson (2003, pp. 84, 105, 108) the proximal *itu* vs. distal *atu* demonstratives also have the interpretation 'that to the left' vs. 'that to the right.'

front of X or *left of X* are potentially ambiguous because they can depend on the relative viewpoint: By contrast, normally we unambiguously understand *above/over X* or *below/under X* if we know the position of X. Because of this (usual) unambiguity of locations along the vertical axis, the anchor point of an observer can shift without difficulty (we will see below what repercussions this has for elevational demonstratives). Levinson summarizes the distinctiveness of the vertical axis by stating that "the intrinsic (canonical position of objects), the relative (perception from an upright stance) and the absolute (as defined by the gravitational axis) tend to coincide" (Levinson, 2003, p. 75; see also Carlson-Radvansky and Irwin, 1993, p. 224 for the same observation). For a detailed explication of the concept of frames of reference in spatial language and its three basic types, intrinsic, absolute and relative, see Levinson (2003, pp. 24–61). Bender and Beller (2014, p. 348) provide useful graphic representations of the basic types and further subtypes.

The intrinsic frame of reference entails that the ground and the origin of the coordinate system that serves as anchoring point are identical and the spatial relation between the figure and the ground is binary. In an absolute frame of reference, there is also a binary relation, but this time between the ground and independently given salient geographical landmarks or cardinal directions that serve as anchoring points (e.g., *north of X*). By contrast, in a relative frame of reference there is a ternary relation because in addition to the figure and the ground (relatum or ground object) there is an anchoring point (=the origin of the coordinate system).

As was just said, the vertical axis is special because of its natural grounding in gravity. However, we can 'escape gravity' in the sense that we can change the frame of reference from absolute to intrinsic or relative. **Figure 2** shows a person stretched out on the ground. The description of object B as *over the head* entails a relative (to an external upright observer) or absolute frame of reference as determined by gravity.[3] We make use of an intrinsic frame of reference when we refer to object A, which is located at the same elevation of the head of the person and aligned with it along the same horizontal axis, as *over the head*. However, in none of the languages in my sample I encountered examples illustrating an elevational demonstrative used with an intrinsic frame of reference (i.e., DEM.UP A). From a logical point of view there is no reason to exclude such usages, but their actual existence has yet to be proven by future research.[4]

The peculiarity of the vertical axis has also been examined in psychology. Vertical spatial relations among objects remain largely constant with respect to a moving observer whereas on the transverse (i.e., horizontal) plane spatial relations change more frequently. Therefore, human beings are faster at retrieving the names of objects located along the vertical axis than along the other two axes (Bryant et al., 1992).

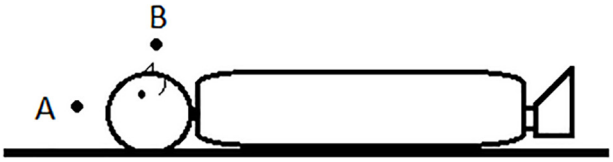


FIGURE 2 | Location 'over the head.'

With respect to the topic of this paper the category of deixis comes into play because the items examined are either categorized as demonstratives themselves or as parts (bound roots, affixes, or clitics) of demonstratives. Following Diessel (1999, p. 2; see also Dixon, 2003), demonstratives are deictic expressions serving specific syntactic and pragmatic functions. Commonly distinguished categories of deixis are person, place, time, discourse, and social deixis (Diessel, 2012, p. 2414), and demonstratives usually express place deixis/spatial deixis (Diessel, 1999, p. 36). The term 'spatial deixis' refers to the localization of a figure relative to a ground (object) in terms of (radial) distance categories by means of language (e.g., *here* vs. *there*), or in combination with a pointing gesture (Levinson, 2003, p. 65). The deictic center is usually egocentric, i.e., the speaker's location serves as the ground, but can also shift depending on the speech situation. According to the survey in Diessel (2013), which included 234 languages, demonstratives are distance-neutral or express up to five distance contrasts (i.e., five positions that differ in terms of distance from the deictic center). In purely distance-based systems, the deictic center is the speaker (i.e., egocentric system) and thus identical for all demonstratives. Among the languages with a three-way distance contrast (88 languages in Diessel, 2013), around one third are so-called person-oriented or person-based systems. This means that one of the three demonstratives expresses proximity to the hearer, and therefore the deictic center is not the speaker, but the hearer.

I will discuss the interaction of deixis with elevation in the Section "The vertical dimension and its relation to deixis" after having described in more detail elevational meanings.

MATERIALS AND METHODS

Language Sample

For this paper I surveyed elevational demonstratives in 50 languages from 20 language families plus one isolate. My sample is a convenience sample because elevational demonstratives are not particularly frequent in the world's languages. Many of the languages have been identified through the works by Diessel (1999); Post (2011, 2017); Sarvasy (2014), and Breunesse (2019). In addition, an unpublished database by Killian (unpublished), which contains data on demonstrative systems in around 1,100 of the world's languages, served as a major reference. According to Killian, the database is not completely unbiased, but it

[3]The relative frame of reference could be changed if the observer stands upside down on his head. In that case, object B would be located *below the head*.

[4]Research in psychology has shown that the acceptability of statement that employ an intrinsic frame of reference for describing the location of a fly with respect to a donkey in a picture comparable to **Figure 2** is lower than the acceptability of statements that make use of an absolute frame (Carlson-Radvansky and Irwin, 1993). See also Friederici and Levelt (1990) for similar results. Friederici and Levelt (1990) also tested two astronauts during space flight and found that if gravity is not present the relative frame of reference as determined from the own position and the head-retinal coordinates is preferred above visual background cues such as the position of trees.

covers all areas of the world and more than half of the world's language families.

The elevational demonstrative systems of the 50 surveyed languages have been coded for a number of formal and semantic properties. The list of languages, schematic overviews of the elevational demonstrative systems together with genealogical and geographical information on the area where the languages are spoken and references are given in the **Supplementary Appendix Table A12**. Language families and subbranches in which elevational demonstratives are attested for many languages are East Caucasian languages, Eskimo-Aleut languages, Sino-Tibetan (in particular Bodic languages, Kiranti languages, Macro-Tani), Timor-Alor-Pantar languages, Nuclear Trans New Guinea, and Omotic languages.

Morphological and Syntactic Features

In general, demonstratives can be bound and unbound forms, whereby the bound forms are normally clitics and not affixes (Diessel, 1999, p. 22–25). They can be morphologically simple and complex. The same is true for the subclass of elevational demonstratives, but with a further complication because elevation constitutes an additional semantic component on top of the basic demonstrative meaning (which is distance-based and/or person-based). This additional semantic component is either not expressed by a separate morpheme and then part of the basic demonstrative stem, or it is expressed by a separate morpheme. For this study, morphemes were considered elevational demonstratives if they combine with a demonstrative stem in a single lexical item, or appear to express both demonstrative and elevational functions.[5] In other words, elevational demonstratives are often morphologically and always semantically complex expressions that constitute single word forms. Based on these considerations, the items under investigation can be divided into three basic types:[6]

(i) Co-expression of elevational and demonstrative meaning in a single morpheme.
(ii) Obligatory co-occurrence of demonstrative morphemes with elevational morphemes in a single word-form.
    (a) No occurrence of elevational morphemes outside these forms.
    (b) Occurrence of elevational morphemes outside these forms.
(iii) Optional co-occurrence of demonstrative morphemes with elevational morphemes in a single word-form.
    (a) No occurrence of elevational morphemes outside these forms.
    (b) Occurrence of elevational morphemes outside these forms.

The elevational morphemes that obligatorily or optionally co-occur with demonstrative morphemes are bound roots, affixes or clitics. Based on the descriptions it is not always possible to distinguish between the subtypes (ii) and (iii) because not all grammars explicitly state whether the elevational morpheme also occurs in some other parts of speech (e.g., as preverb or spatial case affix).

Co-expression of elevational and basic deictic demonstrative meaning at the synchronic level occurs in Muna (**Table 1**), Daga (**Table 9**), Yakkha (**Table 11**), Iaai (32), (33), Jahai, Abui, Tidore, Sougb, Tulil, Hatam, Fore, Usan, Yale, Dadibi, and Zayse. **Table 1** shows the demonstrative system of Muna (Malayo-Polynesian, Sulawesi). Out of six demonstrative forms (with anaphoric and deictic variants), only one (*tatu*) co-expresses the elevational meaning UP and the deictic meaning distal. Only when occurring in opposition with *tatu*, the neutral distal demonstrative *watu* can also mean DOWN or LEVEL.

Eipo (Mek, Eastern Highlands of Papua New Guinea) has two subsets of demonstratives (**Table 2**). The elevational values in both subsets are obligatorily co-expressed with the deictic meaning DISTAL. The second subset, which contains the intensifier *d-*, expresses additional distance or contrast.

There are a number of languages such as Baskeet, Yupno, Makalero (**Table 3**), and Khasi, which obligatorily require further morphology to be added to the elevational demonstrative. This can be gender marking as in the pronominal demonstratives in

TABLE 1 | The structure of Muna demonstratives (van den Berg, 1989/2013, p. 89).

	Anaphoric	Deictic
S-proximal	*ini*	*a-ini*
H-proximal	*itu*	*a-itu*
Away from S, H, but nearby	*maitu*	*a-maitu*
Far (neutral)	*watu*	*a-watu*
Far (high) [UP]	*tatu*	*a-tatu*
Not visible, audible, unspecified for time	*nagha*	*a-nagha*

TABLE 2 | The structure of Eipo demonstratives (Heeschen, 1982, pp. 84–86; Heeschen, 1998, p. 143).

		+ Additional distance
Proximal ('here')	*a-*	*d-a-*
Distal high ('up there, above') [UP]	*ei-*	*d-ei-*
Distal down ('down there') [DOWN]	*ou-, u-*	*d-ou-*
Distal across ('across there') [ACROSS]	*or-, er-*	*d-or-*

TABLE 3 | The demonstrative system of Makalero (Huber, 2011, p. 232).

	Meaning	Nominal demonstratives	Deictic verbs
Same level	Proximal to speaker	*ere*	*e'*
	Proximal to hearer	*uere*	*ue'*
	Distal from speaker and hearer	*umere*	*ume'*
	Higher elevation [UP]	*udere*	*ude'*
	Lower elevation [DOWN]	*ufere*	*ufe'*

[5]Therefore, the study does not include items that have been characterized as deictic, but explicitly not as demonstratives as, for examples, the elevational deictic directionals analyzed by Cauchard (2018). It is of course desirable for future research to also include them and systematically compare them with elevational demonstratives (thanks to an reviewer for pointing that out to me).

[6]I thank an reviewer for suggesting this classification, which differs from my original classification.

Baskeet (8) or in Khasi. Or it can be derivational suffixes for the formation of demonstrative pronouns, adverbs or verbs as in Yupno and Makalero, and Khasi adverbial demonstratives. The elevational demonstrative morphemes themselves cannot be clearly separated further and no unambiguous part with purely elevational meaning can be identified. For example, in Makalero (Alor-Pantar, East Timor) nominal and verbal demonstratives are derived from the same bound roots by means of the nominalizer *-r-* and the verbalizer (glottal stop; **Table 3**).

In a number of languages, the elevational demonstratives are clearly diachronically complex, but synchronically the elevational part cannot be separated or is not treated as a bound root, affix, or clitic. Languages belonging to this type are Sanzhi Dargwa, Hua (Yagaria dialect), Central Alaskan Yupik, Kurtöp (9), (22), and Galo (19).

Languages with morphologically complex elevational demonstratives in which the elevational meaning is expressed by bound roots or affixes and regularly combines with demonstrative stems are Blagar, Tauya, Tanacross (**Table 4**), Koyukon, Andi (**Table 5**), Manambu (**Table 8**), Ngiyambaa (7), and Dyirbal (10). For example, demonstratives in the Athabaskan language Tanacross (Alaska) morphologically and semantically combine deictic meaning (distality) with specific topographic and elevational morphemes. The topographic and elevational morphemes express also directional and locational meanings (e.g., allative).

The demonstratives in the East Caucasian language Andi (Zilo dialect, Caucasus, Russia) are particularly transparent and consist of stems that express distance- and person-based deixis, followed by a range of further optional suffixes such as an emphatic marker, the elevational morphemes and gender suffixes (and/or oblique stem markers and case suffixes not displayed in **Table 5**).

The classification introduced at the beginning of this section makes a distinction between (a) and (b) subtypes, whereby the (a) subtypes refer to elevational morphemes that only combine with demonstrative morphemes, whereas the (b) subtypes of elevational morphemes also occur outside the demonstrative systems. In Andi and Manambu, the elevational markers are only used with the deictic demonstratives and thus belong to the (a) subtype. By contrast, in Dyirbal they can also be added to verbs to form verbs of motion (Dixon, 1972, pp. 57, 322), and thus Dyirbal belongs to the (b) subtype. Similarly, in Eipo, Sougb, Nêlêmwa-Nixumwak, and Abui[7] deictic motion verbs can attach the elevationals.

TABLE 5 | The structure of adnominal and pronominal demonstratives in Andi (Verhees, 2019).

Stem: *hV* distance/ person-based deixis	Emphasis	Elevation	Gender, (number)
Proximal *ho-*	*-n(V)*	Same level [LEVEL] *-dV*	Singular human male *-w*
Medial *he-*		Lower [DOWN] *-gV*	Singular human female *-j*
Distal *hu-*		Higher [UP] *-ɬV*	Singular neuter 1 *-b*
Distal *hi-*			Singular neuter 2 *-r*

However, this cross-categorical formal flexibility is not the rule. There are a few languages in my sample that have specialized motion verbs referring to upward or downward movement, but the elevational markers that those verbs contain are historically unrelated to the elevational demonstratives (Galo, Sanzhi Dargwa, Yupno, and Bantawa).

For the (b) subtypes, the question can be asked what the nature of the elevational morpheme is, in particular, whether they are themselves deictic or non-deictic. However, for this paper the answer to that question is largely irrelevant, because I am only interested in the combined forms, i.e., the co-occurrence of demonstratives and elevational morphemes. This touches upon a problem I encountered during this study. I had to rely on the often implicit assumptions of the linguists whose descriptions I consulted that the items classified as 'elevational demonstratives' represent single lexical units. In languages such as Manambu, Sougb, or Nêlêmwa-Nixumwak, in which the morphemes with the elevational semantics can be readily identified and are sometimes also used with lexical items other than demonstratives (e.g., verbs), the elevationals resemble English non-deictic expressions such as *up*. English *up* can co-occur with adverbial demonstratives (*up there*) and verbs (*climb up*). However, no linguist has ever claimed that English has an elevational demonstrative although such a claim would perhaps be imaginable if we wrote *up-there* or *upthere* instead of *up there*. This means that among the languages studied for this paper there might be languages that are actually not extremely different from

[7]Only the elevational marker with the meaning UP.

TABLE 4 | The demonstrative system of Tanacross (Holton, 2019).

Distance		Topographic, etc.	Allative	Ablative	Punctual	Areal
		Upstream [UP]	*-ndéʔe*	*-ndî·dz*	*-ndé·*	*-ndí·g*
		Downstream [DOWN]	*-ndá·ʔa*	*-ndâ·dz*	*-nda·*	
Proximal	*da-*	Inland	*-ndeg*	*-ndêdz*	*-ndég*	*-ndóg*
Distal 1	*na-*	Waterward	*-tθɛ́nʔ*		*-tθí·*	*-tθúg*
Distal 2	*ya-*	Ahead	*-nɛð*			*-noð*
Distal 3	*yaʔa-*	Across [ACROSS]	*-ná·nʔ*	*-ndáz*	*-ná·n*	*-ndás*
Neutral	*a-*	Away	*-ʔɛ́nʔ*	*-ʔáz*		*-ʔóg*
		Above [UP]	*-deg*	*-dêdz*	*-dé·*	
		Below [DOWN]	*-ʒégʔ*	*-ʒêz*	*-ʒé·*	*-ʒóg*

English, but for which the author of the grammar has reasons to assume that a morphologically and semantically complex expression translates with, e.g., *up-there* or *upthere*, constitutes a single lexical item.

Most of the elevational demonstratives take further optional or obligatory derivational and/or inflectional suffixes (most commonly gender, number, case, nominalizers or adverbializers). They are part of paradigms or subparadigms that consist of three (Andi) to five (Makalero, Manambu, and Buru) items on average, but more than seven members are not exceptional (Daga). For instance, Tanacross has nine items (**Table 4**), and Movima even has 14 basic demonstratives occurring in paradigmatic relationship. See Diessel (1999, pp. 32–33) for general morphological properties of demonstratives, which also apply to elevational demonstratives.

Diessel (1999, p. 57) distinguishes four syntactic contexts in which demonstratives occur. These contexts are (i) pronominal use, (ii) adnominal use (i.e., as determiner), (iii) spatial and manner adverbial use, and (iv) identificational use in copula and non-verbal clauses. The identificational context of use has been and/or is also called 'predicative' use (e.g., in the first typological paper on this topic written by Killian, unpublished.). As stated in the introduction, I focus on adnominal, spatial adverbial, and pronominal elevational demonstratives. Because other forms need further research they will only be mentioned in passing. Examples (1)–(4) illustrate all four contexts.

(1) Sougb (Reesink, 2002, p. 224): adnominal context
tu *gaih*
house DEM.DST.DOWN/east
'that house to the east/down there.'

(2) Sougb (Reesink, 2002, p. 224): pronominal context
[answer to question 'Which one?' or 'Where is X?']
mong-gaih
DERIV-DEM.DST.DOWN/east
'that one down/to the east.'

(3) Sougb (Reesink, 2000, p. 125): adverbial context
Inyomusi / Inyomusi / dara Slora / ingga
Inyomus Inyomus with Slora DEM.DST

lan *la-(e)desa* *gaih*
3DU 3DU-settle DEM.DST.DOWN/east
'Inyomus and Slora, the two of them settled down there (north of Sururei).'

(4) Dime (Seyoum, 2008, p. 76): identificational context
čúú-ná *sugur* *ʔámzi* *dán*
DEM.DOWN-3SG.F Bodi woman COP
'That down there is a Bodi woman.'

The last context (iv) has several subtypes (presentative, identifier, localizer, and copular demonstratives, see Killian, unpublished for the full typology, explanations and examples). Copular demonstratives are cross-linguistically rare (Killian, unpublished; see also Guérin, 2015). Among the languages in my sample Blagar, Makalero, and Tidore have elevational demonstrative verbs with the meanings 'be here/there up/down' that exhibit predicative use:

(5) Makalero (Huber, 2011, p. 393)
. . . *fi-ama=ni* *udere'*
. . . 1PL.INCL-garden=CTR DEM.UP.DXVB
'. . . our garden is the one up there.'

The four basic contexts are attested to various extents for elevational demonstratives. In Maale, only the adverbial use is found. The adverbial context can be considered the minimal context of use probably attested for all languages in my sample. The adverbial use normally refers to the occurrence of elevational demonstratives in the function of spatial adverbs (3), (24). Makalero and Tidore do not have genuine elevational adverbial demonstratives, and the adverbial function is fulfilled by demonstrative verbs (5). Blagar, Galo (26) and all East Caucasian languages in my sample (Avar, Lak, Andi, and Sanzhi Dargwa) have not only spatial elevational demonstratives, but also a further class of elevational demonstratives that function as manner adverbs, e.g., Blagar *do-laŋ* (up.there-as) 'like that/those up there (not necessarily visible)' (Steinhauer, 2014, p. 159).

All languages expect for Tanacross and Maale employ elevational demonstratives in the adnominal context, and this is therefore the second most commonly attested type of usage. In Usan and Eipo, elevational demonstratives can be used as modifiers within a noun phrase (i.e., adnominal use), but not in the syntactic function of determiners. Instead, they co-occur with determiners.

The degree to which the syntactic contexts are expressed by specialized, formally distinct elevational demonstratives varies. I did not come across any language that always distinguishes all four types formally. Nungon makes formal distinctions between the first three syntactic contexts (Sarvasy, 2014, pp. 404–419). Sanzhi Dargwa and other East Caucasian languages formally distinguish elevational demonstrative adverbials (with spatial and manner semantics) from nominal demonstratives by means of derivational suffixes, and also has a separate class of copular demonstratives. Nominal demonstratives can be used adnominally or pronominally in Sanzhi, but they are only case-marked in the latter use (and thus formally distinct). The elevational demonstratives of Baskeet, Tauya, Galo, and Kurtöp seem to pattern alike. In Sougb, the pronominal and/or identificational use requires additional morphology (2), but adnominal and adverbial uses are identically and unmarked (1), (3). In Yakkha, the unmarked forms function as adverbials (12), and the adnominal forms are derived (15). Yale does not formally distinguish between adnominal and adverbial elevational demonstratives (and the author of the grammar does not explicitly mention a pronominal or identificational use).

The Basic Semantic Distinctions of Elevational Demonstratives

Semantically, elevational demonstratives are deictic expressions that also convey elevational or verticality distinctions. Following Schapper (2014), I distinguish four basic concepts for verticality

values and will employ them in the glosses of examples in order to facilitate understanding and comparison (6) (even though individual authors may use alternative terms, e.g., *higher*, *upward*, or *above* instead of UP).[8] The term LEVEL includes a more specific term ACROSS:

(6) UP: up(ward) location of a figure with respect to a ground object.[9]
DOWN: down(ward) location of a figure with respect to a ground object.
LEVEL: same level of altitude.
ACROSS: across a space (e.g., a room, a valley at the same level, a river, a jungle).

I further adopt and simplify the classification of Burenhult (2008) and differentiate between two basic types of elevational systems:

(i) General elevational demonstrative systems:

- The location is determined according to an imagined vertical (longitudinal) axis that runs through the ground (e.g., human body).

(ii) Topographic elevational demonstrative systems:

- The location is determined with respect to the geophysical environment.

The first type ('general') corresponds to Burenhult's 'verticality proper' and 'global elevation,' and the second term ('topographic') to his 'geophysical elevation.' General elevationals are used in accordance with the gravitational axis. They can have very local meanings, which means that they can be applied, for instance, to refer to positions close to the speaker, inside a room or in the immediate environment (7), (8) but they are also used to denote locations in the geophysical environment (9).

(7) Ngiyambaa (Donaldson, 1980, p. 141)
bala *ŋalu-dhar=na* *balaŋ-ga* *waɽa-nha*
head.ABS that.INST-DOWN=3ABS head-LOC stand-PRS
'She is standing head downward, on (her) head.'

(8) Baskeet (Treis, 2019)
án, *zúggóò*, *áyssh* *lokíí*
2.SG.M.VOC Zugga.VOC meat DEM.UP.M

né *núúb* *ímm-ísh*
2SG 1PL.DAT give-2SG.IMP.POL
'[The hyenas said:] "Come on, Zugga, please, give us this meat up there!"' [Meat hanging in the roof].

(9) Kurtöp (Hyslop, 2017, p. 126)
khwe=gi *wome=na=ta* *gari*
water=GEN DEM.DOWN=LOC=EMPH car

yam *nâ=mi* *tshe*
road COP.EXIS.MIR=TAG PRT
'There was a road down near the river, right.' (lit. 'car road', i.e., road usable by cars).

Genuine topographic elevationals refer on the basis of the geophysical environment. There are two types of landmarks outside and generally further away from the speaker that naturally expand along the vertical dimension, namely topographical contour (i.e., mountains including hills or large rocks) and hydrological contour (i.e., rivers and creeks).[10] The vertical dimension of rivers might not be obvious at first glance. But what connects rivers with what was said before about the vertical axis is the fact that gravity causes the flow of the water in a certain direction and the direction is absolute and independent of an anchoring point. I did not find any other types of landmarks defining topographic elevational demonstratives.

Topographic elevational demonstratives basically mean something like 'uphill'/'downhill,' 'upriver'/'downriver' and the like. For instance, Dyirbal has an elaborated set of twelve so-called 'spatial indicators' that are added to demonstratives or other noun markers and express topographic elevation, e.g., 'downhill,' 'uphill,' 'downriver,' 'upriver,' and 'across river' (Dixon, 1972, p. 48; Dixon, 2003, p. 98). Dixon further adds that 'river' is the more specific meaning and the other terms translated by 'hill' rather mean 'not river' and can also refer to locations such as cliffs or trees. The topographic elevationals can be followed by another marker from a smaller set that contains only three items that encode general elevation and the meaning 'out in front' but also seem to have some additional meanings that are not explicitly discussed in the grammar (Dixon, 1972, p. 48).

(10) Dyirbal (Dixon, 1972, p. 102)
bala-n *dɲugumbil* *bani-ɲ*
DEM.DST.VIS-F woman come-FUT

ya-gu-l-baydj
DEM.PROX.VIS-DAT-M-SHORT.DST.DOWNHILL

yara-gu *miyanda-ŋu-gu* *bural-ŋay-gu*
man-DAT laugh-REL-DAT see-?-PURP
'Woman will come to see men laughing just down here.'

[8]One reviewer pointed out that the verticality values in (1) could further be subdivided into those with static meaning and those with directional meaning. The first type expresses static vertical relations and could be used in answers to the question *Where is X?* (*above/below/at the same level/across*). The second type combines the spatial relation with the expression of motion or path, has directional meaning and thus answers the question *Where is X moving to?* (*up/upward/down/downward/along the same level/across*). However, this further semantic distinction is usually not discussed in the reference grammars and descriptions that I used for this paper. The table in the Appendix provides the English translations for all items surveyed (sometimes translated from another language such as Russian). As the table shows, the majority of elevationals have been translated by 'up' and 'down.' A few descriptions use 'above' and 'below,' 'higher' and 'lower,' and some employ both static relational and directional terms. I am not in a position to judge if the translations reflect a cross-linguistic tendency for elevationals to encode more often directional than locational meanings. This question can only be answered by future research and in particular by more detailed transcriptions.

[9]The term 'ground object' is used here in the sense of 'relatum' (Levinson, 2003). It is not necessarily identical with the speaker, because elevationals are not inherently deictic (see the discussion in Section "The Vertical Dimension and Its Relation to Deixis" below).

[10]As one reviewer pointed out, candidate additions might be 'up/down wind' and 'up/down ocean current.' On the other hand, it needs to be clarified by further investigation whether those categories and also 'upriver/downriver' manifest any relation at all to elevational deixis in the sense of whether for these categories the vertical axis as it is determined by gravity plays any role at all.

(11) Dyirbal (Dixon, 1972, p. 227)
bali-dawulu-gu *wandi-n*
there.all-LONG.DST.UPRIVER-ALL motion.up-NON.FUT

budʒay-gu / dʒala-bara-gu
bathe-PURP shallow-COMP-ALL
'(We will) go upriver to bathe, to a shallower place (than Gunbay).'

The distinction between general and topographic elevational demonstratives applies not just to the UP and DOWN meanings but also to LEVEL and ACROSS. For instance, terms that express ACROSS can be topographic and refer to locations across a valley at the same altitude of the opposite mountain as in Yakkha (12), or across the river as in Tanacross (13). They can also be general as in Usan and applied in the local domain (14).

(12) Yakkha (Schackow, 2015, p. 188)
nhaŋ *yunna* *buddhini=ca*
and.then this.ACROSS Buddhist.woman=ADD

eko *pi-ŋ*
one give.PST-1SG.A
'And I gave one to the Buddhist woman (living) over there.'

(13) Tanacross (Holton, 2000, p. 296)
ya'-^{n}dâ'z *ts'eỷ*
DEM.DST-ACROSS.ABL boat

ži' *naʔatłe'ł*
in ITER-PROG-D-go.by.boat
'He's coming back (from) across the river in a boat.'

(14) Usan (Reesink, 1984, p. 112)
tabin *eng* *ire-t* *beg-es-eis*
dish the DEM.DST.ACROSS-at put-for.me-SG.IMP
'Put the dish over there for me!'

Levinson (2018, pp. 27, 35) states that topographic demonstratives make use of an absolute frame of reference because the referent is located "on a notional gradient (upriver/downriver and uphill/downhill) which actually delivers an angle on the horizontal." He adds that such local landmarks do not have the same abstract properties as cardinal directions (Levinson, 2003, p. 90). This hints at one major problem concerning research on elevational demonstratives. Several languages have been claimed to possess topographic elevationals that employ an absolute frame of reference, but these claims are normally not proven by a comprehensive argumentation and detailed data. I suspect that these claims are probably sometimes wrong or at least misleading because, first, the authors do not provide unambiguous evidence that the relevant items refer on the basis of the geophysical environment and not simply to the vertical dimension. Second, the descriptions lack a solid proof of the absolute frame of reference as opposed to the relative or intrinsic frame.

In order to prove that an elevational demonstrative really makes use of an absolute frame of reference one has to explicate the coordinate system that serves as the observer-independent anchoring point in a similar way as cardinal directions. Above I explained that gravity is the natural source for the direction of elevationals and thus for the determination of what counts as UP and what as DOWN independently of an observer or an intrinsic orientation of the ground. This type of absolute frame of reference is also entailed in many usages of adverbs or adjectives such as English *up* vs. *down* or *high* vs. *low* (Clark, 1973), but these items can also be used with a relative frame of reference or an intrinsic frame of reference. What is thus needed when describing elevational demonstratives is to test if they can also refer to the position A in **Figure 2** (intrinsic frame), or relative to an anchor point that is distinct from the observer, e.g., to object A in **Figure 3**, or if such usages are always excluded. Only in the latter case the meaning would truly entail an absolute frame.

Second, a simple translation of a demonstrative as 'uphill' is not a proof for its topographic meaning with an absolute frame of reference. In particular, it is not sufficient if the demonstrative only occurs in example sentences that refer to people, animals, and other relatively big objects such as trees or houses and their location in the outside geophysical environment. If it is really a mountain or river that serves as the absolute landmark, then in a situation such as the one depicted in **Figure 4**, location A is 'downhill' and location B 'uphill' even though on a general vertical axis A is located further away from the ground and thus higher than B. In topographic systems the locations of the points

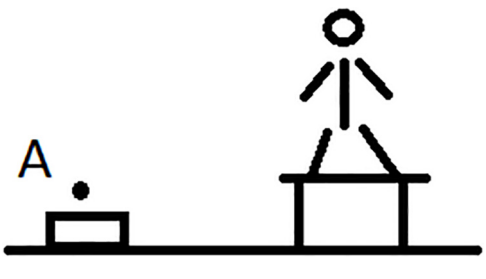


FIGURE 3 | Elevation and the relative frame of reference.

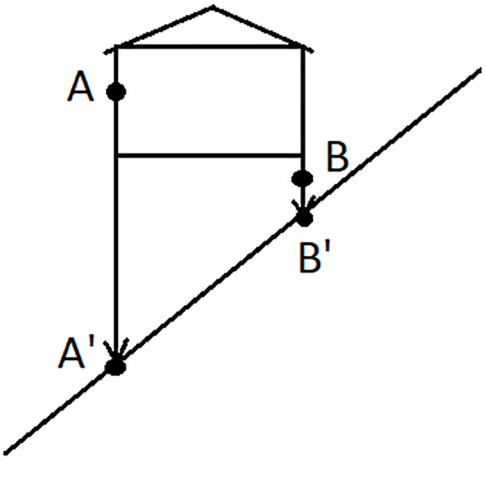


FIGURE 4 | Topographic elevationals.

A and B are projected on the ground and the positions of A' and B' determine the use of the appropriate demonstratives.

An example in point comes from Yakkha, which has two types of topographic elevational adverbials of which elevational demonstratives are formed (see **Table 11** for the first type). In Yakkha, a spider can be referred to as being on the 'downhill' side of the speaker, even if it is located on the same elevation level as the speaker and thus factually not lower than the deictic center (Schackow, 2015, pp. 188–189).

Another important factor to keep in mind when investigating the meaning of elevational demonstratives is scale or domain of use. We have to distinguish at least three domains (which obviously form a continuum and therefore lack clear borders):

(i) The local domain: the minimal local scale is the peripersonal sphere, but it extends to the area inside a house or its immediate surrounding; locations within this area are often visible.
(ii) The local larger, but delimited environment, e.g., a village, a valley, or an island; locations within this area can be visible or not.
(iii) The global scale, e.g., locations on other continents that are never visible from the location of the speaker.

Since general elevationals can normally be used both in the local domain and in the larger domain (and sometimes even at the global scale), they have to be carefully distinguished from topographic elevationals that are projected into the minimal local domain.

Metaphoric usage extensions, projections onto the horizontal plane and conventionalized uses can create problems for the correct categorization of elevational demonstratives as general or topographic because they might obscure the basic elevational meanings. For example, Sanzhi Dargwa has a general elevational system clearly based on an abstract vertical axis (Forker, 2019). However, at the scale of the main modern settlement, which is located in the lowland coastal area close to the Caspian Sea with virtually no differences in height, there is an 'upper' part of the village located closer to the hills and a 'lower' part located closer to the sea coast. When talking about inhabitants of the village, a person might conventionally be referred to by an UP or DOWN demonstrative based on the permanent location of her house within the village, which is mentally divided into an upper part and a lower part, and not on the location of that person with respect to the speaker or another spatial anchoring.

Elevational demonstratives that are characterized as 'topographic' in grammars can be used at the local scale such as within a house or close by a house or, with respect to a tree. For instance, Tanacross and other Northern Dene languages have genuine topographic elevational systems (in addition to the general elevational demonstratives) that conventionally extend to the micro level. This means that within a house there are four directions/locations, namely 'upstream,' 'downstream,' 'inland,' and 'across' because traditionally houses have been built with the door toward the water (Holton, 2000, p. 298). Therefore, an object is, for example, located 'upriver' when its location is referred to with topographic demonstratives. The division of the areas within a house are even used within modern houses that do not always face the water. In this language, 'uphill' location is at the same time away from the river, and 'upriver' ('upstream') is along the river and thus orthogonal to 'uphill'. This means that in terms of cardinal directions and gravity (i.e., location above sea level) 'uphill' and 'upstream' differ (Gary Holton, p.c.).

Similarly, by means of the second topographic system of Yakkha the 'uphill' and 'downhill' elevationals can be mapped onto the human body and teeth are then referred to as uphill, i.e., 'upper teeth' and downhill 'lower teeth' irrespectively of their actual position (even when a person is not in the canonical upright position).

(15) Yakkha (Schackow, 2015, p. 190)
mo=ha *keŋ=ci*
DEM.DOWNHILL=NMLZ.NSG tooth=NSG
'lower teeth'

to=ha *keŋ=ci*
DEM.UPHILL=NMLZ.NSG tooth=NSG
'upper teeth'

The two systems (general and topographic) as portrayed so far are idealized prototypes. Based on the descriptions that I consulted it is not always possible to determine if an elevational system falls into the one or the other category. In addition, it seems that there are systems that cannot be categorized as truly belonging to the one or to the other type, or should be analyzed as combining both types. For instance, the elevationals of Galo are translated as given in (16) (Post, 2007, pp. 349–350).

(16) LEVEL: "on the same or an unknown topographical/riverine level, or to the east, west or an unknown direction of the deictic center or one's home."
UP: "up, upward, upriver, or to the north of the deictic center or of one's home."
DOWN: "downward, downriver, or to the south of the deictic center or of one's home."

They are used at the local scale (17), the larger local scale (18) and the global scale (26). Only when the referent is potentially visible (i.e., within the minimal local and larger local domain) the relevant items encode elevational and riverine meanings.

(17) Galo (Post, 2011, p. 146)
bâə *jəkkâə* *làa?*
bə *jə-kə=ə* *làa*
DEM.DST.DOWN who-GEN=COP.IPFV CQ
'Whose is that (thing down there)?'

(18) Galo (Post, 2007, p. 357)
mootûm *tə*
mootùm *tə*
jungle DEM.DST.UP

rəkênə, *maazídù!*
rə-kèn=əə *maazí-dùu*
exist-AZR:good/easy=COP.IPFV very.much-IPFV
'The jungle (up there) is nice to be in, it really is!' (RmR, CC 118).

When the referent is not potentially visible and also not located on the path of a nearby river, but is separate from the speaker by at least a mountain range (i.e., global scale), then the same items function as labels for cardinal directions (19), and elevational differences are ignored. For instance, the speaker who uttered (19) is located in a village at around 100 m above sea level and Itanagar, where he would like to go, is situated at around 440 m and thus higher, and to the south but not visible from his village. In such a context, the anchor point can be the actual location of the speaker, or her/his home village can serve as conventionalized anchor point (similar to the conventionalized use of Sanzhi Dargwa demonstratives mentioned above). For unknown locations, the LEVEL items can be used as default demonstratives.

(19) Galo (Post, 2007, p. 355)
itanagár *bolò* *jômbə*
itanagar *bolò* *joombə̀*
place DEM.DST.LOC.DOWN how

ŋó *iirə́ nâ?*
ŋó *ì i-rə̀-nà=əə*
1.SG descend-IRR-NMLZ.SUB=COP.IPFV
'How am I to go (south) to Itanagar (having neither car nor money)?'

Four languages in my sample have two separate sets of elevational demonstratives, one set of general and another set of topographic elevationals (Dyirbal, Tanacross, Cora, and Buru, **Table 6**). In Buru (Malayo-Polynesia, Moluccas of Indonesia) topographic elevationals express three elevational values (UP, DOWN, and ACROSS) and general elevational morphemes only two (UP and DOWN) (**Table 6**). Grimes (1991, p. 170) does not provide a precise definition for the term 'emic,' but writes that the concept 'away from an emic center' as it is expressed by the topographic demonstrative *lawe* in Buru indicates 'energy directed away from the actor.' It is possible though not unambiguously clear from the description that this formulation can be translated into 'away from the speaker.'

(20) Buru: general (Grimes, 1991, p. 241)
toho *fi* *saka* *kau* *luke-n* *di* *beka!*
descend LOC DEM.UP tree tip-GEN DEM.DST first
'Come down right now from up in the top of that tree!'

(21) Buru: topographic and distance-based (Grimes, 1991, p. 405)
tu *ana-fina* *dae*
with child-female DEM.UPSTREAM

naa, *fila-n* *ba*
DEM.PROX lightning-GEN DUR

lata, *fila-n* *ba* *leo*
cut lightning-GEN DUR precede
'And as for this girl up there, [she] was gorgeous.'
(idiomatic; lit. 'her radiance was striking, her radiance was preceding').

TABLE 6 | The structure of Buru demonstratives (Grimes, 1991, p. 168).

Distance and definiteness		Topographic and general	
naa	Definite proximal	*pao*	Down, downward [DOWN]
dii	Definite distal (non-proximal)	*lawe*	Downstream/away from emic center/far [DOWN]
saa	Indefinite (specific or non-specific)	*saka*	Up, upward [UP]
		dae	Upstream/toward emic center [UP]
		aki	Across (stream, valley, ridge) [ACROSS]

The Vertical Dimension and Its Relation to Deixis

As stated in Section "Verticality Within the Domain of Spatial Language" above, demonstratives are deictic and express distance-based meanings with the speaker (ego) as deictic center or person-based meanings that additionally consider the position of the hearer. Verticality is not inherently deictic because the ground or anchoring point is not exclusively the speaker (Fillmore, 1982, pp. 39, 51; Diessel, 2012, p. 2,421). Nevertheless, terms expressing verticality can be relational and they can be used with relation to the speaker, which then may lead to the impression that the verticality component in elevational demonstratives is, by itself, deictic.[11] For instance, Kurtöp elevational demonstratives have been glossed as deictic with the speaker as deictic center. However, in (22) the UP-demonstrative occurs together with the hearsay evidential, which means that the speaker has acquired her/his information from the speech of others. This is a clear indication that the speaker cannot be the deictic center that serves as the point of anchoring for the location of the woman. The location of the woman is rather described as being higher than before after she had climbed up to the top of the roof.

(22) Kurtöp (Hyslop, 2017, p. 75)
gonpa=i *yau* *pangkap*
temple=GEN DEM.UP roof

je=do *thrang-wala=ri*
RN.TOP=LOC climb-PFV=REP
'(She) climbed up there on top of the roof (it is said).'

If elevationals were deictic by themselves, they would be ego-centered or only allow for shifting the deictic center to another speech act participant. But several descriptions explicitly mention that the anchor point serving as the ground (=deictic center) for elevational demonstratives can easily shift, e.g., in a story it shifts to a protagonist or to another salient inanimate anchor point [Tulil as analyzed by Meng (2018) and Ma Manda as examined in Pennington (2016)].[12]

[11] In fact, in his earlier work Diessel (1999, p. 41) had claimed that elevational demonstratives are deictic with the speaker being the deictic center.

[12] Most descriptions do not specify how the anchor point is determined. This can perhaps be taken as an argument that the elevational meaning part of demonstratives is, in fact, not deictic, but simply relational, and there is no need to explicitly specify that, as there is no need to specify that adverbs or adpositions such as *above* are relational.

Furthermore, as illustrated by means of **Figures 2**, **3** and in the discussion of the preceding section, when studying elevational morphemes it is necessary to examine whether they allow not only for the absolute frame of reference but if intrinsic and relative interpretations are also available. As I already explained, it is sound to expect the absolute use to be the default such that the interpretation of 'down there' in (17) is normally understood in relation to the position of the speaker and not some other ground object because of gravity. However, since we know that other elevational terms such as ABOVE or BELOW can, in principle, be employed within intrinsic and relative frames of reference, it is desirable in future research on elevational demonstratives to systematically test if there are any elevational demonstratives that can also be used in that way.

RESULTS

The Elevational Hierarchy

In (6), I introduced the basic terms for verticality values. These values can be ordered along the elevational hierarchy that reflects cross-linguistic frequency of occurrence (23):

(23) UP/DOWN > LEVEL/ACROSS

Elevational demonstratives with the meanings UP and DOWN are more commonly found than those with the meanings LEVEL or ACROSS (**Table 7**). All languages with LEVEL or ACROSS elevationals also have DOWN and UP elevationals. A minimal system of elevational demonstratives consists of one item for UP or one item for DOWN, but far more common is to have one term for each of the values UP and DOWN. So far, I did not find any language with both LEVEL and ACROSS elevational demonstratives, so these two values seem to exclude each other (although semantically ACROSS can be considered a sub-category of LEVEL). The more specific value ACROSS (8 languages) occurs only around half as often as LEVEL (19 languages).

Eipo and Andi and have all three types of demonstratives (**Tables 2**, **5**); Manambu has UP and DOWN (**Table 8**), and Muna has just UP (**Table 1**). Hatam has even two terms for UP (*nyo* 'sloping up', *hu* 'vertically up'), but only one for DOWN (*mu*) (Reesink, 1999, pp. 60–61).

Elevation, Distance and the Sagittal Axis

In their demonstrative systems, languages repeatedly combine elevation with distance. This means that the values DISTAL and PROXIMAL (and also MEDIAL for those languages that make a ternary distinction) are either obligatorily co-expressed or optionally combined with elevational items if the elevationals are morphemes that are formally independent of the distance-based deictics. There are languages in which all distance-based deictics can be combined with all elevationals. For example, in Manambu (Ndu, Sepik, Papua New Guinea), three person-based deictic stems take gender, number and the current relevance suffix, followed by the topographic and general elevational morphemes (**Table 8**).

In many other languages there are some restrictions. Daga (Papuan), for instance, has a particularly rich system with 14 demonstratives, of which two are merely person-based, eight co-express three distance-based meanings (CLOSE, DISTAL, and FAR DISTAL) with the elevational values UP, DOWN, and LEVEL, and four more encode only elevational meanings (**Table 9**). Yupno combines MEDIAL and DISTAL but not PROXIMAL with elevationals (Cooperrider et al., 2017, p. 771).

In those languages that optionally or obligatorily conjoin elevational meanings with distance, it is almost always the distal demonstratives that express elevation, whereas medial or proximal demonstratives can lack elevational distinctions. For example, in Andi (**Table 5**), only the distal demonstrative roots can attach elevational suffixes. In Muna and Eipo (**Tables 1**, **2**), elevational semantics and distal deixis are obligatorily co-expressed.

Kewapi (Enga-Kewa-Huli, Southern Highlands of Papua New Guinea) has a rich set of 13 demonstratives of which nine co-express elevational meanings, and relative distance and at the same time additional distance from the speaker ('away from the speaker') (**Table 10**; Yarapea, 2006, pp. 75–79).

TABLE 7 | The frequency of elevational systems classified according to basic elevational meanings[1].

	DOWN	UP	UP + DOWN	UP + DOWN + LEVEL	UP + DOWN + ACROSS
# of languages	1	2	21	19	8

[1] *I include systems that have items with elevational meanings for which I am not entirely sure that they fit the definitions. These items are given in parenthesis in* ***Supplementary Appendix Table A12****, e.g., U/D/(L). A few languages have more than one set, therefore the total for the number of languages in* ***Table 5*** *is higher than the number of languages surveyed. Don Killian (p.c.) pointed out three more languages that are relevant for the topic of this paper, but not included in the Appendix and in the table: Moskona (Gravelle, 2010, pp. 199–200) that has* UP *('upward') and* ACROSS*, which seems to be a very rare combination because it omits* DOWN*. Tepehuan (Willett, 1991, p. 92) has a separate term for* UP *('higher') and combines* DOWN *('lower') and* LEVEL *in one term, and Edolo (Gossner, 1994, pp. 85–87) that has two items for* DOWN *('that below,' 'that far below'), but just one for* UP *and one for* LEVEL*.*

TABLE 8 | Structure of Manambu demonstratives (Aikhenvald, 2015).

Stem	Suffixes	Topographic and general
S-proximal *kə-*	Feminine singular *-ɨ*	Up [UP] *-wur*
H-proximal *wa-*	Masculine singular *-d*	Down [DOWN] *-d(a)*
Distal *a-*	Dual *-bər*	Across [ACROSS] *-aki*
	Plural *-di*	Outwards *-aku*
	Current relevance *-na*	Off-river *-wula*

TABLE 9 | The structure of Daga demonstratives (Murane, 1974, p. 38).

Close to speaker	*ma*	Close to hearer	*ame*
Close higher [UP]	*uta*	Close lower [DOWN]	*ita*
Distal higher [UP]	*utu*	Distal lower [DOWN]	*isi*
Far distal higher [UP]	*use*	Far distal lower [DOWN]	*ise*
Same level [LEVEL]	*ata*	Far distal same level [LEVEL]	*ase*
Overhead [UP]	*oea*	Underneath [DOWN]	*ea*
Up, high [UP]	*ao*	Down, low [DOWN]	*ae*

TABLE 10 | The demonstrative system of Kewapi (Yarapea, 2006, p. 77).

	Relative distance –>	Close		Mid	Far
		Close	Mid		
	Specific location	*gó*	*go*		
	Generic location	*o*	*apo*		
Away	Upward [UP]	*sopo*		*sogo*	*só*
from	Downward [DOWN]	*nopo*		*nogo*	*nó*
speaker	Horizontal [LEVEL]	*mopo*		*mogo*	*mó*

TABLE 11 | The structure of Yakkha demonstratives,/u/-forms (Schackow, 2015, pp. 94, 187).

	Singular	Non-singular/non-count
Proximal	*na*	*kha*
Distal	*nna*	*ŋkha(ci)/nnakha(ci)*
Anaphoric	*honna*	*hoŋkha(ci)*
Proximal-up [UP]	*tunna*	*tukha*
Proximal-down [DOWN]	*munna*	*mukha*
Proximal-across [ACROSS]	*yunna*	*yukha*

As **Table 10** shows, the elevational demonstratives that encode relative proximity and middle distance are morphologically complex in contrast to the elevational demonstratives that encode relative distance. This can be taken as another way of the default co-expression of elevation with further distance as opposed to proximity or middle distance.

Even in a language such as Lak, in which the elevational demonstratives cannot unambiguously be analyzed as co-expressing distance or proximity to the hearer or a third referent, they are not used when the respective locations are so close that the speaker can touch them with her/his finger (e.g., a hat on the head is not located UP).[13] Thus, it seems that elevational demonstratives largely refer to areas outside the peripersonal sphere in a similar way as simple, non-elevational distal demonstratives (e.g., Coventry et al., 2008). I propose that this can be explained in the following way: in the proximal domain, fine grained semantic distinctions are superfluous since this area is accessible to the interlocutors who in the default case of a normal conversation are located in close proximity to each other [(see also Imai, 2003, p. 42) for a similar observation]. I also suggest that the same principle should apply to other semantic distinctions that demonstratives in some languages express such as visibility or audibility since such semantic categories are only relevant when the referent is not near to the speaker.

The only language I found so far that contradicts this otherwise robust cross-linguistic tendency is Yakkha. This language has two cognate sets of basic adverbial elevational roots, which are classified in the grammar as 'topographic.' The first set, which in the grammar is called '/u/-forms' based on their stem vowel, is given in the lower part of **Table 11**. According to Schackow (2015, p. 187), the '/u/-forms' combine with the proximal demonstrative (singular *na*, non-singular *kha*), but not with the distal or anaphoric demonstratives (**Table 11**). Thus, items such as *tunna* or *tukha* are morphologically complex, consisting of a morpheme with elevational meaning, followed by a morpheme with (originally) proximal demonstrative meaning.[14]

This type of co-expression or combination of distance and elevation in demonstratives is not obligatory because there are languages such as Makalero (**Table 3**), Hatam, Iaai, Hua, Tidore, and Baskeet (8), in which elevational demonstratives are unmarked for distance and cannot be co-expressed with distance. However, those languages constitute a minority.

I encountered only very few cases of elevational demonstratives that combine with person-based deictic systems and therefore express person-based elevational meanings, e.g., Manambu (24) (**Table 8**).

(24) Manambu (Aikhenvald, 2015, p. 213)
wakuli *wa-na-d*
mouse DEM.PROXH-CURR.REL-DOWN

rə-na
sit-ACT.FOC+3F.SG.SUBJ
'A mouse is sitting here close to you in the mentioned (downstream) location.'

If languages have elevationals and person-based deictics, these meanings are more commonly separately expressed as, for instance, in Muna, Daga (**Table 9**) or Sanzhi Dargwa. The reason for the relative rareness of person-based elevational demonstratives is probably unnecessary specificity. In practice, locations above the speaker and above the addressee during a conversation largely coincide.

On the horizontal plane, the genuinely vertical dimension can, in principle, be translated into FURTHER/NEARER (or FRONT/BACK) along the sagittal axis (Bender and Beller, 2014). This means that FURTHER is equated with UP and NEARER with DOWN. This kind of projection happens at least in Sanzhi Dargwa (Forker, 2019), Tulil (Meng, 2018, p. 266), Nungon (Sarvasy, 2014, p. 413) and Belhare (Bickel, 1997), and has been called 'person-morphic mapping' by Bickel (1997, pp. 58–60, 68). In Sanzhi, the projection occurs not only within the local, peripersonal sphere, for example, items on a table in front of the speaker are located as UP when they are further away and DOWN when they are closer to the speaker (but always in front of the speaker). The same kind of projection is applied at the global scale on an imagined map, e.g., Estonians are located UP with respect to Latvians because Estonia is further to the north (Forker, 2019). The projection can be explained by the fact that due to their upright position human beings have to move the

[13]This has been tested by the author during fieldwork in Dagestan in September 2019. It is not described in any of the grammars of Lak.

[14]D. Schackow (email from May, 08, 2020) confirmed that it is the proximal demonstratives that combine with the elevationals, and not the distal ones, but that for reasons yet to be clarified the singular forms contain a double 'n' instead of the expected 'n' (e.g., *tunna* instead of the expected form *tuna*). Furthermore, Schackow writes that the function of the demonstrative roots when attached to the elevationals in **Table 11** is in the first place nominalization and that the semantic component of distance is rather neutralized such that the composite form can be used for both proximal and distal reference.

head downward in order to look at proximal items whereas the gaze goes upward in order to look at distal items (see Bickel (1997 and references therein). An alternative explanation could be that positions further away from the speaker are (almost) unlimited in the sense that there is no clear and unambiguous natural boundary or limit (e.g., if we climb up a mountain we can see even further away). Similarly, there is no unambiguous natural boundary or limit for the direction upward of the vertical axis. By contrast, the direction downward is limited by the ground as are locations near or close to the speaker limited by the position of the speaker.[15]

Cardinal Directions

There are a number of languages whose elevational demonstratives also encode cardinal directions, but these meanings seem not to be available within the local domain. Examples were given in (16) and (19) from Galo. Other languages are Makalero, Bantawa, Baskeet, and Sougb. Usually only two opposite cardinal directions are encoded. Which elevational expresses which compass direction depends on the local position of the mountains that serve as anchor points and thus varies from language to language. For instance, as (25) shows, in Galo we have UP = north, DOWN = south, and LEVEL = east or west. The first two equations are also found in Bantawa. In Makalero and Baskeet, the relation is UP = east and DOWN = west, (and Baskeet has additionally 'over there' = north/south). In Sougb the equation is the opposite, namely UP = west and DOWN = east (26). In Iaai, the elevationals are in complementary distribution with other items that also convey compass points.

(25) Galo (Post, 2011, p. 147)

ʔastrée lijáa	*bəmbə̀*	*kaí rə̀m*,
ʔastreelijáa	*bəmbə̀*	*kaí-rə̀=ə̀m*
place	DST.MAN.DOWN	big-IRR=TSR

buppɨ̂ɨ.	*minə́ rələ̀m*	*paarûu rə́.*
buppɨ̂ɨ	*minərə́l=ə̀m*	*pàa-rûu-rə́*
all	mineral(<Eng)=ACC	get-CERT-IRR

'If Australia down there [in the south, D.F.] is so big, certainly you'll find all (manner of) minerals.'

(26) Sougb (Reesink, 2002, p. 225)

d-odo	*dig*	*gahi-da*
1SG-carry	to	DEM.DST.DOWN/east-go

'I carried (it) in an eastern direction/down.'

Temporal Reference

In three languages of my sample, the UP-demonstratives carry the temporal meaning FUTURE, whereas the DOWN-demonstratives encode PAST (Tulil, Ma Manda, and Towet dialect of Nungon). The languages are spoken in Papua New Guinea, but in different areas of the country, and they belong to two different language families. In Iaai, an Oceanic language from New Caledonia, only the second equation, i.e., DOWN = PAST exists. In the following, I will provide examples from the four languages and discuss this type of spatial metaphor. I will also mention a few other languages in which spatial verticality metaphorically maps onto time.

Tulil (Taulil-Butam) has three morphologically complex demonstrative stems with elevational meaning that can be used for temporal expression (Meng, 2018, pp. 240, 263, 271). The first two demonstratives are formed by reduplication and the third one by compounding:

- *mə* 'down, downhill, downstream' > *pmə* 'down distal, back' > '(near/far) past.'
- *bo* 'up, uphill, upstream' > *pbo* 'front, up near' > '(near/far) future.'
- *mu* 'far from speaker and hearer' + *mə* 'down' > *mumə* 'down distal' > 'far past/future.'

When functioning as demonstrative determiners, they can be employed with nouns such as *vənu(=a)* 'day,' *atade(=e)* 'week,' *vəgam(=e)* 'month,' or *laləng(=a)* 'year,' whereby demonstratives can precede or follow the noun (27). They are also used as independent demonstrative pronouns. The temporal meaning of the first two elevational demonstratives can be schematized as DOWN = BACK = PAST and UP = FRONT = FUTURE, and it is possible that the temporal meanings are, in fact, based on the 'front'/'back' meanings. It is well known that words for 'front' and 'back' are commonly used as temporal metaphors in a wide range of different languages and cultures (e.g., Traugott, 1978; Haspelmath, 1997, pp. 56–63; Bender and Beller, 2014). I do not have an explanation for the third demonstrative and the grammar provides only one example (27), in which its meaning seems to correspond to the meaning of the first and is thus in accordance with the DOWN = BACK = PAST schema.

(27) Tulil (Meng, 2018, p. 271)

be	*laləng=a*	***a-pmə***,
at	year=SG.CLF.M	3SG.M-DOWN

a-mu-mə,	*məte*
3SG.M-DST-DOWN	like

təgət=a	*me*	*nereita*	*ko*
one=SG.CLF.M	and	six	plus

mukəm.magərung,	*me*	*libəti*	*me*
three.SG.CLF.M	and	five	and

nereita	*ko*	*mukəm*,	*laləng=a*	*a-bət.*
six	plus	two,	year=SG.CLF.M	3SG.M-PROXH

'In a year **before**, like 1985, that year.'

Note that in the following example the elevational morpheme is actually an adverbial demonstrative with originally spatial function (due to the locative prefix *nə-* > *nə-p-bo* 'up there'), but it has been translated with a temporal meaning.

(28) Tulil (Meng, 2018, p. 272)

məte	*nga-nəkən*	*idə*	*məgət*,
like	1SG.NPST-plant	3N	today

io	*avar*	***nə-pbo***	*avar*
then	again	**LOC**-DEM.UP	also

[15]I thank a student in the audience of the LingConLab (*Linguistic Convergence Laboratory*) seminar of the HSE Moscow on May 12, 2020, for suggesting that to me.

i-tu *be* *a-taem*
3N.NPST-grow at ART-time

a-vi *to* *ngə-məkən.*
3SG.M-PROXS SR 1SG.PST-plant
'Like I plant them today, then **next year** it will grow at the same time that I planted.'

Ma Manda (Finisterre–Huon), has a three-level contrast in elevation (DOWN/UP/LEVEL), in contrast to Tulil, which has only terms for UP and DOWN, co-expressed with distance such that we arrive at six items (Pennington, 2016, pp. 287–295). The demonstratives also express temporal meanings similar to Tulil, i.e., UP = FUTURE and DOWN = PAST, and the items with the LEVEL-meaning do not cover temporal functions. Moreover, Ma Manda speakers gesture upward and downward in accordance with the meaning of the demonstratives when they refer to future and past, respectively.

(29) Ma Manda (Finisterre–Huon) (Pennington, 2016, p. 294)
gulat *ban* *kum=slong* *laai* *ku-go-it*
year a DEM.DST.DOWN=ALL PN go-REMPST-1SG
'A year ago, I went to Lae.'

(30) *gulat* *kan=slong*
year DEM.PROX.UP=ALL
fentagût *naandû-maa-de-m*
all know-CMPL-IRR.DU-1NSG
'Next year we (two) will know it all.'

(31) *kun* *fafaan*
DEM.DST.UP who.have.already.died
'future ancestors.'

The temporal usage of the elevational demonstratives in the Towet dialect of the related language Nungon is identical to that of Ma Manda (Sarvasy, 2014, pp. 413–414).

In Iaai (Oceanic, Loyalty Islands) the deictic particle *jii* 'down (and toward the sea)' can express the meaning 'past' (32), and also serves to introduce relative clauses with past time reference. According to Ozanne-Rivierre (2004, p. 135), there are other Austronesian languages such as Taba with the same temporal extension DOWN = PAST.

(32) Iaai (Oceanic) (Ozanne-Rivierre, 2004, p. 135)
hnyi *bong* *e-jii* ...
in day LOC-DEM.DOWN
'the day before'

(33) *Haba* *jii* *me* *ogee* *haa* *kö* *u*
TOPIC DEM.DOWN COOR 1SG.ACC say to you
'I had told you before.'

There are three other languages in my sample that do not employ their elevational demonstratives with temporal meaning, but make use of the same or a very similar type of metaphor, namely Yupno (which belongs to the same language family as Tulil), Avar, and Lak.[16] Yupno speakers have been found to consistently use topographic (i.e., geocentrically anchored) gestures toward the ground for referring to the present, uphill for reference to the future and downhill for past (Núñez et al., 2012). The language has also one temporal expression employing a spatial metaphor *omo-ropmo bilak* (down.there.other.side year) 'a couple years ago, a few years ago.' In Avar, the adverbials *ʁorɁisa* 'last year' and *ťadejałːu* 'next year' originate from the adverbs *ʁorɁ* 'down(ward), under' and *ťade* 'up(ward),' respectively, and in Lak *jalunè'in* 'next year' is derived from *jalu(w)* 'up(ward).' Finally, in Tzeltal, which does not have elevational demonstratives, the topographic terms *-ajk'ol* 'uphill' and *-anil/alan* 'downhill' are also employed with the meanings 'later' and 'ahead of time, before.' Brown (2012, p. 10) analyzes those expressions as providing evidence for the metaphor 'time moves uphill' or 'the future is up(hill).'

I take the examples (27)–(33) as metaphors that map spatial expressions onto a temporal dimension: the future is located above or higher than the deictic center, and the past below. The metaphor can be explained by the direction of the biological growing process of upright human beings in the course of time. During the first years of their life human beings become taller as they get older, which means that if we compare one and the same person across time in the past the same person was smaller (=DOWN) whereas in the future s/he will be taller (=UP). The same applies to many other animals and plants with an upright position (e.g., trees).[17]

These findings are particularly interesting in view of the widely debated use of Mandarin Chinese spatial terms *shang* 'upper, up, over, above' for past events and *xia* 'lower, down, below, under, for future events, which show the opposite metaphorical extension (e.g., Yu, 1998, pp. 110–112; Boroditsky, 2001). Yu (1998, p. 111) argues that this conceptualization can be explained if one presupposes that on the horizontal plane the sagittal FRONT (or FURTHER) corresponds to EARLIER and BACK (or NEARER) to LATER. This metaphorical correspondence is said to result from the fact that if human beings moved by crawling on the ground their head would be in front and their feet would come last. The same applies to other animals that move with legs – the head is normally in front and turned into the direction of movement. Yu adds that in Western cultures family trees are arranged in a similar fashion: the oldest (earliest) generations are placed on the treetop and the last generation on the bottom. Radden (2003) hypothesizes that the cultural importance of the Yangtze River may have also played a role: the river flows downward and any objects moving on it would be located higher at an earlier period

[16]Don Killian (p.c.) drew my attention to yet another language from New Guinea that employs an elevational demonstrative for the expression of temporal meaning, namely Edolo. In this language, the phrase *salele elö alogogi* (week other over.there.LOC/week other up.across) refers to either last week or next week (Gossner, 1994, p. 87).

[17]Another possible explanation for the UP = FUTURE connection, which was suggested to me by Michael Daniel, is that it is mediated by the commonly found FRONT = FUTURE link. This equation takes into account the projection of the vertical axis onto the horizontal plane. This means that what is up is at the same time in front or further and that provides the link to future, i.e., UP = FRONT/FURTHER = FUTURE. However, I do not see how a similar equation can explain DOWN = PAST.

of the journey and lower at a later period [(see also Bender and Beller, 2014, p. 369), who call this the 'river model' of time]. Furthermore, gravitation might be seen as providing a 'natural direction' to the vertical dimension, which goes again from up downward (Bender and Beller, 2014, p. 349).

The spatial metaphors for the vertical dimension mentioned so far are not the only ones attested for elevational demonstratives in my sample. In Tidore, the elevational deictic verbs *ine* 'upward' and *tora* 'downward' are used in two temporal expressions, namely *mulamula ine* 'early morning, at sunrise' (morning + upward) and *lobino tora* 'early evening, shortly after the sunset prayer' (lit. night downward). In these expressions, the demonstratives most likely refer to the path of the sun with its apparent rising and setting. In Daga, there seems to exist a correlation such that FUTURE/PAST = UP because *yampoa utu-pa* (third up.there-out.of.sight) means 'next Wednesday' and *wataget utu-p* (before up.there-out.of.sight) means 'long ago' (Murane, 1974, pp. 101–102).

To sum up, temporal uses of elevational demonstratives show once more how the mapping from space to time differs across languages and cultures. It is important to keep in mind, however, that these verbal metaphors are not necessarily indications or proofs that speakers of those languages have a vertical mental time line.

Social Deixis, Evaluation, and Other Non-spatial Extensions

Perhaps surprisingly, it does not seem to be common to employ elevational demonstratives for the expression of social deixis, at least not in the languages surveyed for this study. So far, I encountered only two languages that are spoken in the Melanesia/West Papua area and have this type of semantic extension.

The first example comes from Tidore (North Halmahera), in which the elevational with the meaning UP is used to refer to locations and movements in the direction of the sultan's palace even though the palace is located rather low.[18] van Staden (2018) calls this usage 'royal up' and shows that in certain cases it includes *de facto* downward movement. Speakers showed some reluctance to use the 'royal up' when the referent was a dog because in the local Muslim culture dogs are not appreciated. She adds that there are other conventionalized usages that cannot be explained in terms of verticality or social deixis (e.g., Papua, which is located to the southeast of the island of Tidore, is referred to as UP because of sea currents and historical trading routes). The correlation between the UP-elevational and the conventional position of a powerful person represents an example of the metaphor CONTROL/POWER IS UP, for which cognitive evidence has been found by psychologists and psycholinguists (Schubert, 2005; Valenzuela and Soriano, 2009).

Bril (2004, p. 120) provides another example from Nêlêmwa-Nixumwak (Oceanic), where so-called 'directionals,' which are regularly added to deictic or anaphoric suffixes, which, in turn, are added to pronouns or determiners to form demonstratives, can be used for respectful reference to people of a higher social status. Sentence (34) is the only example that she cites for this use and it shows the elevational UP-directional *da* 'up' (without a preceding pronoun, deictic or anaphoric suffix). Bril further writes that it is generally improper to address others by name. Directional are used instead, e.g., *hey! the man up there*.

(34) Nêlêmwa-Nixumwak (Oceanic) (Bril, 2004, p. 120)
I *thovi* *da*
3SG ladle DEM.UP
'She serves him.' ('up' refers to the higher status of that person).

In Manambu, the noun phrases *a-da-wur du* (DEM.DST-M.SG-UP man) (Aikhenvald, 2008, p. 53) and *a-na-wur numa-də du* (DEM.DST-CURR.REL-UP big-M.SG man) (Aikhenvald, 2015) are used to refer to God (in addition to their literary sense '(big) man up there.'

Elevationals as Parts of Rich Demonstrative Systems

Demonstrative systems that encode elevation are, in general, already larger than the more common systems that express only (person-based) distance. A number of languages in my sample have not only elevational demonstratives but some more terms.[19] Other semantic distinctions with which elevational demonstratives are combined or are in complementary distribution in languages with rich demonstrative systems are

- Direction/movement: TOWARD vs. AWAY FROM[20]
  - Toward (Daga, Movima, and Lepcha).
  - Yonder/away (Ngiyambaa, Buru, Tanacross, Koyukon, and Movima).
  - Ahead (Tanacross and Koyukon).
  - Transverse (Nêlêmwa-Nixumwak).
- INWARD vs. OUTWARD (or INTERIOR vs. EXTERIOR)
  - Exterior (Jahai).
  - In/out (Central Alaskan Yupik and Eastern Canadian Inuktitut).
  - Out-of-field (Eastern Canadian Inuktitut).
  - Out in front (Dyirbal).
  - Outward (Manambu).
- Position (standing vs. non-standing) (Movima).
- Perception.
  - Invisible (Muna, Khasi, Baskeet, and Daga).
  - Visible (Daga).
  - Audible (Muna and Dyirbal).

[18] For some speakers even the governor's offices not very far from the sultan's palace conditions the same type of usage of the 'up' elevational.

[19] As one reviewer pointed out, it is an interesting question for future research to explore whether there is a hierarchy of co-expression of demonstrative semantics with these categories, and if there would be such a hierarchy, then we could examine where elevationals fall in.

[20] The meanings that fall under TOWARD/AWAY FROM and INWARD vs. OUTWARD are conceptionally related to the basic NEAR/FAR distinction, which demonstratives in a large number of languages make (Diessel, 2005). All three distinctions are radial concepts that refer to half-axes that are radiating out from a central point (Bender and Beller, 2014, p. 345).

- Other topographic meanings.
  - Seaward/landward (Iaai, Tidore, Tanacross, and Koyukon).
  - Off-river (Manambu).
- Temporary possession (Movima).
- Non-past vs. past (Movima).
- Referential or confidential (Dawro).

Most of the meanings are well-known from the literature on demonstratives (see, e.g., the lists by Diessel, 1999, p. 51; Dixon, 2003; Imai, 2003; Levinson, 2018, p. 35). Among the languages examined in this paper, Movima is particularly rich in demonstratives with unusual meanings such as 'temporary possession' or 'standing position' (Haude, 2006, pp. 177–186). Visibility has attracted some attention (Diessel, 1999, pp. 41–51; Dixon, 2003, pp. 90–91; Levinson, 2018, p. 30), but also terminological confusion (Breunesse, 2019, pp. 91–93). Levinson (2018, pp. 30, 37) suggests that in some languages invisibility might in fact better be analyzed as indirect evidentiality or simply audibility.

Areal Distribution of Elevational Demonstratives

The examined languages come from all around the world. As explained in the Section "Materials and Methods," the sample is a convenience sample, but based on a rather systematic and comprehensive survey of all areas of the world and more than half of the language families (Killian, unpublished). It is thus possible to suggest some generalizations concerning the areal distribution of elevational demonstratives.

There are clear areal hotspots in which there is a particular dense concentration of languages with elevational demonstratives. These areas are the New Guinea Highlands, the Himalayas, the Ethiopian Highlands and the Eastern Caucasus. Furthermore, a number of languages spoken on volcanic islands of Southeast Asia have elevational demonstratives. However, only on the island of New Guinea and immediately adjacent islands, in particular in the New Guinea Highlands, elevational demonstratives are found across a large range of different language families. In the Himalayas, only Sino-Tibetan languages have elevationals. If we consider the entire greater Hindu Kush Himalayan Region, we have to add some more Indo-Aryan languages. In the Caucasus, only East Caucasian languages, and in Ethiopian Highlands only some Omotic languages possess elevational demonstratives. By contrast, the mountainous areas of the Americas largely lack languages with elevational demonstratives with the exception of Cora and Pacaraos Quechua. The other American languages in my sample that have elevational demonstratives are spoken in rather flat areas (Movima in the Bolivian plains, Eskimo-Aleut and Na-Dené languages in Alaska and Greenland).

It has been hypothesized several times that there is a correlation between the presence of elevational demonstratives and the location of the speech community, more specifically, that the respective languages are spoken in hilly or mountainous areas (e.g., Imai, 2003, pp. 36, 38; Post, 2011, p. 152; Breunesse, 2019, p. 90; Ratliff, 2019). With respect to the languages of this paper, this claim is only partially confirmed. Five of the surveyed languages are spoken in lower hills (in general lower than 500 m above sea level), and seven languages on flat territory. All 38 remaining languages are spoken in mountainous locations mostly between 1,000 and 3,000 m (see **Supplementary Appendix Table A12** for more details). This proves Holton (2019) remark that "elevation does not require mountains." The definitions of the general elevational demonstratives given in (6) do not refer to salient landmarks. Only topographic elevational systems make a straightforward reference to mountains or hills, but as I stated above, most languages have general elevational systems and genuine topographic systems are rare. Even among the few languages which clearly have topographic elevationals, there are three languages not spoken in the mountains, but in lower hills (Dyirbal), on a flat island (Iaai) and in a flat area of Alaska (Tanacross). Holton (2019), who discusses the Eskimo-Aleut and Na-Dené languages spoken in the Artic, which is generally rather flat, notes that even though the Alaska territory includes some of the highest mountains in North America, the speakers of Na-Dené languages, which have elevational demonstrative, do not live in the mountains.

DISCUSSION

In this paper, I have examined elevational demonstratives, mainly focusing on their semantic and pragmatic properties. The main results of this study can be summarized as follows. The basic semantic values that elevational demonstratives encode can be ordered along a hierarchy (UP/DOWN > LEVEL/ACROSS) that reflects cross-linguistic tendencies in the frequency of the respective elevational values. Furthermore, the importance of the peripersonal sphere is linguistically reflected by elevational demonstratives because they predominantly co-express distance as opposed to proximity to the speaker. Another important finding of this study concerns the metaphorical extension of spatial elevational demonstrative meanings to the domain of time: the future is metaphorically located higher than the deictic center, and the past below. This metaphorical extension is the opposite of what has been found in Mandarin Chinese. I have proposed that the metaphor can be explained by the direction of the biological growing process of humans, many animals and plants.

In addition, I have also shown that elevational meaning *per se* is not deictic, because it does not depend on the speaker's (or addressee's) location, but simply relational and needs an anchor point, which can be a location that is independent of speaker or addressee. Items expressing elevational meaning can combine with deictics, in particular with demonstratives. If the combination is tight such that the items are synchronically monomorphemic, this leads to the deceptive impression that the elevational component is also deictic. Second, elevational demonstratives only rarely refer to geomorphic landmarks and they do not make use of an absolute frame of reference comparable to cardinal directions. They seem to be 'absolute' because normally gravity determines the direction and thus what is up and down, but the same is true for relational adverbials referring to the vertical axis.

Finally, I have argued that with respect to elevational demonstratives genealogical affiliation is more predictive than areal location. Languages with elevational demonstratives are found in flat, hilly, and mountainous regions, and they are a characteristic feature of a few language families worldwide (East Caucasian, Eskimo-Aleut, Sino-Tibetan, Timor-Alor-Pantar, Nuclear Trans New Guinea, and Omotic). New Guinea is the only area in which a wide range of languages with different genealogical affiliations that are spoken in mountain settlements have elevational demonstratives and thus geography or even language contact might have played a role in the development of those systems. In relation with that finding one possible direction for future research is to clarify whether the languages with elevational demonstratives, which were discussed in this paper, confirm the *Topographic Correspondence Hypothesis*. The latest version of the *Topographic Correspondence Hypothesis*, which is called *Sociotopographic Model*, states that languages spoken in similar topographic environments tend to have similar systems of absolute spatial reference, whereby social and cultural factors also play a role (Palmer et al., 2017). The hypothesis has been supported by data from atoll-based languages (Palmer, 2015; Palmer et al., 2017), and two languages spoken in the Hindu Kush mountain range (Heegård and Liljegren, 2018). What concerns the distribution of elevational demonstratives of the language sample used for this paper, they do not show evidence of topographical correspondence. First, there are many mountainous areas in the world without languages that have elevational demonstratives (e.g., almost all languages spoken in the American Cordillera, the Alps, the Great Dividing Range in Australia, the Atlas Mountains in North Africa, the slopes of the Great Escarpment in Southern Africa, and many more). Second, in two of the major mountain areas with elevationals, the elevationals are restricted to only one or two families. Except for East Caucasian none of the other language families spoken in the Caucasus has elevational demonstratives. In the Hindu Kush-Himalayas region, elevational demonstratives have been found so far only in Sino-Tibetan languages and a few Indo-Aryan languages (e.g., Palula, see Heegård and Liljegren, 2018 for more references). Therefore, my preliminary conclusion is to agree with Holton (2019) by suggesting that geography is less relevant than language structure and genealogy when it comes to elevational demonstratives. However, this hypothesis might obviously be rejected by new data and future studies.

There are many other open questions left for future studies of elevational demonstratives. In this paper, I have largely ignored the morphological and syntactic properties of elevational demonstratives as well as their use in discourse (e.g., as anaphors or cataphors). In order to be able to accomplish a detailed typological study we need more comprehensive descriptions of language-particular systems that are based on natural corpus data such that not only formal properties are covered but also the actual use and possibly frequency estimations can be detected. Another fruitful direction of research are various experimental approaches. The role of demonstratives in spatial cognition has been mainly investigated with respect to peripersonal space and distance as well as pointing, and the vast majority of controlled, experimental studies that I am familiar with examine languages with small demonstrative systems (English, Dutch, Italian, Spanish, Hungarian, Turkish, etc.). In the future, this line of research should be extended to languages with rich demonstrative systems such as the languages discussed in this paper.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The publication of this article was supported by the Thüringer Universitäts- und Landesbibliothek Jena.

ACKNOWLEDGMENTS

I am grateful to Don Kilian for helpful suggestions and discussions of elevational demonstratives and advice with respect to the data. I also want to thank two reviewers, my colleagues Gary Holton and Diana Schackow who answered my questions on their languages of expertise and the audience at the LingConLab (*Linguistic Convergence Laboratory*) seminar of the HSE Moscow on May 12, 2020, in particular Michael Daniel, for many questions and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01712/full#supplementary-material

REFERENCES

Adelaar, W. (2019). "Elevational deixis in Pacaraos Quechua," in *Paper Presented at the 52th Annual Meeting of the SLE*, Leipzig.

Aikhenvald, A. (2008). *The Manambu Language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Aikhenvald, A. (2015). Distance, direction, and relevance: how to choose and use a demonstrative in Manambu. *Anthropol. Linguist.* 57, 1–45. doi: 10.1353/anl.2015.0003

Amha, A. (2001). *The Maale Language*. Doctoral dissertation, University of Leiden, Leiden.

Amha, A. (2012). "Omotic," in *The Afroasiatic Languages*, eds Z. Frajzyngier and E. Shay (Cambridge: Cambridge University Press), 423–504.

Bender, A., and Beller, S. (2014). Mapping spatial frames of reference onto time: a review of theoretical accounts and empirical findings. *Cognition* 132, 342–382. doi: 10.1016/j.cognition.2014.03.016

Bickel, B. (1997). "Spatial operations in deixis, cognition, and culture: where to orient oneself in Belhare," in *Language and Conceptualization*, eds J. Nuyts and E. Pederson (Cambridge: Cambridge University Press), 46–83. doi: 10.1017/cbo9781139086677.003

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 42, 1–22. doi: 10.1006/cogp.2001.0748

Breunesse, M. (2019). *Demonstratives in Space and Discourse: A Synchronic and Diachronic Analysis*. Doctoral dissertation, University of Jena, Jena.

Bril, I. (2004). "Deixis in Nêlêmwa," in *Deixis and Demonstratives in Oceanic Languages*, ed. G. Senft (Canberra: Pacific Linguistics), 99–128.

Brown, P. (2012). Time and space in Tzeltal: is the future uphill? *Front. Psychol.* 3:212. doi: 10.3389/fpsyg.2012.00212

Bryant, F. J., Tversky, B., and Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *J. Mem. Lang.* 31, 74–98. doi: 10.1016/0749-596x(92)90006-j

Burenhult, N. (2008). Spatial coordinate systems in demonstrative meaning. *Linguist. Typol.* 12, 99–142.

Burenhult, N. (2018). "The Jahai multi-term demonstrative system: what's spatial about it?" in *Demonstratives in Cross-Linguistic Perspective*, eds S. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 361–380. doi: 10.1017/9781108333818.018

Carlson-Radvansky, L. A., and Irwin, D. F. (1993). Frames of reference in vision and language: where is above? *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-j

Casad, E. H. (1982). *Cora Locationals and Structured Imagery*. Ann Arbor, MI: UMI.

Casad, E. H. (1984). "Cora," in *Studies in Uto-Aztecan Grammar 4: Southern Uto-Aztecan Grammatical Sketches*, ed. R. W. Langacker (Dallas, TX: SIL), 152–459.

Casad, E. H. (1985). Vignettes of Cora locatives. *SIL Mexico Workpap.* 7, 101–117.

Cauchard, A. (2018). *Spatial Expression in CAAC: An Oceanic Language Spoken in the North of New Caledonia*. Berlin: de Gruyter.

Charachidzé, G. (1981). *Grammaire de la Langue Avar (Langue du Caucase Nord-Est)*. Paris: Farvard.

Clark, H. H. (1973). "Space, time, semantics, and the child," in *Cognitive Development and the Acquisition of Language*, ed. T. Moore (New York, NY: Academic Press), 27–63.

Cooperrider, K., Slotta, J., and Núñez, R. (2017). Uphill and downhill in a flat world: the conceptual topography of the Yupno house. *Cogn. Sci.* 41, 768–799. doi: 10.1111/cogs.12357

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2008). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Denny, J. P. (1982). Semantics of the Inuktitut (Eskimo) spatial deictics. *Int. J. Am. Linguist.* 48, 359–384. doi: 10.1086/465747

Diessel, H. (1999). *Demonstratives: Form, Function, and Grammaticalization*. Amsterdam: Benjamins.

Diessel, H. (2005). "Distance contrasts in demonstratives," in *World Atlas of Language Structures*, eds M. Haspelmath, M. Dryer, D. Gil, and B. Comrie (Oxford: Oxford University Press), 170–173.

Diessel, H. (2012). "Deixis and demonstratives," in *An International Handbook of Natural Language Meaning*, Vol. 3, eds C. Maienborn, K. von Heusinger, and P. Portner (Berlin: de Gruyter), 2407–2431.

Diessel, H. (2013). "Distance contrasts in demonstratives," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Dixon, R. M. W. (2003). Demonstratives: a cross-linguistic typology. *Stud. Lang.* 27, 61–112. doi: 10.1075/sl.27.1.04dix

Donaldson, T. (1980). *Ngiyambaa. The Language of the Wangaaybuwan*. Cambridge: Cambridge University Press.

Doornenbal, M. (2009). *A Grammar of Bantawa*. Utrecht: LOT.

Ebert, K. (1994). *The Structure of Kiranti Languages*. Zurich: ASAS.

Epps, P. (2008). *A Grammar of Hup*. Berlin: de Gruyter.

Fillmore, C. J. (1982). "Towards a descriptive framework for spatial deixis," in *Speech, Place, and Action*, eds R. J. Jarvella and W. Klein (Chichester: John Wiley), 31–59.

Forker, D. (2019). Elevation as a category of grammar: Sanzhi Dargwa and beyond. *Linguist. Typol.* 23, 59–106. doi: 10.1515/lingty-2019-0001

Fortescue, M. (1984). *West Greenlandic*. London: Croom Helm.

Friederici, A. D., and Levelt, W. J. M. (1990). Spatial reference in weightlessness: perceptual factors and mental representations. *Percept. Psychophys.* 47, 253–266. doi: 10.3758/bf03205000

Friedman, V. (1994). "Ga in Lak and the three "there-s": deixis and markedness in Daghestan," in *NSL 7: Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, ed. H. I. Aronson (Chicago, IL: Chicago University Press), 79–93.

Gossner, J. D. (1994). *Aspects of Edolo Grammar*. MA thesis, University of Texas at Arlington, Arlington, TX.

Gravelle, G. (2010). *A Grammar of Moskona: An East Bird's Head Language of West Papua, Indonesia*. Doctoral dissertation, Vrije Universiteit, Amsterdam.

Grimes, C. E. (1991). *The Buru Language of Eastern Indonesia*. Doctoral dissertation, Australian National University, Canberra.

Guérin, V. (2015). Demonstrative verbs: a typology of verbal manner deixis. *Linguist. Typol.* 19, 141–199.

Haiman, J. (1980). *HUA: A Papuan Language of the Eastern Highlands of New Guinea*. Amsterdam: Benjamins.

Haspelmath, M. (1997). *From Space to Time: Temporal Adverbials in the World's Languages*. Munich: Lincom.

Haude, K. (2006). *A Grammar of Movima*. Doctoral dissertation, Radbound University, Nijmegen.

Hayward, R. J. (1990). "Notes on the Zayse language," in *Omotic Language Studies*, ed. R. J. Hayward (London: SOAS), 210–355.

Heegård, J., and Liljegren, H. (2018). Geomorphic coding in Palula and Kalasha. *Acta Linguist. Hafniensia* 50, 129–160.

Heeschen, V. (1982). "Some systems of spatial deixis in Papuan languages," in *Here and There: Cross-Linguistic Studies on Deixis and Demonstration*, eds J. Weissenborn and W. Klein (Amsterdam: Benjamins), 81–110.

Heeschen, V. (1998). *An Ethnographic Grammar of the Eipo Language Spoken in the Central Mountains of Irian Jaya (West New Guinea), Indonesia*. Berlin: Dietrich Reimer Verlag.

Holton, G. (2000). *The Phonology and Morphology of the Tanacross Athabaskan Language*. Doctoral dissertation, University of California, Santa Barbara, CA.

Holton, G. (2018). "Place naming strategies in lower Tanana Dene," in *Language and Toponymy in Alaska and Beyond: Papers in Honor of James Kari*, eds G. Holton and T. F. Thornton (Honolulu, HI: University of Hawai'i Press), 121–138.

Holton, G. (2019). "Elevation and spatial orientation in Alaska," in *Paper Presented at the 52th Annual Meeting of the SLE*, Leipzig.

Huber, J. (2011). *A Grammar of Makalero: A Papuan Language of East Timor*. Utrecht: LOT.

Hyslop, C. (1993). *Towards a Typology of Spatial Deixis*. Bachelor's thesis, Australian National University, Canberra.

Hyslop, G. (2017). *A Grammar of Kurtöp*. Leiden: Brill.

Imai, S. (2003). *Spatial Deixis*. Doctoral dissertation, State University of New York, Buffalo, NY.

Kratchovil, F. (2007). *A Grammar of Abui: A Papuan Language of Alor*. Utrecht: LOT.

Leer, J. (1989). "Directional systems in Athapaskan and Na-Dene," in *Athapaskan Linguistics: Current Perspectives on a Language Family*, eds E. Cook and K. D. Rice (Berlin: de Gruyter), 575–622. doi: 10.1515/9783110852394-016

Levinson, S. C. (2003). "Deixis," in *The Handbook of Pragmatics*, ed. L. Horn (Oxford: Blackwell), 97–121.

Levinson, S. C. (2018). "Introduction: demonstratives: patterns in diversity," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781108333818.002

MacDonald, G. E. (1976). *Dadibi Grammar: Morpheme to Sentence*. Master's thesis, University of Papua New Guinea, Port Moresby.

MacDonald, L. (1990). *A Grammar of Tauya*. Berlin: de Gruyter.

Meng, C. (2018). *A Grammar of Tulil*. Doctoral dissertation, La Trobe University, Melbourne.

Miyaoka, O. (2012). *A Grammar of Central Alaskan Yupik: An Eskimo Language*. Berlin: de Gruyter.

Murane, E. (1974). *Daga grammar: From Morpheme to Discourse*. Norman: SIL.

Núñez, R., Cooperrider, K., Doan, D., and Wassmann, J. (2012). Contours of time: topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. *Cognition* 124, 25–35. doi: 10.1016/j.cognition.2012.03.007

Overall, S. (2007). *A Grammar of Aguaruna*. Doctoral dissertation, La Trobe University, Melbourne.

Ozanne-Rivierre, F. (2004). "Spatial deixis in Iaai (Loyalty islands)," in *Deixis and Demonstratives in Oceanic Languages*, ed. G. Senft (Canberra: Pacific Linguistics), 127–137.

Palmer, B. (2015). "Topography in language: absolute frame of reference and the topographic correspondence hypothesis," in *Language Structure and Environment: Social, Cultural, and Natural Factors*, eds R. De Busser and R. J. LaPolla (Amsterdam: Benjamins), 177–226. doi: 10.1075/clscc.6.08pal

Palmer, B., Lum, J. T. S., Schlossberg, J., and Gaby, A. R. (2017). How does the environment shape spatial language? Evidence for sociotopography. *Linguist. Typol.* 21, 457–491.

Pennington, R. (2016). *A Grammar of Ma Manda a Papuan Language of Morobe Province, Papua New Guinea*. Doctoral dissertation, James Cook University, Cairns.

Plaisier, H. (2007). *A Grammar of Lepcha*. Leiden: Brill.

Post, M. W. (2007). *A Grammar of Galo*. Doctoral dissertation, La Trobe University, Melbourne.

Post, M. W. (2011). "Topographical deixis and the Tani languages," in *North East Indian Linguistics*, Vol. 3, eds G. Hyslop, S. Morey, and M. W. Post (Delhi: Foundation Books), 137–154. doi: 10.1017/upo9788175968875.011

Post, M. W. (2017). "The inheritance, change and loss of topographical deixis in Trans-Himalayan languages," in *Paper Presented at the 12th Biennial Conference of the Association for Linguistics Typology* (Canberra: Australian National University). doi: 10.1017/upo9788175968875.011

Rabel, L. (1961). *Khasi, a Language of Assam*. Baton Rouge, LA: Louisiana State University Press.

Radden, G. (2003). The metaphor TIME AS SPACE across languages. *Z. Interkulturellen Fremdsprachenunterricht* 8, 226–239.

Ratliff, M. (2019). "On the development of elevation deictics in Hmong-Mien," in *Paper Presented at the 52th Annual Meeting of the SLE*, Leipzig.

Reesink, G. P. (1984). *Structures and Their Functions in Usan: A Papuan Language of Papua New Guinea*. Doctoral dissertation, University of Amsterdam, Amsterdam.

Reesink, G. P. (1999). *A Grammar of Hatam: Bird's Head Peninsula Irian Jaya*. Canberra: Australian National University.

Reesink, G. P. (2000). "Sougb," in *Studies in Irian Languages: Part II*, ed. G. P. Reesink (Jakarta: Universitas Katolik Indonesia Atma Jaya), 105–129.

Reesink, G. P. (2002). "A grammar sketch of Sougb," in *Languages of the Eastern Bird's Head*, ed. G. P. Reesink (Canberra: Australian National University), 181–275.

Sarvasy, H. S. (2014). *A Grammar of Nungon: A Papuan Language of the Morobe Province, Papua New Guinea*. Doctoral dissertation, James Cook University, Cairns.

Schackow, D. (2015). *A Grammar of Yakkha*. Berlin: Language Science Press.

Schapper, A. (2014). "Elevation in the spatial deictic systems of Alor-Pantar languages," in *The Alor- Pantar Languages: History and Typology*, ed. M. Klamer (Berlin: Language Science Press), 247–284.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *J. Pers. Soc. Psychol.* 89, 1–21. doi: 10.1037/0022-3514.89.1.1

Scott, G. (1978). *The Fore Language of Papua New Guinea*. Canberra: Australian National University.

Seyoum, M. (2008). *A Grammar of Dime*. Utrecht: LOT.

Steinhauer, H. (1990). "Demonstratives in the Blagar of Dolap," in *Papers in Papuan Linguistics*, ed. T. Dutton (Canberra: Pacific Linguistics), 179–223.

Steinhauer, H. (2014). "Blagar," in *Papuan Languages of Timor-Alor-Pantar: Sketch Grammars*, Vol. I, ed. A. Schapper (Berlin: de Gruyter), 115–172.

Traugott, E. C. (1978). "On the expression of spatio-temporal relations in language," in *Universals of Human Language*, Vol. 3, ed. J. H. Greenberg (Stanford, CA: Stanford University Press), 369–400.

Treis, Y. (2019). "The expression of elevation in demonstratives of the Omotic-Lowland East Cushitic contact zone," in *Paper Presented at the 52th Annual Meeting of the SLE*, Leipzig.

Valenzuela, J., and Soriano, C. (2009). "Is control really UP? A psycholinguistic exploration of a primary metaphor," in *Trends in Cognitive Linguistics: Theoretical and Applied Models*, eds J. Valenzuela, A. Rojo, and C. Soriano (Frankfurt: Peter Lang), 31–50.

van den Berg, R. (1997). "Spatial deixis in Muna (Sulawesi)," in *Studies in Austronesian and Papuan Languages*, ed. G. Senft (Oxford: Oxford University Press), 197–220.

van den Berg, R. (1989/2013). *A Grammar of the Muna Language*. Ukarumpa: SIL.

van Staden, M. (2018). "Tidore: non-contrastive demonstratives," in *Demonstratives in Cross- Linguistic Perspective*, eds S. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 343–360. doi: 10.1017/9781108333818.017

Verhees, S. (2019). "Demonstratives in Zilo Andi," in *Paper Presented at Linguistic Convergence Laboratory*, Moscow.

Whitby, C. (1990). *Dadibi, Tok Pisin, English, Po Dage Dabe*. Ukarumpa: SIL.

Willett, T. L. (1991). *A Reference Grammar of Southeastern Tepehuan*. Dallas, TX: SIL.

Woldemariam, H. (2001). Demonstratives in Dawuro. *Afrikanistische Arbeitspapiere* 65, 157–167.

Yarapea, A. M. (2006). *Morphosyntax of Kewapi*. Doctoral dissertation, Australian National University, Canberra.

Yu, N. (1998). *The Contemporary Theory of Metaphor: A Perspective from Chinese*. Amsterdam: Benjamins.

Zhirkov, L. I. (1955). *Lakskij Jazyk: Fonetika I Morfologija [The Lak Language: Phonetics and Morphology]*. Moscow: Izdatel'stvo Akademii Nauk SSSR.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a shared affiliation, though no other collaboration, with the author DF at the time of the review.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 01 September 2020
doi: 10.3389/fpsyg.2020.02016

Targeting in Language: Unifying Deixis and Anaphora

*Leonard Talmy**

Department of Linguistics and Center for Cognitive Science, University at Buffalo, The State University of New York, Buffalo, NY, United States

This article proposes that a single cognitive system underlies the two domains of linguistic reference traditionally termed anaphora and deixis. In anaphora, the referent is an element of the current discourse itself, whereas in deixis, the referent is outside the discourse in its spatiotemporal surroundings. This difference between the lexical and the physical has traditionally led to distinct theoretical treatments of such referents. We propose instead that language engages a single linguistic/cognitive system–"targeting"–to single out a referent, whether it is speech-internal or speech-external. To outline this system: As a speaker communicates with a hearer, her attention can come to be on something in the environment–her "target"–that she wants to refer to at a certain point in her discourse. This target can be located near or far in either the speech-external (deictic) or the speech-internal (anaphoric) environment. She thus needs the hearer to know what her intended target is and to have his attention on it jointly with her own at the relevant point in her discourse. The problem, though, is how to bring this about. Language solves this problem through targeting. First, at the intended point in her discourse, the speaker places a "trigger"–one out of a specialized set of mostly closed-class forms. English triggers include *this/these*, *that/those*, *here*, *there*, *yonder*, *now*, *then*, *therefore*, *thus*, *so*, *such*, *yay*, *the*, personal pronouns, relative pronouns, and tense markers. Next, on hearing the trigger, the hearer undertakes a particular three-stage procedure. In the first stage, he seeks all available "cues" to the target. Such cues belong to 10 distinct categories, representing 10 different sources of information about the target. In the second stage, he combines these cues so as to narrow down to the one intended target and rule out alternative candidates. In the third stage, he maps the concept of the target he has found back onto the original trigger for integration with the sentence's overall reference. This article is based on the overview portion of a book–*The Targeting System of Language*, MIT Press, 2018.

Keywords: deixis, anaphora, targeting, cues, trigger, demonstratives, pronouns, joint attention

OPEN ACCESS

Edited by:
Holger Diessel, Friedrich Schiller University Jena, Germany

Reviewed by:
Peter Gärdenfors, Lund University, Sweden
Kenny Coventry, University of East Anglia, United Kingdom

***Correspondence:**
Leonard Talmy
talmy@buffalo.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *27 February 2020*
Accepted: *20 July 2020*
Published: *01 September 2020*

Citation:
Talmy L (2020) Targeting in Language: Unifying Deixis and Anaphora. Front. Psychol. 11:2016. doi: 10.3389/fpsyg.2020.02016

INTRODUCTION

This study proposes that a single linguistic/cognitive system underlies two domains of linguistic reference, those traditionally termed anaphora and deixis. Broadly, an anaphoric referent is an element of the current discourse, whereas a deictic referent is outside the discourse in the spatiotemporal surroundings[1]. This is a distinction made between the lexical and the physical, one that has traditionally led to distinct theoretical treatments of the corresponding referents. Our proposal, on the contrary, is that language engages the same cognitive system to single out a referent whether it is speech-internal or speech-external. This single system, here named "targeting," can be outlined as follows.

As a speaker communicates with a hearer, her attention can come to be on something in the environment that she wants to refer to at a certain point in her discourse. This object of her attention will be called her *target*. Such a target can be located near or far in either the *speech-internal* or the *speech-external* environment–that is, in traditional terms, can be either anaphoric or deictic. To communicate about such a target, she needs the hearer to know what it is and to have his attention on it jointly with her own at the relevant point in her discourse. The problem, though, is how to bring this about. She cannot somehow directly reach into the hearer's cognition, take hold of his attention, and place it on her selected target at the intended moment.

A particular language-mediated process solves this problem. In this process, the speaker places a specialized lexical form at the relevant point in her discourse and, on hearing this, the hearer undertakes a specialized procedure. Her form is here called a *trigger* because it initiates, or "triggers," his procedure. Every language has a particular set of mostly closed-class triggers. The English set includes *this*/*these*, *that*/*those*, *here*, *there*, *yonder*, *now*, *then*, *therefore*, *thus*, *so*, *such*, *yay*, *the*, personal pronouns, relative pronouns, and tense markers. Such triggers are not simply static "placeholders," as some linguistic approaches view them, but in effect actively direct the hearer to undertake his procedure.

For its part, in turn, the hearer's procedure has three stages. In the first stage, the trigger directs the hearer to find certain elements of information to which he does have ready access. These elements of information function as *cues* to the speaker's intended target. Such cues have so far been found to belong to 10 distinct categories, representing 10 different sources of information. This first stage can thus be called the *trigger-to-cues* stage.

In the second stage, equipped with the cues that he has ascertained, the hearer uses them in combination to determine the speaker's intended target. The cues together thus guide him toward the target, which he could not have known directly. Generally, each cue rules in some candidates for target status while ruling out others. In association, the cues thus enable the hearer to narrow down to the target, singling it out from alternative candidates. This second stage of the procedure can accordingly be called the *cues-to-target* stage.

In the third stage, having determined the target, the hearer maps his concept of it back onto the trigger in the speaker's sentence. He relates this concept to the full conceptual content of the sentence in accord with the trigger's syntactic relation to the sentence. He thus has his attention on the target jointly with that of the speaker at the point of the discourse, and with the relationship to it, that she had intended. This stage can then be called the *target-back-to-trigger* stage.

This whole interaction rests on a coordination of the speaker's and the hearer's cognitive processing. As part of her cognitive processing, the speaker aims to get the hearer's attention jointly with her own on her target at a particular point in her discourse, selects the appropriate trigger to insert at that point, and ensures that cues in sufficient quantity and informativeness are available for the hearer to use to determine that target[2]. In turn, as part of his cognitive processing, the hearer perceives the trigger and, in consequence, carries out the three-stage procedure in which he finds the cues, determines the target, and integrates the concept of it back into the discourse, there to join his attention on it with that of the speaker.

This entire sequence–including the selection of a trigger, the three stages with their use of cues, and the cognitive processing of both speaker and hearer throughout–will be called *targeting*. Such targeting is understood as a linguistic/cognitive system that equally underlies both anaphora and deixis, in which they are unified as an essentially single phenomenon.

This targeting system is, then, the central topic of the present study. Its distinguishing features can be summarized as follows. Deixis and anaphora both rest on a trigger-initiated three-stage procedure–engaged in by a speaker and a hearer–in which the hearer finds cues, uses them to determine the speaker's intended target, and maps the concept of that target back onto the trigger and into its sentence. The cues to the target fall into 10 categories representing 10 different sources of information. This "targeting" process is a single linguistic and cognitive system in which deixis and anaphora are unified. The cognitive processing of both speaker and hearer in this targeting system can, in many respects, be inferred and built into the analysis.

The present analysis can be distinguished from others, first, with regard to the relation between anaphora and deixis. As Consten (2003) suggests, most approaches highlight the differences between the two domains or simply focus on one of them. For example, Mitkov (1999) and Ariel (2014) focus on anaphora, while Diessel (1999, 2013), Levinson (2003), and Chilton (2014) focus on deixis.

[1]In another system of terminology, one not prominent in the United States, "anaphora" and "deixis" refer instead to an attentional timeline, distinguishing, respectively, between whether the hearer's attention is already on the target when the speaker refers to it or is first brought to it by the speaker's utterance. Talmy (2018) extensively analyzes this attentional distinction and the properties of the joint attention that either precedes or ensues but does so using terms other than "anaphora" and "deixis."

[2]To ensure such *cue adequacy*, the speaker cannot be heedless as to which cues may happen to be available but must function as a proactive agent so as to enable the hearer to determine the target. Any cue inadequacy, as with communication deficiency generally, can be repaired, e.g., if the hearer indicates the need for more information and the speaker then supplies it.

To be sure, a few treatments have also highlighted the similarities or the commonalities between the domains. Authors with this approach include Bühler (1934), Peirce (1955) within semiotics, Silverstein (1976) and Hanks (2011) in their treatment of indexicality within linguistic anthropology, Consten himself (who sees a fuzzy boundary, parallelism, and coordination between the two domains), and Recanati (2005) within the tradition of language philosophy. Even Halliday and Hasan's (1976) labeling of anaphora and deixis, respectively, as "endophora" and "exophora"–with prefixes referring to "inside" and "outside"–suggests an awareness of the two domains' relatedness. However, this minority group has no counterpart of the explanatory system for unifying the two domains proposed in the targeting account.

Furthermore, the three-stage targeting procedure proposed here seemingly has no counterpart in other approaches. Specifically, no counterpart exists for the first stage, where a trigger directs a hearer to search for cues to a target–cues in 10 categories with distinct information sources. None exists for our second stage where such cues are all combined by the hearer in accord with certain governing principles to zero in on the target, and none exists for our third stage, where the hearer maps the concept of the target back onto the trigger for integration into its sentence.

The present article is based on the introductory sections of a book–*The Targeting System of Language* (Talmy, 2018; the MIT Press has kindly permitted this adaptation). Those sections serve there as an overview of an extensively laid-out framework. Standing alone, this article is also intended as an overview to the full framework. As an overview, it necessarily omits mention of numerous immediately relevant issues, but many of these are analyzed in detail in the book. Both the article and the book are set within the theoretical framework of "cognitive semantics" as put forth in Talmy (2000a,b, 2011).

As in traditional linguistics, the examples presented below and in the book as a whole are not observed but constructed, an option based on the following consideration. In science generally, two main methods might be cited for engaging with an area under examination, both of them valuable. In one, the researcher adopts the perspective of ecological validity, observing the naturally occurring patterns of interacting elements of the area within a context extended in both space and time. In the other method, the researcher controls the elements of the area and systematically manipulates them–for instance, holding other elements constant while varying one so as to investigate it in isolation. This method can reveal certain deep features of the area's organization and operation not readily accessible otherwise. It is this second method that is realized in our technique of constructing examples.

The second method is further realized by psycholinguistic experimentation. The beginnings of such experimentation have been reported, e.g., on deixis by Bangerter (2004); Coventry et al. (2008, 2014); Imai (2009), and Peeters et al. (2015) and on anaphora by McKoon and Ratcliff (1980) and Sanford and Garrod (1989). The book, with its detailed theoretical framework, serves as a call for much further experimentation.

SURVEY OF THE TEN CUE CATEGORIES

As mentioned, cues to a target can be analyzed as belonging to 10 distinct categories that represent 10 different sources of information about the target. These ten categories are outlined here. They can in turn be placed into five groups of two categories each. For simplicity in the survey, the categories are illustrated only with speech-external targets, but speech-internal targets are treated in section "Interaction of Compatible Cues to a Speech-Internal Target" below and, extensively so, in the book that this article is based on.

Each example in this survey includes one or more cues additional to the cue being illustrated. In fact, two or more cues are always needed in any given case for a hearer to determine a speaker's intended target. Such cues are usually in different categories but sometimes are in the same category.

Any two such cues will have one of three *concordance relations* to each other. In two of these relations, the cues are compatible. The cues then either *corroborate* each other, providing the same information about the target, or *complement* each other, providing different information about the target. In the present survey, the cues included in each example have one of these two compatible relations. In the third relation, two cues *conflict* with each other, providing incompatible information about the target, but such conflict typically initiates a constructive resolution in the hearer that again helps guide him to the target. Section "Interaction of Incompatible Cues to a Speech-External Target Broader Systems" illustrates this conflict relation.

The Lexical Cue Categories

In one group of two categories–the *lexical cue categories*–the cues to the target are provided by lexical forms in the speaker's utterance. In one category, the cues are provided by the trigger and, in the other, by forms around the trigger.

Core Cues

The trigger that a speaker includes in an utterance not only initiates the three-stage targeting procedure in the hearer but, in addition, is always lexicalized to provide cues to certain characteristics of the target. These are here called *core cues*. For example, a speaker, without using manual or ocular gestures, might say either (1a) or (1b) while opening the door to his lab to let a visitor peer inside, where a woman and several machines are located.

(1) a. She is new here.
b. These are new here.

If he says (1a), the trigger *she* provides the core cues that the target has the characteristics of being uniplex, an entity, animate, female, and third-person (i.e., not the speaker or the hearer). In surveying the lab, the hearer perceives that one part of its contents, the woman there, exhibits these five characteristics. These perceivable characteristics then function as targetive cues (see below). The hearer combines these two types of cues, the core cues and the targetive cues–which corroborate

each other–and settles on the woman as the speaker's probable intended target.

If the speaker instead says (1b), the trigger *these* provides the core cues that the target has the characteristics of being multiplex, entities, proximal[3], third-person, and–in the present construction–inanimate. In surveying the lab, the hearer will now likely select the machines as the speaker's intended target.

In the third stage of the targeting process for either sentence, the hearer, having determined the target, maps the concept of it back onto the trigger. In accord with the trigger's syntactic relation to the sentence[4], the hearer then integrates that concept into the overall conception expressed by the sentence. In the present cases, where the trigger is a subject nominal in construction with the predicate adjective *new*, the hearer ascribes the concept of "newness" to the concept of the target–that is, the woman or the machines. This third stage will not be described in the rest of the survey.

Co-form Cues

The linguistic constituents located around a trigger are here called its *co-forms*. A *co-form cue*, then, consists of any information provided by a co-form that helps the hearer determine the target of that trigger. The further a constituent is from a given trigger, the less likely it is to provide a co-form cue relevant to that trigger, and the less it would be regarded as a co-form of it.

To illustrate, suppose that a customer in a pet shop that has only one parrot among its animals goes up to the clerk behind the counter and, without gesturing manually or looking, says (2):

(2) That is the kind of parrot I like.

The trigger in the speaker's utterance, *that*, directs the hearer, that is, the clerk, to look for cues and use them to find a target. It also provides the core cues that that target has the properties of being uniplex, an entity, distal, and third-person. However, the hearer cannot narrow down to the target with these core cues alone because too many components in the scene have these properties. In the same utterance, the co-form *parrot* provides the co-form cues that the target has the properties of being a single entity with the identity of a parrot. These core and co-form cues corroborate each other in one respect–in indicating that the target is a unitary entity. They also complement each other, with the core cue indicating that the target's location is distal and the co-form cue indicating that its identity is that of a parrot. As hearer, the clerk will combine these cues to single out the one parrot in the shop as the speaker's intended target.

The Bodily Cue Categories

In another group of two categories–the *bodily cue categories*–the cues are provided by the body of one of the speech participants. Those of the gestural cue category consist of movements or configurations of parts of the speaker's body that she produces volitionally, while those of the corporal cue category consist simply of the location of the speaker's or the hearer's whole body.

Gestural Cues

Apart from the use of the mouth for speaking, any movement and/or configuration that a speaker volitionally produces with her body to communicate to a hearer is here considered a "gesture." A gesture that a speaker produces in association with a trigger specifically in order to provide a cue to a target is then called a *targeting gesture*. The cue that such a targeting gesture provides is a *gestural cue*.

To illustrate, suppose that a speaker says (3) to a guest standing beside her, while pointing toward one corner of a table across the room from them. That corner is clear in front, but a bottle of wine is standing about a foot back from its edge. (An exclamation point placed before a word here indicates heightened stress on that word).

(3) You can put your glass down right!-there.

The trigger *there* in the speaker's utterance not only alerts the hearer to find a particular target but also provides the core cues that this target is distal and is a location, not an entity. However, this cue by itself is not enough, given the multitude of distal locations in the situation. The speaker's gesture also provides a gestural cue to the target. By our analysis, this gesture leads the hearer to imagine an intangible line extending from the speaker's finger to the table's corner where the target is.

[3]As analyzed in Talmy (2018, Chapter 2), many triggers across languages indicate a target's *degree of remove* from the speaker (for a speech-external target) or from the trigger (for a speech-internal target). Such remove is not solely spatial but can be divided into five types: spatial, temporal, personal, social, and experiential; and the experiential type can be further divided into four subtypes: memorial, attentional, recognitional, and affective/perspectival. The number of different degrees that triggers distinguish varies by language and type of remove. But for spatial, temporal, and certain subtypes of experiential remove, English triggers distinguish only two degrees: *proximal* and *distal*. English triggers providing a core cue that a target is at a proximal degree of spatial remove (such as the trigger of the present example) include *this*/*these* and *here*, while those for a distal degree include *that*/*those* and *there*.

Such triggers, like closed-class forms generally [see Talmy (2000a), Chapter 1], express concepts whose character is "topological" rather than "Euclidean." In particular here, the proximal or the distal degree of a target's spatial remove is "magnitude-neutral." The degree of remove becomes more "magnitude-specific" only to the extent that other targeting cues (including elements of the context) narrow down the full possibility range. For example, other cues may set the distance between the two targets in (ia) and in (ib), respectively, at meters and parsecs, but the use of the proximal triggers *this*/*here* and the distal triggers *that*/*there* is unaffected by that difference. Their proximal/distal distinction only indicates that their target is, in effect, on the speaker-side or the non-speaker-side of a conceptual partition imaginally located midway between the two targets. These triggers are also neutral to the magnitude of their targets, which here–whether entities or locations–ranges from centimeters to megameters.

(i) a. This ball/The ball here is bigger than that ball/the ball there.

b. This planet/The planet here is bigger than that planet/the planet there.

[4]In some cases in some languages, the morphological form of the trigger itself helps determine its semantic relation and hence how to integrate its concept. For example, the English triggers *he*, *him*, and *his* in a portion of discourse might all target the same male entity but indicate that the concept of that entity is to function within that discourse as, say, an agent, patient, or possessor, respectively.

Such an intangible line is one instance of an elaborate system of *fictive chains*[5].

In combination, these core and gestural cues corroborate each other in indicating that the target is distal. They also complement each other in indicating, respectively, that the target is a location and that this location is situated where the imaginal line terminates at the table's corner. Integrating these cues, the hearer will select the surface of the table at the corner as the intended target–singling it out from other regions of space in the room. He will not select the bottle as the target–though it is equally included by the pointing gesture–because it is an entity, not a location.

The hearer is additionally guided toward this intended target by co-form cues from the phrase *put your glass down* and by the epistemic cues that they evoke (see below), namely, the knowledge that a glass is normally placed by resting it on a clear horizontal surface. This knowledge corroboratively rules in the table's surface and rules out the bottle as the target.

Suppose now that the speaker, while pointing as before, instead says (4):

(4) Could you please bring!-that over to me?

Here the trigger *that* provides the core cues that the target is uniplex, distal, third-person, and an entity, not a location. The gestural cue is the same as above. Combining these core and gestural cues, the hearer will now select the bottle of wine in the corner portion of the table as the intended target, singling it out from other entities in the room. He will not select the portion of the table's surface included by the gesture because it is a location, not an entity. The hearer is now additionally guided by co-form cues from the expression *bring over* and the epistemic cues they evoke: the knowledge that a person can bring only something he can readily hold. This knowledge here rules in the bottle and rules out the table's surface.

Laterally ambiguous gestures

In a comparable example, a gestural cue and a core cue again combine to provide complementary information about the target, thus enabling the hearer to single it out. Here ambiguity and its resolution are brought into the analysis. Suppose that a speaker in conversation with a companion wants to target a woman standing together with a man across a room with other women and men. The speaker points toward the couple while saying (5a):

(5) a. She is the director of our lab.
b. That is the director of our lab.

The gesture rules in the couple while ruling out the other people in the room, but the speaker is far enough away that it cannot indicate which member of the couple is intended. We will say that this gesture has *lateral ambiguity*. The trigger *she* provides the core cues that the target is uniplex, an entity, animate, female, and third-person. This cue then provides the additional information needed for the hearer to narrow down to the woman of the couple as the intended target. We will say that the core cue here enables the *lateral disambiguation* of the gestural cue. Note that if the speaker had said (5b), the trigger *that*–indicating only that the target is uniplex, an entity, distal, and third-person–would not provide enough information for such lateral disambiguation.

The process of disambiguation here can also be regarded as proceeding in the reverse order. Thus, the trigger *she* in (5a) rules in all the third-person women in the room as candidates for target status while ruling out all the men (as well as other entities), but it does not narrow this selection down to the woman in the couple. The gestural cue then provides the additional information needed to zero in on that particular woman.

Corporal Cues

The sheer presence of the speaker's–or, in certain cases, the hearer's–body at a particular location in space at the time of the speaker's utterance can serve as a *corporal cue* to the speaker's intended target. If the hearer is not already aware of this location, he must determine it perceptually or establish a mental image of it (as in a phone conversation) in order to make use of it as a cue. A corporal cue does not need to be accompanied by a gestural cue.

To illustrate, a woman in a booth at a fairground could reply as in (6) when asked by someone standing in front of her about the whereabouts of a certain man, Fred.

(6) Fred was here earlier.

In her utterance, the trigger *here* has its "corporal reading" (some languages have a distinct morpheme for this sense). Specifically, in addition to directing the hearer to find a target, it provides the core cue that that target is the spatial region surrounding the speaker's current location. The hearer combines this core cue with the corporal cue consisting of the speaker's actual location, which he perceives directly (or would imagine if on the phone with her). This cue combination allows him to select the region immediately around the woman out of all possible regions as the target. The speaker did not need to gesture–for example, by pointing to the ground in front of herself–but relied on the hearer's determining her bodily location. The utterance goes on to indicate that the man asked about was previously situated in the region now being characterized.

[5]A speaker's targeting gesture is always at a different spatial location than her target. The hearer must have a cognitive mechanism that associates the gesture with the target (unlike, say, a cat that at most just looks at the gesture). Talmy (2018), Chapter 5 proposes that the hearer forms this connection by generating a fictive chain: a succession of imaginal constructs–possibly from a relatively closed universal inventory–that are either schematic (largely geometric) structures or operations that affect such structures.

Such a fictive chain may have three properties of a physical mechanical system: (a) It is fully connected without gaps; (b) It forms progressively from the gesture to the target, not in place all at once nor from target to gesture; (c) It is causal: the gesture gives rise to the first fictive construct, the first construct to the second, etc.

In the present example, the pointing finger may be schematized as a straight line with a front point that coaxially emits a straight one-dimensional projection that progresses quickly through space to terminate at and intersect with the location on the table to mark it as the intended target.

The Collateral Cue Categories

When a speaker initiates a targeting event, the entities that she talks about and to–namely, the target and the hearer–can be regarded as categories generated by and collateral to the speaker. Cues to the target provided by these two collateral entities, then, belong to a group of two *collateral cue categories*. Cues provided specifically by the target and by the hearer are treated next in order.

Targetive Cues

A speaker's target generally exhibits characteristics that the hearer can discern, whether immediately or after a search. These characteristics can serve as cues for the hearer. A speech-external target in particular can produce sensory stimuli that provide the hearer with perceptual cues. These cues help guide the hearer toward the very entity producing them, which he can then single out as the target. These cues that the target itself provides will be called *targetive cues*. Two main types of such cues, the feature type and the salience type, are discussed next in order.

Targetive Feature Cues

Any intrinsic or contingent feature that a target exhibits, such as its own identity or its current distance away, can serve as a *targetive feature cue*. Its use was already seen in the "she is new" example in (1a). The trigger *she* in the speaker's utterance there provided the hearer with the core cues that the target had the features of being uniplex, an entity, animate, female, and third-person. In processing the visual scene before him, the hearer perceived one element with those same features–the woman in the lab. Taking these latter features as targetive feature cues then allowed him to combine them with the core cues–they corroborated each other–and to settle on that entity as the intended target.

A related example rests on a co-form cue instead of core cues. As they round his house onto an open field with a tractor, a horse, and a car spaced apart in the distance, a farmer says (7) to a visitor without gesturing or looking at the tractor:

(7) That is my tractor.

In addition to the usual core cues from the trigger *that* here, the co-form cue from the noun *tractor* ascribes to the target the feature of being a tractor in its identity. The physical tractor in the field, in turn, provides the visual stimulus of being a tractor in its identity. This is then the targetive feature cue. The two cues corroborate each other and help the hearer zero in on the tractor in the field–not, say, on the horse or the car–as the intended target.

Targetive Salience Cues

Where the features that a speaker's utterance ascribes to the target are insufficient for its determination, the hearer can instead search his perceptual environment for the most salient phenomenon within it, as judged on the basis of some 20 salience-associated parameters (Talmy, 2018; section 7.2.2). He entertains this phenomenon as a target candidate that exhibits *targetive salience cues*.

To illustrate, an experienced camper at a lake with a novice companion might, without gesturing, say (8) just after what seemed like a long plaintive sound could be heard:

(8) That is a loon.

The trigger *that* in the speaker's utterance directs the hearer to search for a target and provides the core cues that it is a uniplex distal third-person entity. The noun *loon* provides few co-form cues to the hearer, who is less familiar with the word. These cues together do not ascribe enough features to the target for the hearer to determine it. He instead performs a salience search of his environment.

The sound that the hearer has just heard has several forms of salience. It is unique in its surroundings, non-prototypical for its category, and unfamiliar to the hearer. This salience of the sound constitutes a targetive salience cue. This cue then tends to rule that sound in as a target candidate and to rule out other concurrent sounds or, for that matter, non-sonic phenomena without salience. Furthermore, the temporal nearness of the sound to the trigger's moment of occurrence provides a perichronal cue (see below) that tends to rule the sound in as a target candidate while ruling earlier-occurring sounds out. Combining these cues, the hearer is likely to settle on the long plaintive sound as the speaker's intended target[6].

Hearer-Focus Cues

A *hearer-focus cue* is a cue, metacognitively available to a hearer, indicating that her own current object of attention may be the speaker's intended target. In that case, she must also be sure that both the object and her attention on it are perceived by the speaker. To illustrate, a speaker who sees his friend looking fixedly at one particular car among others on the road might say (9) to her without himself gesturing:

(9) That is a Ferrari.

The trigger *that* in his utterance directs the hearer, his friend, to ascertain any available cues and use them to determine a target that he has in mind. The trigger itself provides her with the core cues that the target is a uniplex, distal, third-person entity. However, she does not find a gestural cue or a targetive cue from some especially salient object in the environment to help with her search. Among the additional cues she can check for is her own current focus of attention. She metacognitively notes that the object of her attentional focus is the car she is gazing at, and she is aware that the speaker can see both the car and her fixed look at it. In the absence of more compelling cues, she accepts the direction of her attention as a hearer-focus cue. She combines it with the core cues to settle on the car she is looking at as the target that the speaker aims to communicate about with his utterance.

The Background Cue Categories

In a still further group of two categories, the *background cue categories*, cues to the target arise from an extended field of

[6]The trigger *that* targets this sound, itself in a metonymic relation with the loon–a relation that would be literal if the speaker had said *That is a loon's call*.

phenomena–from the surrounding environment in one category and from the hearer's own cognitive infrastructure in the other.

Environmental Cues

The "environment" is everything that extends out from the speaker in the speech-external domain and from the trigger in the speech-internal domain. An *environmental cue* then is any information provided by a component of the environment that helps the hearer determine the target. In the two domains, respectively, such information consists of physical stimuli that the hearer can perceive and of syntactic properties that the hearer can discern. Environmental cues are here chiefly divided into ones that help a hearer either locate a target or bound it, addressed next in order. A secondary division that crosscuts the first rests on whether an environmental cue involves content or structure.

Environmental locating cues

Environmental locating cues come from aspects of content and structure within the total environment that guide the hearer in narrowing down to just a certain subenvironment that the target is located in. This reduction process thus limits the search space that the hearer must check through to find the target. We here sketch two variants of such reduction. As one of the variants, the subenvironment can be a continuous region that encompasses the target. To illustrate, a speaker on a farm might, without gesturing, say either (10a) or (10b) to a visitor:

(10) a. That Cessna in the field is Jane's.
b. That Cessna in the air is Jane's.

The core cue from the trigger *that* and the co-form cue from the noun *Cessna* direct the hearer to search for a uniplex, distal, third-person entity with the identity of a Cessna. The prepositional phrase in each utterance provides further co-form cues that help the hearer limit her search. Both phrases direct the hearer to attend perceptually to the surrounding environment and to abstract out certain aspects of its structure and content. For both phrases, in fact, this abstraction here includes the horizontal layer of space directly above and adjacent to the horizontal plane of the land. More specifically, the hearer knows from the phrase in (10a) that she can limit her search to the horizontal layer of space at her own eye level just above the land and can dispense with looking up or down. She knows from the phrase in (10b) that she can limit her search to the space overhead and omit looking through the space at eye level or below.

In the second variant, the subenvironment is not continuous but consists of a set of distinct elements, one of which will be the target. To illustrate, a speaker might, without gesturing manually or looking, say (11) to a hearer as they stand in a field with a number of cows and horses, where one of the latter is gray:

(11) That gray horse is Jane's.

The co-form cue from the noun *horse* may first lead the hearer to reduce her attention down to all the elements of the environment with the identity of a horse–together constituting the subenvironment–thus excluding the cows and other entities. Guided by another co-form cue from the adjective *gray*, the hearer needs then only look through that subenvironment–that is, through the set of horses–to find the gray one as the target. This succession of reductions constitutes a *nested search*. The hearer need not search directly through the entire environment for all occurrences of a gray color.

Environmental bounding cues

Environmental bounding cues are aspects of content and especially structure in the environment that help a hearer determine the outer boundary of what the speaker intends as his target. The hearer is generally guided to a particular set of such environmental aspects by cues of other categories.

To illustrate, as they stand atop a hill near a lagoon, a speaker might say (12) to a hearer while pointing at the middle of the lagoon:

(12) !-Mist forms there at night.

The trigger *there* in the speaker's utterance initiates the hearer's targeting procedure and provides the core cue that the target is a distal location. At the same time, the speaker's pointing gesture may lead the hearer to imagine an intangible line extending from the finger to one point at the lagoon's center. The hearer interprets this as a gestural cue to the target, but is that target to be the one point or some larger area around it? The hearer's general knowledge provides the epistemic cue (see next) that mist does not form at a single point but over some area; but then, what area? An environmental bounding cue provides this final information about the target. The hearer perceives that an area of roughly uniform appearance extends from the gesturally indicated point out to the lagoon's perimeter. This perimeter is a structural delineation within the environment. The hearer thus settles on the target as being not the spot pointed at but the entire surface of the lagoon as bounded by its outer perimeter.

Epistemic Cues

An *epistemic cue* is any information that a hearer derives from his own knowledge and beliefs that then helps him determine the speaker's intended target. Two main types of this cue category are knowledge about entities and knowledge about discourse, illustrated next in order.

Epistemic entity cues

To illustrate entity knowledge, after they get off a train, a speaker might say (13) to a companion beside her while pointing toward three people–two men and a woman–waiting for her in the station. One of the men looks substantially older and the other younger than the speaker.

(13) That is my father.

As before, the trigger *that* in the speaker's utterance initiates a targeting procedure in the hearer and provides the core cue that the target is a uniplex, distal, third-person entity. By itself, this cue does not much reduce the set of target candidates since there are many such entities in the scene. The gesture does narrow this set down to the three people in its scope. Since its distance away gives rise to lateral ambiguity, it does not indicate which of these three is the intended target.

In addition, though, the word *father* provides the co-form cue that the target is a man who has sired a child. The phrase *my*

father provides the further co-form cue that the targeted man has sired the speaker herself. The "man" component within the semantics of these co-form cues provides further complementary information that rules out the woman and narrows the target pool down to the two men of the trio, but neither of these co-form cues by itself distinguishes between the two men.

However, the word *father* also activates the conceptual category "father" in the hearer's knowledge store, which, besides other information, provides the epistemic cue that a father is older by some years than his child. The combination of this epistemic cue with the phrasal co-form cue indicates specifically that the targeted man is older than the speaker. The further combination of this result with the environmental content cues provided by the two men's appearances finally leads the hearer to rule out the younger-looking man and to conclude that the target is the man in the pair of men who looks older than the speaker.

Epistemic discourse cues

To illustrate discourse knowledge, two zoo visitors are in front of an enclosure with a giraffe and a straight-horned antelope standing close together, and one says (14) while pointing with lateral ambiguity at the pair of animals:

(14) That is an oryx.

In the hearer's knowledge of discourse management, a speaker would not state as new information something that the hearer would be expected to already know–here what a giraffe and the word for it are. Using this as an epistemic cue for lateral disambiguation, she concludes from the speaker's use of the unusual word *oryx* that the trigger *that* targets the animal in the pair other than the giraffe.

The Temporal Cue Categories

In this final group of two *temporal cue categories*, cues to the target arise from the temporal characteristics of elements present in an event of targeting. Cues of this sort from the trigger are in one category and cues from non-trigger elements are in the other.

Chronal Cues

The sheer occurrence of a trigger in a speaker's utterance at a particular location in time can serve as a *chronal cue* to the speaker's intended target. For certain triggers, such as English *now*, this target is itself an interval of time, one that extends through the moment of that trigger's occurrence. The speaker's utterance, furthermore, regularly identifies a particular state or event that occurs within or throughout this targeted interval. We can illustrate with a speaker saying (15) to a guest in her house:

(15) The bathroom is free now.

The trigger *now* in her utterance directs the hearer to determine the target, and it provides the core cue that this target is a temporal interval. It further indicates that this interval extends through her trigger's moment of occurrence. This latter indication, in turn, rests on what can be analyzed as an independent process, namely, determining that trigger's moment of occurrence. That is, the hearer must determine the chronal cue–the moment at which the speaker's trigger is uttered–and combine it with the core cue so as to center the interval around that trigger moment. As it happens, determining this chronal cue is straightforward, consisting simply of the hearer's taking cognizance of the moment at which he just heard the trigger.

In addition, the hearer's knowledge about bathroom use provides the epistemic cue that the length of the targeted interval should be reckoned in minutes–rather than, say, hours or, for that matter, decades, as would be the case for the interval targeted by the *now* in the sentence *We are in the age of the Internet now*[7].

The hearer concludes that the target is an interval of some minutes passing through and centered on the trigger he has just heard. He then temporally locates the state referred to by the utterance–the bathroom's availability–as occurring throughout this interval.

Perichronal Cues

A *perichronal cue* is any temporal property of an element other than the trigger that helps the hearer determine that trigger's target. In the majority case, though, perichronal cues do not help determine the target directly. They are rather the temporal properties of elements near a trigger that help determine which of those elements can serve as cues to its target, ruling some of them in and others out on the basis of their timing.

To illustrate, suppose that two joggers are running along the sidewalk past successively parked cars spaced amply apart, each car in turn to the left of them. At one point, one runner says (16a) to her companion while pointing leftward and, a few moments later, says (16b) while again pointing leftward:

(16) a. That is my car.
b. And that is my sister's car.

To examine the second communication, the trigger *that* in the speaker's (16b) utterance directs the hearer to find a particular target and provides the core cues that it is a uniplex, distal, third-person entity. In addition, the co-form cue from the word *car* tells him to look for a car as that target. Furthermore, the gestural cue from the speaker's second pointing movement and the targetive cue from the car appearing directly in view on the left both provide the perichronal cues that their occurrence is close enough in time to that of the trigger for them to be relevant.

By contrast, the previous (16a) pointing movement and the car it pointed at are ruled out as providing gestural and targetive cues relevant to the present (16b) trigger. Though the hearer might, in principle, entertain them as potential cues, their time of occurrence is too distant from that of the current trigger, so he concludes not to use them as indications of the speaker's currently intended target. Thus, through his observation of perichronal cues, the hearer takes into account only the concurrent gesture and the car immediately to his left as pertaining to the present event of targeting and disregards the earlier gesture and other cars along the curb.

[7]This trigger *now* also exhibits magnitude neutrality (see text Footnote 3)–but here with respect to the size of the temporal interval it targets.

INTERACTION OF COMPATIBLE CUES TO A SPEECH-EXTERNAL TARGET

While the preceding survey focused on one cue category at a time, each example in the following sections includes numerous cues of different categories to allow attention to their interactions as well as to the hearer's processing of them.

Our account of this processing, it must be emphasized, is in the form of a *regularized description*–consisting of a succession of discrete steps–both for clarity of presentation and to reflect logical relations. Such a description is not based on assumptions about any actual operations in cognitive processing, which may well occur in parallel or in other sequences as well as being more gradient than discrete and which must be determined experimentally. We do posit, though, that the elements and the relationships presented in the description are *somehow* represented in cognition since any absence among them would disrupt targeting. The detailedness of the description, furthermore, may help identify the numerous potential points of articulation in cognition relevant to such processing. Their presence in cognition is indeed posited on the assumption that little in cognition just happens by itself, no matter how quickly and unconsciously a hearer's processing may proceed or how self-evident the result may seem.

The examples in the present and the next section have only compatible cues. The cues in these two examples are, respectively, to a speech-external and to a speech-internal target, but the example in section "Interaction of Incompatible Cues to a Speech-External Target Broader Systems" includes conflicting cues.

In the present illustration involving a speech-external target, a couple walking along a sidewalk stop a foot in front of a gift shop window and look in. The speaker knows that her companion had wanted to learn what the color puce looks like among other colors that he was unclear about, such as vermilion and chartreuse. The speaker spots puce coloring among the gift items on display and–wanting to direct the hearer's attention to it–says (17) while facing toward the interior and pointing:

(17) Those boxes are puce-colored.

In the display's setup, a platform extends back behind the shop window. A single cluster of gift items appears in the front portion of the platform, while three separate clusters are arrayed left to right along the rear portion. The middle cluster in the rear includes some boxes that are red, some boxes that are of a hue which is unknown to the hearer–hue number 1, a single box with unknown hue number 2, and some statuettes of unknown hue number 3. The front cluster has some boxes of unknown hue number 4. While saying (17), the speaker points toward the rear middle cluster, but her gesture is laterally ambiguous, unable to single out specifically what within the cluster is puce-colored.

The trigger *those* in the speaker's utterance sets the hearer off on a three-stage targeting procedure. In addition to other information, the trigger provides the core cue that the target is distal. This core cue rules out any proximal entities from being the target. However, the hearer cannot use this cue by itself because the distal/proximal distinction that it entails is topological. It could distinguish between entities separated by inches as readily as by miles. It must be combined with information from another cue, an operation described below.

At the same time, the orientation of the speaker's head and body provides a gestural cue, namely, that the target is situated within the corridor of space extending forward from her front. This gestural cue thus rules out all entities located outside this corridor as candidates for target status.

The platform that the hearer perceives nearby within this corridor of space can then provisionally suggest an environmental locating cue. This cue is that the perimeter defining the platform's expanse is also the boundary of the region in which the target is located. If confirmed, this cue would then eliminate any other regions in the corridor as areas in which the target might be found.

The hearer can now combine this environmental locating cue with the earlier-mentioned core cue that the target is distal. Now anchored on the platform, that cue loses its topological relativity and rules in the rear portion of the platform while ruling out the front portion. This fact also eliminates the possibility that the target is in the cluster of gift items located in that front portion. Accordingly, puce cannot be the unknown hue 4 of the boxes in that cluster.

A second gestural cue–the one provided by the speaker's pointing finger–is not precise enough to pinpoint the target, but it is precise enough for certain other indications. First, it corroborates the environmental locating cue that the target is situated within the perimeter of the platform, thus confirming a cue that, in this regularized description, was previously only provisional. Second, it corroborates the just-seen indication that the target is located in the rear of the platform. This indication itself had been derived by combining the environmental locating cue with the core cue. Third, it provides the new indication that the target is located in the region of the middle cluster out of the three clusters along the platform's rear. This gestural cue thus rules out both side clusters from consideration.

In addition to its core cue that the target is distal, the trigger *those* provides the core cue that the target has an "entity" character. This cue thus rules out the possibility that the target is a location, among other non-entity-like options. If it had been viable, such a location, in accordance with the preceding cues, could have been the volume of space occupied by the middle cluster or the portion of surface it rests on. What this additional core cue does rule in as target candidates, then, is either a particular physical object or objects within the cluster or the full ensemble of the cluster.

A third core cue provided by the trigger *those* is that the target is third-person, which excludes the speaker and the hearer as target candidates. This core cue thus corroborates the same exclusion indicated by the two gestural cues, which located the target where the speaker pointed at within the corridor defined by her bodily orientation.

Additional information next comes from the morpheme *box* that the speaker uses in her utterance. It provides the co-form cue that the target has a "box" identity. At the same time,

through the visual stimuli they produce, the statuettes in the middle cluster provide the environmental content cue that they have a "statuette" identity, while the remaining items in that cluster provide the environmental content cue and potentially targetive cue of having a "box" identity. In conjunction with these additional cues, then, the co-form cue rules out the statuettes in the cluster as candidates for target status–thus eliminating unknown hue number 3 as puce–but rules in the remaining items. This co-form cue also corroborates the preceding two core cues in the elimination of the collocutors and of any locations from target candidacy.

An additional co-form cue comes from the plural morpheme *-es* on *box*, which indicates that the target is multiplex. At the same time, the trigger *those*–in addition to its core cues that the target is distal, entity-like, and third-person–provides the core cue that the target is multiplex. This final core cue and the co-form cue from the plural morpheme *-es* thus corroborate each other in their indication that the target is multiplex. This redoubled indication thus rules out the single box of unknown hue number 2 as a candidate for the target, but they still leave the red boxes and the boxes of unknown hue number 1 as candidates.

What then distinguishes between these two candidates is an epistemic cue from the hearer's knowledge of discourse principles. He knows that the speaker, following an informativeness principle, would not present as new information something that her addressee would be assumed to know already, such as what red is. On the basis of this epistemic cue, the hearer reasons that the speaker could not have been informing him about the red boxes and hence must have been referring to the last remaining target candidate, the boxes of unknown hue number 1.

The hearer still has more cues to note and narrowing down to do. The present tense of the verb *are* can be regarded as a second trigger in the speaker's sentence. It provides the chronal cue that the target occupies its location during an interval that extends through the moment of the trigger's utterance–that is, the current moment. The target thus does not occupy its location during an interval wholly before or after this current moment. This cue thus eliminates from potential target status any puce-colored boxes that may have been present in the past or might be present in the future at the indicated location. If the speaker had intended such a target in, say, the past, she would instead have said something like (18a or b). With such temporally displaced boxes ruled out, the presently appearing boxes of unknown hue number 1 continue to be ruled in.

(18) a. Those boxes were puce-colored (yesterday).
b. Some boxes there (yesterday) were puce-colored.

Finally, suppose that the couple had also been stopping at shop windows at other locations along the street, pointing at and commenting on items in them. Some factor in the hearer's cognition must be present that leads him to deal only with the gestural and the targetive cues concurrent with the present trigger and associate these all with each other rather than to use cues from the recent past now in memory. The concurrentness of the present gestural and targetive cues is a perichronal cue that rules them in, while the non-concurrentness of the previous cues is a perichronal cue that rules them out.

In sum so far, the hearer is prompted into a targeting procedure by the trigger *those* in the speaker's utterance. In the first stage of this procedure, he discerns some dozen-specific cues from eight different categories. In the second stage, he integrates these cues to the point where they enable him to narrow down to the speech-external target evidently intended by the speaker. This target turns out to be the boxes of unknown hue number 1. The narrowing-down process has ruled out any other items currently visible in the window display, any items that were or will be in that display, any items that were or will be seen in other displays along the way, and, generally, any non-items or items outside the display.

In the third stage of the procedure, the hearer next maps the concept of the now-identified target back onto the trigger *those*. In accord with the trigger's syntactic relation to the sentence–namely, as the determiner of a subject nominal within a construction of predicate–adjective attribution–he integrates that concept into the sentence's overall conception. As a final result, he concludes that the boxes he has perceptually narrowed down to, tinted with one of the hues he had not known, are in fact puce-colored.

INTERACTION OF COMPATIBLE CUES TO A SPEECH-INTERNAL TARGET

Shifting focus from a speech-external to a speech-internal target, we provide an illustration in which a man and a woman are alone in a room, and he says the two consecutive sentences in (19) to her.

(19) a. My sister led her mare down the hill toward some cowboys.
b. She was dappled.

The trigger *she* in the speaker's (19b) utterance directs the hearer to undertake a three-stage targeting procedure. It also provides the core cues that the target is uniplex, an entity, animate, female, and third-person. Although the hearer herself has the first four of these characteristics, she lacks the fifth. Relative to the speaker, the hearer is not third-personal but second-personal and, if the speaker had intended her as the target, he would have instead used the trigger *you*. Accordingly, the core cue that the target is third-person rules out the hearer herself as a possible target but rules in other female animate entities as candidates for target status.

The hearer might accordingly look for the speaker's intended target in her physical surroundings. From her perception of those surroundings, environmental content cues arise with the information that no such female beings are present there. These environmental cues thus suggest ruling out the speech-external environment as the target's location.

This suggested exclusion may then be corroborated in the hearer's cognitive processing by the state of the cues from three further categories. First, there is an absence of gestural cues–the speaker does not produce targeting gestures as he speaks. This

suggests that, in the immediate physical surroundings, the target is not present to be gestured at. Second, in full corroboration with the environmental content cues, there is an absence of targetive cues provided by perceptual stimuli from another woman. This again suggests that the target is not present in the immediate physical surroundings. Third, the tense of the verb *was* in (19b)–which can be regarded as a second trigger in that sentence–provides the core and the chronal cues that the target's time of occurrence is not that of the trigger itself, that is, the present. If it had been, it too might have indicated a target in the current immediate physical surroundings. These multiple indications to rule out the speech-external environment as the target's location then strongly suggest ruling in the speech-internal environment as its location.

To find the target, the hearer thus directs her attention to the speech-internal environment. This consists of the discourse–both its formal and its semantic aspects–that the trigger *she* occurs in. The perichronal cues from this discourse increasingly rule out portions of it the further they are in time from the trigger's occurrence and increasingly rule in portions of it the closer they are. These perichronal cues may finally narrow the location of the target down to the utterances in (19) themselves and eliminate utterances outside them.

The hearer may next consider environmental content and structure cues present in the discourse surrounding the trigger (rather than considering them in her physical surroundings, as she had done earlier). These cues consist of information about the formal and the semantic components of the utterances that have been ruled in through perichronal cues and indicate that the target is likely to be one of those components.

To consider the formal components of an utterance first, they generally consist of its morpheme, word, and phrase constituents as well as the grammatical relationships that these bear to each other. The formal components of (19a) include four noun phrases, a verb, and two prepositions as well as their contained and containing constituents and all the grammatical interrelationships present. Those in (19b) include a trigger, a verb, and an adjective as well as their grammatical relationships.

For their part, the semantic components of an utterance generally consist of the meanings of the formal components and their relationships as well as of their pragmatic implications. The semantic components of (19a) include the speaker's sister, her mare, a hill, some cowboys, an act of leading, and a path of descent and approach, among other indications and relationships. The semantic components of (19b) include a quality of dappledness and, from the trigger *she*, a directive to find a target together with core cues to that target.

These core cues can now interact with the remaining environmental content cues. A first result comes from the core cue that the target is animate. This cue actually eliminates all the formal components of both utterances from the possibility that the target is one of them. The reason is simply that formal linguistic components are never animate. This elimination leaves only the semantic components within the two utterances as candidates for target status.

But then, this same core cue that the target is animate further eliminates some of these semantic components as well. To begin with, it eliminates all the semantic components of (19b) since none of them–e.g., neither the dappledness nor the trigger's directive or core cues–is animate. This core cue eliminates itself as well because it is, in fact, not itself an animate but rather part of a directive to find an animate.

These exclusions then leave only the semantic components of (19a) as contenders for target status. Now another core cue provided by the trigger *she*–that the target has an "entity" character–rules out such semantic components as the act of leading and the downward approaching path. It rules in only four semantic components: the speaker's sister, her mare, a hill, and the cowboys.

With respect to these four ruled-in components, the core cue that the target is animate again comes into play to rule out the hill. At the same time, the core cues that the target is uniplex and female both rule out the cowboys. Among the semantic components of the first utterance, then, the core cues together rule in the speaker's sister and her mare as candidates for target status but rule out the rest.

The hearer's ability to select a single target from these two remaining candidates is furthered by information from the co-form *dappled*. This adjective provides the co-form cue that the target has the property of being "dappled". This adjective has two main meanings: one involves spots of different shades intrinsically present on the skin or fur of a non-human animal; the other involves spots of light being reflected off of any surface. If the first meaning is in effect, the co-form cue from the adjective is enough to finally narrow the selection down to the mare because the mare is a non-human animal while the sister is human; but if the second meaning is in effect, the co-form cue does not distinguish between the two remaining candidates since they both present surfaces.

An epistemic cue then finally enables the hearer to zero in on the target. This cue is the hearer's linguistic knowledge that, for the second meaning of the adjective *dappled* to be in effect, the adjective must be accompanied by a phrase referring to light or shade. This meaning would be evoked, for example, in a sentence like that in (20).

(20) My sister was dappled in the sunlight (that filtered through the leaves of the trees).

However, the utterance in (19b) did not include such a phrase. Accordingly, the hearer concludes that only the first meaning of *dappled* can be in effect. Thus, the hearer finally settles on the mare as the speaker's intended target.

In sum to this point, the trigger *she* in the speaker's second utterance initiates a targeting procedure in the hearer. In the first stage of the procedure, the hearer discerns over a dozen either negative or positive cues–that is, cues that are missing or that are present with specific content–from seven different categories. In the second stage, her processing of these cues enables her to narrow down to the speech-internal target intended by the speaker. This target turns out to be the mare referred to in the speaker's first utterance. This narrowing-down process has ruled out everything in the current speech-external environment

and, within the speech-internal environment, all the formal components and all but one of the semantic components.

In the third stage of the procedure, the hearer next maps the concept of the now-identified target back onto the trigger *she*. Since, as a semantic component, the target "mare" is already a concept, the hearer simply maps this concept–or a copy of it–onto the trigger. In accord with the trigger's syntactic relation to the sentence–namely, as a subject nominal within a construction of predicate-adjective attribution–she integrates that concept into the sentence's overall conception. She concludes that the mare in the discourse just referred to by the speaker has naturally dappled skin.

INTERACTION OF INCOMPATIBLE CUES TO A SPEECH-EXTERNAL TARGET

In the preceding two illustrations, the cues were all compatible, but here some are incompatible, providing conflicting information about the target (itself again speech-external). This incompatibility is a well-formed feature of the speaker's production, designed to initiate cognitive processing in the hearer that resolves the conflict so he can proceed to determine the target.

In our illustration–sketched here but detailed in the book–a woman sits across a restaurant table from a man. For two initial control examples with unconflicted cues, she either says (21a) and points to the right side of her own mouth or says (21b) and looks at his mouth while extending her finger across the table to point directly at it:

(21) a. I have got something in my teeth right here.
b. You have got something in your teeth right there.

In the conflicted example in (22)–which has the same import as (21b)–she again looks at his mouth but, as in (21a), points to a spot on the right side of her own teeth.

(22) You have got something in your teeth right here.

In (22), the trigger *here* directs the hearer to determine the speaker's intended target and, to that end, to determine the available cues to it, including the following five. The trigger itself provides the core cue that the target is a location proximal to the speaker. An environmental locating cue from the perceivable surroundings indicates that the speaker's body is the setting for this targeted location. One of the speaker's two gestural cues, the manual one from her finger, indicates that the targeted location is a spot on the right side of her teeth. Another gestural cue, the ocular one from her gaze, indicates that the targeted location is at the hearer's mouth. The phrase *in your teeth* provides the co-form cue that the targeted location is at the hearer's teeth.

The hearer's processing of these cues he has assembled–again using a regularized description–begins with an assessment phase. Within this phase, an initial operation of *consistency checking* examines the cues for their mutual compatibility. It proceeds on the basis of a certain set of principles, including one of plausibility. Here taking all the cues at face value can lead to an implausible conception, such as one in which some of the hearer's teeth are in the speaker's mouth. He thus concludes that some of the cues in fact are in conflict.

A second operation of *clustering* within the assessment phase segregates the cues into groups that are each internally compatible but that are incompatible with each other. Here the co-form cue and the ocular gestural cue are compatible with each other in one group, both indicating that the targeted location is at the hearer. Incompatible with this first group is a second group that includes the core cue, the manual gestural cue, and the environmental cue–all compatibly indicating that the targeted location is at the speaker.

A third operation of *evaluation* within the assessment phase assigns opposite states of validity to the two incompatible groups mainly on the basis of a *problem-avoidance principle*. Control example (21a) lacks problems, but control example (21b) has some: her reaching gesture might be considered as socially inappropriate, physically awkward, or incapable of precision. By contrast, the conflicted example in (22), in which the speaker touches her own teeth, avoids such problems: it evades social stigma, is easier to perform, and permits precision (the hearer can now use his vision to learn the exact location).

The hearer concludes that the speaker has resorted to the conflicted utterance to avoid the problems of the direct communication in (21b) and hence that its import equates to that communication rather than to the unproblematic direct one in (21a). By the evaluation operation, then, the group of two cues that the conflicted communication shares with the direct one in (21b)–the co-form and ocular gestural cues–is assessed as *valid*, while the other group of three cues is assessed as *anomalous*.

The hearer's processing next proceeds to a resolution phase. This phase retains the valid cues, that is, the co-form cue and the ocular gestural cue; but it adjusts the anomalous cues–that is, the core, environmental, and manual gestural cues–so that they become compatible with the valid cues. The full coherent set of cues that results can then lead to the intended target.

The main operation in this resolution phase is that of *mapping*. This operation acts on the location at the teeth in the speaker's body that is seemingly targeted by the anomalous cues–the *initial target*. Through this mapping, the hearer imaginally translates that location to the structurally homologous location on his own body–the *final target*. *Rotational* mapping targets a location on the right side of his teeth, while *reflective* mapping targets a location on the left side.

BROADER SYSTEMS

We conclude by observing that the targeting phenomena presented in this overview are generally part of broader systems significant in their own right. We here cite some of these, all examined in Talmy (2018), as noted below.

First, the joint attention on a target that a speaker aims to achieve with a hearer is part of a broader system of *common attention* (Section 13.3, "Taxonomy of Common Attention"). Common attention is characterized by different combinations of settings along four parameters: participation, recognition, elicitation, and epistemic parameters. Although "joint attention" seems to be the prototype in communication and is the term usually found in the literature, in our analysis, it is not an elementary phenomenon but the most elaborated endpoint in a hierarchy of attentional patterns. It is a conjunction of the highest settings on all four parameters: elicited mutually recognized common attention based on observation.

Next, the pointing gesture of earlier illustrations is merely the prototype within a prodigiously extended system of targeting gestures (Chapter 5). In this system, the fictive chain imaginally generated by a gesture can alternatively intersect with, enclose, pervade, coprogress with, parallel, access, "behold," neighbor, or contact a target at different levels of precision.

Furthermore, the system of fictive chains is part of a more general cognitive system of *spatial fictivity* (Section 5.15, "A Cognitive System of Spatial Fictivity"). This system is often engaged within visual perception, linguistic representation, and cultural constructs.

In addition, the phenomenon of conflict and its resolution was seen engaged with respect to cues to a target in section "Interaction of Incompatible Cues to a Speech-External Target Broader Systems" above. It is part of a much broader cognitive system. When at work in language, this system enlists *constructive discrepancy* and underlies all tropes, including that of metaphor (Section 14.4.1, "Conflict and Resolution in Tropes").

Targeting can also be, in part, incorporated into other preexisting theories, but then it requires their expansion (Section 1.8.2, Specific Approach Comparisons). For example, triggers, as a linguistic category, constitute a certain type of construction within the broader theory of construction grammar, as articulated, for example, by Fillmore et al. (1988) and Goldberg (1995, 2006); but the trigger construction has novel properties whose inclusion necessitates the theory's extension or again, indexicality theory as articulated, e.g., by Silverstein (1976, 2003) includes an index and its object, corresponding to our trigger and target. Where their index simply "stands for," "represents," "points to," or "indexes" its object, targeting inserts an entire stratum between index and object, a stratum consisting of cognitive processes that ascertain cues and integrate them so as to narrow down to that object/target.

It can thus be seen that, through its own properties and their generalizations like those above, targeting is a significant area for exploration within cognitive science because of the window it opens onto the nature of cognition overall.

Note: All the author's publications except the 2018 book are available on his website: https://www.acsu.buffalo.edu/~talmy/talmy.html.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Ariel, M. (2014). *Accessing Noun-Phrase Antecedents. RLE Linguistics B: Grammar*. London: Routledge.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi: 10.1111/j.0956-7976.2004.00694.x

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.

Chilton, P. (2014). *Language, Space, and Mind: The Conceptual Geometry of Linguistic Meaning*. Cambridge: Cambridge University Press.

Consten, M. (2003). "Towards a unified model of domain-bound reference," in *Deictic Conceptualization of Space, Time, and Person*, ed. F. Lenz (Amsterdam: John Benjamins Publishing Company), 223–248. doi: 10.1075/pbns.112.12con

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition* 75, 209–235.

Diessel, H. (1999). *Demonstratives: Form, Function, and Grammaticalization*. Amsterdam: John Benjamins Publishing Company.

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Lang. Cogn.* 5, 239–249. doi: 10.1515/langcog-2013-0017

Fillmore, C., Kay, P., and O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64, 501–538.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Halliday, M., and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hanks, W. F. (2011). "Deixis and indexicality," in *Foundations of Pragmatics*, eds W. Bublitz and N. Norrick (Berlin: Mouton de Gruyter), 313–346.

Imai, S. (2009). *Spatial Deixis: How Finely Do Languages Divide Space?*. Saarbruecken: V.D.M. Verlag, Dr. Mueller Aktiengesellschaft & Co.

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.

McKoon, G., and Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *J. Verb. Learn. Verb. Behav.* 19, 668–682. doi: 10.1016/s0022-5371(80)90355-2

Mitkov, R. (1999). "Anaphora resolution: The state of the art," in *Paper based on the COLING'98/ACL'98 Tutorial on Anaphora Resolution*, (Wolverhampton: University of Wolverhampton).

Peeters, D., Hagoort, P., and Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition* 136, 64–84. doi: 10.1016/j.cognition.2014.10.010

Peirce, C. S. (1955). *Philosophical Writings of Peirce*. New York, NY: Dover Publications.

Recanati, F. (2005). "Deixis and anaphora," in *Semantics vs. Pragmatics*, ed. Z. Szabo (Oxford: Oxford University Press), 286–316. doi: 10.1093/acprof:oso/9780199251520.003.0007

Sanford, A. J., and Garrod, S. C. (1989). What, when and how?: Questions of immediacy in anaphoric reference resolution. *Lang. Cogn. Process.* 4, 235–263.

Silverstein, M. (1976). "Shifters, linguistic categories, and cultural description," in *Meaning in Anthropology*, eds H. A. Selby and K. H. Basso (New York, NY: Harper & Row), 11–57.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Lang. Commun.* 23, 193–229. doi: 10.1016/s0271-5309(03)00013-2

Talmy, L. (2000a). *Toward a Cognitive Semantics, volume I: Concept Structuring Systems. i-viii*. Cambridge, MA: MIT Press, 1–565.

Talmy, L. (2000b). *Toward a Cognitive Semantics, volume II: Typology and Process in Concept Structuring. i-viii*. Cambridge, MA: MIT Press, 1–495.

Talmy, L. (2011). "Cognitive semantics: an overview," in *Semantics: An International Handbook of Natural Language Meaning*, eds C. Maienborn, K. vonHeusinger, and P. Portner (Berlin: Mouton de Gruyter, HSK series), 622–642.

Talmy, L. (2018). *The Targeting System of Language*. Cambridge, MA: MIT Press.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fpsyg.2020.01779

Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention

*Anja Stukenbrock**

Chair of German Linguistics, University of Lausanne, Lausanne, Switzerland

The paper investigates the use of gaze along with deictics and embodied pointing to accomplish reference and joint attention in naturally occurring social interaction. It assumes that deixis, in its primordial use in face-to-face interaction, is an embodied phenomenon that involves gestural pointing as well as visual perception, thus giving rise to recurring gaze practices of the participants. The analysis draws on a model of the interactional organization of deictic reference and joint attention that serves as a sequential framework for investigating the functions of eye gaze. The analysis focuses on two meta-perceptive practices: gaze following and gaze monitoring. It shows that the use of these practices in naturally occurring social activities is context dependent, positionally sensitive, tied to participant roles, and temporally fine-tuned to the stream of the participants' verbal and embodied conduct. The sequential analysis of these practices further documents that meta-perceptive gaze practices contribute to the constitution of joint attention as mutually known by the participants. The data for this study were recorded with two pairs of mobile eye tracking glasses and an external camera. Methodologically situated within the framework of conversation analysis and interactional linguistics where video recording is used, the study breaks new ground by employing a technology almost exclusively applied in experimental frameworks to record ordinary social activities "in the wild." In striving for ecologically valid and precise eye gaze data, it also contributes to a refinement of concepts developed in experimental paradigms by adapting them to qualitative research within the field of multimodal conversation analysis and interactional linguistics.

Keywords: reference, deixis, joint attention, gaze following, mutual monitoring, sequential organization, social interaction

OPEN ACCESS

Edited by:
Holger Diessel, Friedrich Schiller University Jena, Germany

Reviewed by:
Katharina J. Rohlfing, University of Paderborn, Germany
Pia Bergmann, Friedrich Schiller University Jena, Germany

***Correspondence:**
Anja Stukenbrock
anja.stukenbrock@unil.ch

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *15 February 2020*
Accepted: *29 June 2020*
Published: *11 September 2020*

Citation:
Stukenbrock A (2020) Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention. Front. Psychol. 11:1779. doi: 10.3389/fpsyg.2020.01779

INTRODUCTION

Across languages, demonstratives, or deictics, constitute a particular class of linguistic items. They are defined by a range of features that distinguish them both from grammatical and lexical items: they are universal (Himmelmann, 1997; Diessel, 1999, 2006); they have developed so long ago that they cannot be reconstructed diachronically from either lexical or grammatical items (Diessel, 1999, 2006); in ontogenesis, they are among the first words that children acquire (Clark, 1978); they constitute the only linguistic class that is inextricably connected to gestures (Bühler, 1965 [1934];

Diessel, 2006; Stukenbrock, 2015); and they fulfill one of the most central functions in human communication: the establishment of joint attention (Diessel, 1999, 2006).

There are various ways in which we can direct a person's attention to an external object. An important attention-directing device is gestural pointing, which precedes the acquisition of verbal deictics and is considered to be universal (Povinelli and Davis, 1994; for a different view cf. Wilkins, 2003) and uniquely human (Butterworth, 2003; for a differentiated overview of great apes' capacities for imperative vs. declarative/referential pointing cf. Tomasello and Call, 2019). Whether established verbally and/or gesturally, joint visual attention constitutes a triadic relationship between two participants and an object. This implies that they shift their gaze to the object. In face-to-face interaction, these gaze shifts may be observed and interpreted by the participants with respect to the ongoing activities. Eye gaze thus assumes an interactional function (Rossano, 2012, 2013). It shapes participants' actions and mutual understandings of those actions. Gaze may also serve to establish joint attention in the absence of verbal and/or gestural pointing. By following another person's line of regard (Flom et al., 2007), and by observing and inferring what the other person sees, we may share attention with that person. Note that joint attention involves more than following another person's gaze, or gesture, to an object. Joint attention presupposes two (or more) persons focusing on the same phenomenon and being aware of that act (Clark and Marshall, 1981; Moore and Dunham, 1995). In other words, joint attention must be mutually known in order to become part of the participants' common ground (Clark, 1996).

How do participants in the course of demonstrative reference know that joint attention has occurred? How do they make sure that they are seeing the same phenomenon, and see it in the same way? How do they achieve a mutual understanding of the referent? And how does gaze help them shape their actions according to the actions of the other? The present paper is concerned with these questions. It investigates the use of deictics, pointing, and gaze to accomplish joint attention in naturally occurring social activities. It assumes that the use of deictics in face-to-face interaction is inexorably connected to embodied, visible acts of demonstration. It proposes that eye gaze constitutes an integrative component of how joint attention is cooperatively accomplished. Coordinating joint actions is less successful when participants have limited or no visual access to the same objects and/or to each other (Clark and Krych, 2004). In the process of jointly attending to objects, participants look at each other at key moments in order to design and time their actions with respect to the actions of the other.

This paper aims at identifying those moments by focusing on the gaze behavior that participants themselves orient to in the course of demonstrative reference. By checking the gaze orientation of their coparticipant, they attribute relevance to the other's perception. The inferences that participants draw from perceiving the other's perception are displayed subsequently in their own behavior. In turn, these displays of understanding shape the participants' next actions and as such become accessible to sequential analysis (Schegloff, 2007). Gaze practices that turn perception into an object of perception will be termed meta-perceptive gaze practices:[1] a person (*ego*) looks at the eyes of another person (*alter*), perceives the other's gaze orientation as an index of visual perception, and interprets it within the context of its occurrence. Two types of gaze practices will be investigated: *gaze following* and *gaze monitoring*. While the former is tied to the participant role of the addressee who follows the speaker's gaze, hence *speaker gaze following*, the latter is performed by the speaker to check the gaze orientation of the addressee, hence *addressee gaze monitoring*.

I begin by discussing previous research on deixis, gesture, and gaze. Subsequently, I introduce a model of deictic reference that provides the framework for analyzing the meta-perceptive gaze practices in the context of demonstrative reference and joint attention (section "A Model of the Interactional Organization of Deictic Reference and Joint Attention"). Next, a note on materials and methods explains how mobile eye tracking was applied within the framework of conversation analysis (section "Materials and Methods"). This is followed by sequential analyses of deictic reference and meta-perceptive gaze practices within everyday activities (section "Analysis: Meta-Perceptive Practices and Joint Attention in Deictic Reference"). The paper concludes with a discussion of the findings (section "Discussion").

Conceptualization of Deixis as an Embodied Phenomenon

In research on deixis, two theoretical traditions can be distinguished. In the Anglo-American tradition, the terms *deixis* and *indexicality* are used "coextensively" (Levinson, 2004, p. 97) and apply to the broader phenomenon of context dependency. In contrast, the European tradition within which my paper is situated favors a narrow definition. Following Bühler (1965 [1934]), the concept of *deixis* refers exclusively to the grammatical encoding of context dependency in a closed set of linguistic items (deictics/demonstratives). Deictics have grammaticalized the space-, time-, and person-bound structure of the participants' subjective orientation (*origo*) in the speech event (Bühler, 1965 [1934]). Bühler distinguishes three modes of deictic reference: (1) *demonstratio ad oculos et ad aures* (reference to visible phenomena in the surroundings), (2) *anaphora* (reference to elements in the context of speech), and (3) *Deixis am Phantasma* (reference to absent phenomena that have to be imagined). Only *demonstratio ad oculos* is relevant for the present paper. The Latin syntagma that Bühler chose (*demonstratio* = "pointing" and *ad oculos* = "to/for the eyes") links the speaker's gesture to the addressee's eyes, thus anticipating an understanding of deixis as an embodied, interpersonal phenomenon.

Bühler postulated that gestures constitute an indispensable component of verbal deixis (Bühler, 1965 [1934], p. 93). In the Anglo-American tradition, this was also acknowledged by Fillmore who distinguished gestural from symbolic and anaphoric usages: gesturally used deictics "can be properly interpreted only by someone who is monitoring some physical

[1]The term "practice" is used in conversation analysis and interactional linguistics to distinguish actions from the practices that are used to implement those actions. It denotes "recurrent ways in which linguistic (and other) resources are used for particular purposes, for instance, in constructing turns, organizing turn taking, initiating repair" (Couper-Kuhlen and Selting, 2018, p. 29).

aspect of the communication situation" (Fillmore, 1997 [1971], p. 62), notably, the speaker's gesture, body orientation, gaze direction, etc. Various strands of video-based research have since contributed to an understanding of deixis and gesture in spoken interaction. They provide a point of departure for including gaze in the discussion.

In gesture studies, Kendon (1988, 2004) laid out a continuum of gesture types, later termed "Kendon's continuum" (McNeill, 1992, McNeill, 2000): different types of pointing gestures were systematically described (Kendon and Versante, 2003; Kita, 2003; Kendon, 2004), including the use of different body parts such as lips (Sherzer, 1973; Enfield, 2001), nose (Cooperrider and Núñez, 2012), and eye gaze (Kendon, 1967; Streeck, 1988, 1993, 2002; Stukenbrock, 2015). Conversation analytic studies have sharpened our understanding of demonstrative reference as an interactional achievement that requires coordinating talk, gestures, gaze, and body movements (Streeck, 1988, 1993, 2002; Goodwin, 2000, 2003; Hindmarsh and Heath, 2000; Hausendorf, 2003; Mondada, 2005, 2007, 2012, 2014; Stukenbrock, 2008, 2009, 2010, 2015; Eriksson, 2009). These works show that the use of deictics and concurrent gestures form multimodal packages, or *Gestalts*, that are recipient designed (Sacks et al., 1974; Hindmarsh and Heath, 2000), and coupled with the environment in which they occur (Goodwin, 2003, 2007). An integrative account of demonstratives and gestures as closely coupled, temporally flexible resources includes describing the interaction between speaker and addressee gaze.

Multimodal Deixis and Gaze: Temporality, Interactivity, and Intersubjectivity in Context

In face-to-face interaction, gesturally used deictics demand that addressees shift their gaze away from the speaker and to the target object. They constitute a summons that makes a response of the addressee relevant, the response being visual attention by means of gaze allocation. The idea that features in the speaker's talk instantiate a summons for addressee gaze was first formulated by Goodwin (1980, 1981): restarts, pauses, and hesitations produced by the speaker request the gaze of a non-gazing hearer (Goodwin, 1981, p. 280). A similar relationship holds for the use of deictics and addressee gaze, with an important difference: the gaze summons implemented by a demonstrative signals that the addressee is expected to look at the speaker's body to gather visual information on the location of the object (Stukenbrock, 2018c). In short, whereas restarts, pauses, and hesitations summon the addressee to establish mutual gaze and a *dyadic relationship* with the speaker, deictics summon the addressee to participate in a *triadic relationship*, i.e., to look at the speaker and pick up embodied cues to locate the object.

Although conversation analytic studies have revealed the multimodal complexity of deixis and have shed some light on gaze (Hindmarsh and Heath, 2000; Goodwin, 2003; Eriksson, 2009; Stukenbrock, 2009, 2015; Mondada, 2014), to date, they rely exclusively on video recordings that do not allow to zoom in on the details of gaze. In contrast to video recordings, mobile eye tracking technology delivers information on the location and duration of target fixations, the trajectories of gaze shifts, and, last but not least, on the interaction between speaker and addressee gaze.

While eye trackers have been used in experimental studies to examine joint attention and reference (Louwerse and Bangerter, 2005, 2010; Land, 2006; Hanna and Brennan, 2007; Clark and Gergle, 2010; Gergle and Clark, 2011), mobile eye tracking studies on deictic reference in naturally occurring interaction are practically non-existent (see, however, Stukenbrock, 2018a,b; Stukenbrock and Dao, 2019). In contrast, experimental studies on gaze and reference are undertaken in highly controlled settings, most of them stationary with the participants seated. For instance, the question how participants' gaze patterns interrelate was tackled in dual eye tracking experiments on the coordination of gaze in a programming task in which two programmers worked on the same code on two different computers placed opposite one another. The concept of "gaze cross-recurrence" (Jermann and Nüssli, 2012) was introduced to capture how much participants looked at the same spots simultaneously, and a method was developed for the automatic extraction of gaze cross-recurrence in large amounts of experimental data.

In everyday life, however, reference often involves participants on the move as well as moving objects. Mobile configurations are both shaped by and shaping the participants' use of verbal and visual resources (De Stefani, 2010; Haddington et al., 2013; De Stefani and Gazin, 2014; De Stefani and Mondada, 2014). Their investigation cannot be treated separately from the dynamically changing context that they help constitute. A few experimental studies have applied mobile eye tracking to compare reference in stationary and mobile settings. In a conversation elicitation task (Clark and Gergle, 2010; Gergle and Clark, 2011), participants in seated and standing mobile conditions were asked to discuss LEGO objects according to their likelihood of being replicas of modern art. The results showed differences between mobile and seated participants, i.e., mobile pairs used a higher proportion of local deictics for reference than seated participants but showed a lower proportion of gaze overlap (Gergle and Clark, 2011, p. 442). These results point to the context dependency and flexibility of participants' solutions to the problem of establishing joint attention and thus underline the need to study them in the social contexts in which they are embedded.

This also holds for gaze following (Flom et al., 2007). It has been noted that gaze following includes "an inference about perception: The observer follows to see what the gazer perceives" (Brooks and Meltzoff, 2014, p. 171). In ordinary activities, these inferences are based on context- and activity-related attributions of intentionality and meaning. In experimental settings, however, gaze following is examined as a mechanism that involves an observer following directional gaze shifts of an experimenter in a decontextualized way. Its investigation plays a prominent role in developmental studies on the age relatedness of infants' capacity for joint attention and for cooperative behavior (Scaife and Bruner, 1975; Carpenter et al., 1998; Flom et al., 2007; Tomasello et al., 2007), on the relationship between gaze following and language development (Brooks and Meltzoff, 2008), and on infants' abilities to use first-person experiences to understand the visual experiences and minds of others

(Brooks and Meltzoff, 2014). From an interactional perspective, gaze following in experiments is initiated by the experimenter who first establishes (eye) contact with the gaze follower and then shifts gaze to the object. However, in everyday life, gaze following emerges in particular spatial, temporal, and social contexts and assumes interactional significance within those contexts. Alternative paths such as eye–hand coordination between child-eye and adult-hand instead of eye gaze following may be preferred by social partners in more natural, free play contexts with several targets (Yu and Smith, 2013). Note that, in the flow of everyday activities, gaze following may just be an evanescent event that does not become relevant in interaction. Without uptake, however, participants may not be aware of its occurrence. This, then, does not constitute joint attention as mutually known by the participants (Clark, 1996).

In sum, experimental studies face three problems that affect the transferability of their results to human interaction "in the wild:" the problem of temporality, the problem of interactivity, and the problem of intersubjectivity. Depending on the interactional, cognitive, and perceptual availability of the participants, the interactional accomplishment of joint attention follows different temporal orders. Participants establish joint attention not only *simultaneously* but also *successively*, i.e., when speakers withdraw their gaze from the object before addressees look at it. Whereas instances of the first case lead to "gaze cross recurrence" (Jermann and Nüssli, 2012) in the eye tracking data, instances of the second case do not. Accomplishing joint attention implicates more complex, reciprocally adaptive gaze patterns of the participants than looking at the same target. These include mutual gaze as well as meta-perceptive practices such as gaze following and gaze monitoring. For instance, a common practice that infants have yet to acquire (Franco and Butterworth, 1996; Liszkowski et al., 2004) involves a pointing speaker (P) checking the visual orientation of the addressee (A) at the moment when A is looking at the target. In the eye tracking data, this appears as a fixation of P's gaze cursor on A's eyes at the moment in which A's gaze cursor reveals a fixation on the target. This moment of P perceiving A's perception constitutes an interactional mechanism for P to make inferences about A's perception and understanding.

Given that the eye tracking data reveal that the gaze of P and A is on the same target, the participants themselves may not be aware of the fact that they are looking at the same thing. Even if they know that they do, they may (or may not) find out subsequently that they have constructed different referents. As targets and referents are not the same, they need to be distinguished (Quine, 1960, p. 29ff; Clark et al., 1983, p. 245ff; Stukenbrock, 2015, p. 72–85, 282–313) to avoid naive conclusions about reference solely on the grounds of gaze fixations in the eye tracking data. Even though these fixations are technically precise, they do not reveal *prima facie* what the participants "really" see and what they referentially construct from what they see (Goodwin, 1994).

To resume, the problem consists of how to ground claims that (1) joint attention has occurred in the course of demonstrative reference, (2) mutual knowledge about the occurrence of joint attention exists, and (3) an intersubjective understanding of the referent has been achieved. I argue that participants themselves are continuously confronted with these problems and have developed routine solutions for them in social interaction. These solutions constitute the sequential orderliness of demonstrative reference in naturally occurring interaction. They are temporally flexible, context-sensitive, and thus serve participants' practical needs in everyday life. The following model reconstructs those solutions as interactional jobs that participants fulfill to establish reference and joint attention.

A MODEL OF THE INTERACTIONAL ORGANIZATION OF DEICTIC REFERENCE AND JOINT ATTENTION

In this section, I present a model that conceptualizes deictic reference as an interactional accomplishment (Stukenbrock, 2015, p. 495) and accounts for the multimodality, temporality, and intersubjectivity of demonstrative reference in face-to-face interaction. The model specifies the interactional jobs that participants have to fulfill depending on their roles as referring/pointing speaker and addressee. The jobs are conceived of as sedimented solutions to participants' concrete problems of sharing attention on visible phenomena in everyday life. While their jobs are complementary, both participants actively contribute to joint attention. The model was developed empirically on the basis of a large video corpus, with methods from conversation analysis and interactional linguists (Stukenbrock, 2009, 2015, 2016). The usability of the model was documented by developmental studies on infants' capacities to establish reference (Heller and Rohlfing, 2017; Heller, 2019). For the present study, it provides the framework for an investigation of gaze practices used by participants in the course of demonstrative reference.

The jobs (1–10) are accomplished by the participants in temporally flexible ways. If, for instance, participants already maintain "an eye-to-eye ecological huddle" (Goffman, 1963, p. 95), they will not have to establish focused interaction in order to jointly attend to an object. In contrast, speakers will have to mobilize additional resources to summon disattending addressees with whom they are currently not engaged. The first job thus defines the interactional precondition for attention sharing. The subsequent jobs detail the specific tasks that participants are faced with according to their role as referring participant (P) and addressee (A).

(1) Establishing Focused Interaction

In order to jointly attend to a visible phenomenon in their surroundings, participants may have to establish *focused interaction* first (Goffman, 1963). In contrast to *unfocused interaction*, in which persons are merely together in the same situation and may glean information about one another without getting engaged, *focused interaction* "occurs when persons gather close together and openly cooperate to sustain a single focus of attention, typically by taking turns at talking"

(Goffman, 1963, p. 24). When persons are momentarily not in focused interaction but within reach of one another, they sometimes use demonstratives to (re-)engage another person in focused interaction.

(2) Projecting a Domain of Pointing

Participants who are in focused interaction may use demonstratives to share attention to visible objects with their addressees. In order to direct A's visual attention to an object, P first has to orient him- or herself in space. Prior to initiating joint attention, P can thus be seen to look at the object before sharing it with A. In other words, P projects a *domain of pointing*.

Depending on the local context, P's self-orienting practices may be witnessed by A. We can therefore expect that P's self-orientation may be used by A to anticipate a relevant domain by following P's line of regard. This meta-perceptive gaze practice of observing the visual orientation of the partner and following his or her gaze to a new domain will be investigated in the section on "Deictic Reference, Speaker Gaze Following, and Joint Attention."

(3) Establishing the Perceptual Relevance of the Body as a Semiotic Resource

In order to direct A's visual attention to an object, P has to make sure that his or her gesture is visible to A. P has to establish the perceptual relevance of his or her own *body as a semiotic resource* to be attended to by A. This can be done through verbal summons, salient body movements, touch, etc. and, most importantly, verbal deictics. As has been argued, verbal deictics constitute a central gaze-summoning device in face-to-face interaction (Streeck, 2002; Stukenbrock, 2009, 2011, 2018a,c).

We can expect P to monitor the success of his or her attempt to secure addressee gaze (Goodwin, 1980, 1981). Gaze shifts from P to A may not only occur at the end of P's utterance. Instead, we can expect A-gaze monitoring by P while his or her referential action is still emerging. This meta-perceptive practice of monitoring the addressee's gaze will be investigated in the section on "Deictic Reference, Addressee Gaze Monitoring, and Joint Attention." In the model, it is defined as job 9 (below).

(4) Demonstratives and (5) Pointing Gestures as Multimodal *Gestalts*

Gesturally used deictics need a pointing gesture to direct A's visual attention to the target. The unity of verbal and gestural components in demonstrative reference has been conceptualized as a multimodal *Gestalt* (Heath, 1985; Goodwin, 2003; Streeck, 2009; Stukenbrock, 2011, 2015; Mondada, 2015). Note that, while the demonstrative has to be heard and understood, the concurrent gesture has to be seen by the addressee. In order to maximize the opportunity for successful reference, demonstratives and gestures are deployed in context-sensitive, temporally flexible, and recipient-designed ways (Hindmarsh and Heath, 2000). Their flexible use serves to synchronize the performance of P's gesture with A's gaze allocation to P (Stukenbrock, 2018c). Their multimodal packaging and local timing crucially depends on A's activities. Again, monitoring the addressee (job 9) helps P to maximize interpersonal coordination with A (cf. section "Deictic Reference, Addressee Gaze Monitoring, and Joint Attention").

(6) Constituting a Domain of Scrutiny

In general, it is assumed that A uses P's pointing gesture to extrapolate a linear vector to the target (Fillmore, 1982, p.46). However, in naturalistic settings with dense perceptual and cognitive ecologies, locating the target is a complex task (Stukenbrock, 2009, p. 305f.); it cannot be reduced to geometrical operations as in controlled experimental settings (Butterworth, 2003). Instead of extrapolating a vector, A first has to constitute a *domain of scrutiny* (Goodwin, 2003, p. 221) within which the *target* is to be found.

(7) Identifying the Target

After the domain of scrutiny has been established, A has to identify the *target*. Depending on contextual factors such as distance, complexity, accessibility, and transparency of the domain of scrutiny (Goodwin, 1996), identifying the target may either be unproblematic or lead to repair sequences. The perceptual task of identifying the target is intimately connected to the cognitive task of constructing the referent. Monitoring A while he or she is fulfilling this task may help P to anticipate success, or failure, of the referential act.

(8) Constructing the Referent

Standard accounts do not distinguish between target and referent. Following Quine (1960) and Clark et al. (1983), I distinguish between the perceptual task of establishing the target and the cognitive task of identifying the referent (Stukenbrock, 2009, p. 307–309). Although these tasks are normally accomplished as one, the distinction is licensed by repair sequences. These document categorically different problems (repairables) leading to problem-specific repair (Stukenbrock, 2015, p. 302–313). The distinction between target and referent accounts for locally emerging problems that participants themselves orient to as distinct trouble sources.

According to the literature on embodied reference (Hindmarsh and Heath, 2000, p. 1872f.), repair occurs after misunderstandings have been revealed in A's subsequent turn[2]. However, repair may also occur earlier in the course of a referential action when P looks back at A, anticipates problems, and repairs by dynamically enhancing target visibility (job 7) or facilitating referent construction (job 8). We can expect that addressee gaze monitoring provides a critical resource for participants to foresee trouble and repair early and thus to assure intersubjectivity and progressivity (Heritage, 2007). The following paragraph explains this job (9) in general terms.

[2]This is, in general, the sequential position for other-initiated repair, i.e., repair launched not by the speaker, but by the addressee, and then left to the speaker to accomplish the actual repair (other-initiated self-repair according to Schegloff et al., 1977).

(9) Monitoring Addressee Gaze

For demonstrative reference to be successful, P has to make sure that it is comprehensible and efficient, i.e., that A perceives what P wants A to perceive (i.e., P's body as a semiotic resource, domain of scrutiny, target). Monitoring A in the course of demonstrative reference is a powerful instrument for P to shape his or her action in recipient-designed ways. When P gazes at A to monitor A's gaze orientation, A's perception becomes the object of P's perception. Perceiving A's perception (*Wahrnehmungswahrnehmung*; cf. Luhmann, 1984, p. 560; Hausendorf, 2003) at key moments in the interaction allows P to adapt ongoingly to what P perceives and infers about A's perception. Addressee gaze monitoring will be investigated in the section on "Deictic Reference, Addressee Gaze Monitoring, and Joint Attention."

(10) Display of Understanding

The referential sequence is brought to a close when A displays understanding in the subsequent turn. A's display of understanding implies a successful identification of target (7) and referent (8), as well as an unproblematic completion of the preceding tasks. However, A's response may also document trouble and initiate repair. As has been argued above (jobs 8 and 9), response turns are not the only position in which problems may surface and lead to repair. This may also happen earlier in the sequence.

Interim Summary

In this section, a model was introduced that explains the interactional organization of deictic reference in terms of the jobs participants have to fulfill in order to establish joint attention on the referent. While formulated on an abstract level, the model is grounded in detailed observations on participants' actions in face-to-face interaction (Stukenbrock, 2015). The jobs make particular gaze patterns expectable: first, *gaze shifts to objects*, and second, *gaze shifts to coparticipants*. When P sees an object and wants to share it with A, he or she will have to shift gaze between the two, and when A is summoned by P to attend to that object, A will likewise shift gaze between P and the object. According to the model, the two types of gaze patterns, gaze shifts to objects vs. gaze shifts to coparticipants, perform different functions according to their sequential position and the participant role of the gazer (P or A). Notably, P-gaze to an object of joint attention projects a future domain of pointing and serves as self-orientation (job 2) before P initiates joint attention. In contrast, A gaze to the domain of scrutiny (job 6) aims at identifying target and referent. This is accomplished by A on the grounds of his or her perception of P (jobs 3 and 4) and inferences about P's communicative intentions.

Furthermore, the model states that gaze shifts to coparticipants serve different, context-specific functions. P-gaze to A, and A-gaze to P, when aimed at the other's eyes, implement meta-perceptive functions. The partner's perception is turned into the object of perception. According to the model, the interactional function of meta-perceptive gaze practices differs with respect to participant role, sequential context, and temporal implementation. The subsequent analysis is based on the expectation that P-gaze to A, before, during, and after a deictic referencing act serves different functions than A-gaze to P in the course of that act.

MATERIALS AND METHODS

Participants

The study is based on two corpora of video data recorded with mobile eye tracking glasses worn by participants in naturally occurring interaction. The participants were friends or colleagues who engaged with each other as part of their lives. This naturalistic approach to data collection is derived from the endeavor of conversation analysis to reconstruct the "endogenous organization of social activities in their ordinary settings" (Mondada, 2013, p. 33). Two different settings were selected from the range of activities represented in the corpus: shopping together at a market (three dyads), and visiting a museum together (three dyads). The choice was based on the observation that, in both settings, mobile and static configurations emerge naturally as participants collaboratively move on or stop to share attention on objects they consider noticeable in the local context.

Data Collection and Analytic Procedure

For the first corpus, SMI glasses (sampling rate, 30 Hz) were used. The data for the second corpus (SNSF project DeJA-VI) were recorded with Tobii Pro Glasses II (sampling rate, 50 Hz) and an external video camera to offer a full shot on the activities. The eye tracking recordings and the video from the external camera were frame-precisely synchronized with *Adobe Premiere Pro* and exported as a single split-screen video. The split-screen video and the corresponding wave file were imported into ELAN (Wittenburg et al., 2006) in order to make verbal transcriptions and multimodal annotations. Each split-screen video lasts between 30 and 45 min. Altogether, the verbal transcripts cover roughly 4 h. Talk was transcribed according to GAT2 (Selting et al., 2009; see **Supplementary Material**). The videos were precoded for all occurrences of gesturally used deictics and concurrent gestures. Gesturally used personal, temporal, and modal deictics were excluded from this study. The detailed analyses drew on a subset of demonstrative reference in which joint visual attention was (1) attested in the eye tracking data (gaze fixations on the target), (2) displayed by the participants in social interaction, and therefore (3) mutually known to the participants (Clark, 1996), as well as (4) accessible to sequential analysis. Twenty sequences were analyzed using methods from multimodal conversation analysis (Schegloff, 2007; Sidnell and Stivers, 2013) and interactional linguistics (Selting and Couper-Kuhlen, 2001). The multimodal annotation was adapted from Mondada (2019; see **Supplementary Material**). The extracts presented in the

subsequent section cover both corpora and exemplify context-specific variations in the use of meta-perceptive gaze practices along with demonstratives and embodied conduct in the course of demonstrative reference.

ANALYSIS: META-PERCEPTIVE PRACTICES AND JOINT ATTENTION IN DEICTIC REFERENCE

Gaze shifts to objects and gaze shifts to coparticipants are part of the sequential orderliness of demonstrative reference and joint attention in ordinary interaction. Both types of gaze shifts may pass unnoticed, or they may be perceived and oriented to by the participants. In this section, I investigate gaze practices that participants use in order to gain access to the coparticipant's visual perception. As members' practices, these meta-perceptual practices constitute procedural solutions to the problem of coordinating bodies and minds for demonstrative reference and joint attention. The deployment of meta-perceptive gaze practices is sensitive to spatial, temporal, and interactional affordances and constraints of the context. The aim of the analysis is to uncover how participants' mutual access to one another's gaze helps them shape their actions moment by moment according to the action of the other. The analysis starts with speaker gaze following. With respect to the sequential jobs described in the model above, gaze following occurs early and thus shapes the way in which subsequent jobs are accomplished. Next, addressee gaze monitoring will be investigated. It is closely coupled with the speaker's production of deictics and thus occurs later than gaze following within the sequential ordering of jobs. Last, I will present instances in which meta-perceptive gaze practices are absent, and propose that in the local context, participants' fine-tuned interpersonal coordination allows them to establish joint attention in implicit ways.

Deictic Reference, Speaker Gaze Following, and Joint Attention

This section is concerned with the following problems: How is gaze following organized in everyday activities? When does it

Extract 1: "Hokkaido" (WM02_00:13:20)

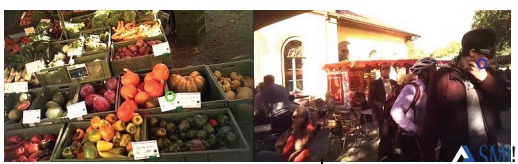


Figure 1, left: P-gaze at domain of pointing; right: A-gaze to P

```
01          (1.0) (1.0) (1.0) (1.0) (1.0)
    P-gz    ~~~scans stall|-hokkaido--|..
    A-gz    ~~~looks around~~|.|-P-|..GF..>
```

Figure 2, left: P-gaze to A; right: gaze following by A

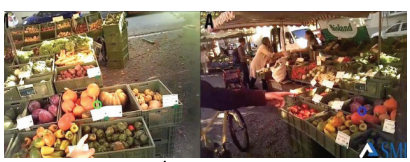


Figure 3 : simultaneous joint attention

```
02  A-vb    [SOllen wir,  ]
             shall we
03  P-vb    [dA steht AUCH] hokkaido dran;
            there it also says hokkaido
    P-ge                    ......|PG|,,,|
    P-gz    -A-|.........|--hokkaido------>
    A-gz    ..GF..........|--hokkaido---->
04  A-vb    JA;
            yes
    P-gz    -hok-->
    A-gz    -hok-->
05          (0.6)
    P-gz    -hok-->
    A-gz    -hok-->
06  A-vb    hier SOllen wir mal einfach irgendwAs ähm besorgen schon mal?
             here should we PTCL just get something uhm in the meanwhile
    P-gz    --------------------------------|,,|~scans diff.pumpkins~~~~~~>
    A-gz    --------------------------------|,,|~scans diff.pumpkins~~~~~>

07          (0.3)
```

occur, how does it contribute to demonstrative reference and joint attention, and how is it integrated in the sequential order of social actions? To answer these questions, gaze following is conceptualized as an interactional gaze practice of tapping into a coparticipant's gaze and exploiting it as a resource to gather information on *where*, *how*, and *for how* long he or she is looking, and to infer *what* he or she is looking at, and *why*.

The first extract ("Hokkaido") is from the market corpus. The figures in the transcript were extracted from the split-screen video and represent the participants' respective eye tracking recordings. In the figures, P's perspective is displayed on the left (green gaze cursor) and that of A on the right (blue gaze cursor). The color coding in the transcripts corresponds to the participants' gaze cursor. The following abbreviations are used: P: referring participant; A: addressee; GF: gaze following; PG: pointing gesture; vb: verbal; gz: gaze; and ge: gesture (see **Supplementary Material** for details).

The extract exemplifies the projective force of self-orientation by P to a domain of pointing (job 2) and the interactional significance of early gaze following by A for reference and joint attention. It starts after two friends, P and A, have arrived at a local farmers' market. They have talked about different sorts of pumpkins for sale at the first stall. A's preference for Hokkaido constitutes the common ground for P's subsequent noticing. The noticing refers to a sign indicating Hokkaido. It contains a demonstrative and is accompanied by a pointing gesture (l. 3). However, the resource used by A to co-orient with P is not his gesture but his gaze. An analysis of the temporal details demonstrates that A taps into P's gaze orientation and follows his line of regard early. Consequently, her gaze is already in place when the reference occurs.

At the beginning of the extract (l. 1), the participants are in an open state of talk (Goffman, 1981, p. 74, 134). P is looking at a sign indicating Hokkaido (**Figure 1**, left) when A turns toward him (**Figure 1**, right) and begins to follow his line of regard. Focused interaction (job 1) is established when they simultaneously start talking (l. 2 and 3). While A's utterance (l. 2: sollen wir/"shall we") projects a proposal and is broken off in the course of the overlap, P's turn is continued (l. 3). It contains a spatial deictic coupled with a gesture and followed by a nominal phrase (l. 3: dA steht AUCH hokkaido dran;/"there it also says hokkaido").

At the beginning of his utterance, P shifts his gaze to A (**Figure 2**, left). He can see that she is moving her gaze toward the domain he has been looking at before (**Figures 2**, **3**, right). When P shifts his gaze back to the pumpkins (**Figure 3**, left), joint attention has been established (**Figure 3**, left and right). A's confirmation (l. 4: JA/"yes") reveals successful, unproblematic reference. The participants' focus of attention is sustained as they synchronously scan the objects for a while and A ventures a buying proposal (l. 6).

Due to the temporality of its occurrence, gaze following plays a privileged role among the practices that contribute to the moment of joint attention. A taps into P's line of regard early and follows it to the phenomenon that P has been looking at before he refers to it. At this point, A is perceptually already orienting to the domain of scrutiny (job 6). Thus, P's referential act (jobs 4 and 5) helps her identify target and referent, but does not initiate her gaze shift to the target.

The second extract ("Balcony") is from the market corpus as well. The participants return from shopping together and arrive at the researcher's house. A is talking about her party when P interrupts her to comment on the facade of the house. In contrast to the first extract, the participants are already in focused interaction (job 1). The example demonstrates the occurrence of gaze following in an ongoing conversation. Whereas in extract 1, gaze following was initiated by A before the participants started talking, it is now set off by P's interruptions and occurs concurrently with talk. In the figures in the second transcript, P's perspective is situated on the left (green cursor) and that of A on the right (blue cursor).

A has complained about a friend who cannot come to her party. She closes the topic with an assessment (l. 13–14), and continues planning (l. 16) when she is interrupted by P who utters a noticing (l. 17). In the course of the interruption, A looks at P's eyes (**Figure 4**, right) who is, however, not reciprocating her gaze but looking upwards (**Figure 4**, left). A continues her turn and projects an account (l. 19), but P interrupts her again by proposing an improvement related to the house (l. 20).

Upon the second interruption (l. 20), A shifts her gaze once more to P's eyes (**Figure 5**, right). She can now see that he is still scanning the facade (**Figure 5**, left). A follows his line of regard (job 6) and also looks at the facade (l. 20–23) (**Figure 6**, right). After a pause (l. 21), she utters a response token (l. 22: ja:_a/"yeah"), which displays her understanding (job 10). It implies a successful identification of target (job 7) and referent (job 8). A's subsequent assessment (l. 23) aligns with P's stance toward the house. Thus, joint attention (**Figure 6**, left and right) as a mutually known visual, cognitive, and affective orientation to an object has been accomplished.

Several practices contribute to joint visual attention and intersubjectivity: P's noticing (l. 17) and his subsequent assessment (l. 20) are both sequentially placed as interruptions and solicit A's gaze (job 3). They entail, however, different gaze patterns. Whereas the first instance of gaze shift to P does not prompt A to follow P's gaze to the target, the second instance does. In contrast to the noticing (l. 17) that contains a deictic reference to the participants' present location in space and is interpretable without visual evidence, the subsequent assessment (l. 20) can only be responded to by A on the grounds of visual evidence. It necessitates perceiving what the speaker perceives in order to understand what he refers to.

Therefore, P's upward gaze acquires a different interactional status in the second instance: it is perceived by A as displaying relevant spatial information (job 3) and used as a directional cue. A taps into P's gaze and disregards the vague gesture that he performs too low for her to attend to. Due to the unperspicuousness of the gesture and the temporal precedence of gaze following, the former remains interactionally irrelevant. Instead of the gesture, speaker gaze following contributes to the emergence of joint attention in this sequence.

Extract 2: "Balcony" (WM02_00:34:41)

```
11  A-vb  <<lachend>der_is ja sO:_ne [verpeilte S]OCke,
                    he is such a space cadet
12  P-vb                             [((lacht))  ]
                                     ((laughs))
13        (0.2)
14  A-vb  das_is unGLAUBlich;>
          that is incredible
15  A-vb  ((schnupft))
          ((snuffs))
```

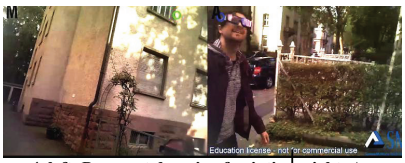


Figure 4, left: P-gaze at domain of pointing; right: A-gaze shift to P

```
16  A-vb  nee aber [es is gAnz PRAKtisch,]
          nah but it is quite practical
17  P-vb           [ah: hier SIM_ma;     ]
                    ah here we are
    A-gz  ..............|--Ps eyes------->
    P-gz                |..moving up.....>
18        (0.1)
          weil ähm: [dAdurch-
19  A-vb  because uhm in this way
```

Figure 5, left: P-gaze at target; right: A-gaze at P, subsequent gaze following

```
    P-vb              [ja_da_könnte_ma_noch_ma schö:ne balkOne hier HINbauen;
20                     PTCL you could PTCL build a nice balcony there
    P-ge                          ......lifts arm..|--|,,,??
    P-gz  ..|---scans facade---------------------------,,,.....|-As eyes-|
    A-gz  ----|,,,,,,,...|--Ps eyes-----|...GF.....|---facade----------->>
```

Figure 6: simultaneous joint attention

```
21        (0.7)
22  P-gz  ,,..|-facade->
    A-gz  ------------->
    A-vb  <<creaky>ja:_a,>
                yeah
    P-gz  -facade---------->
    A-gz  ----------------->
23  A-vb  das wÄr nicht SCHLECHT;
          that would not be bad
    P-gz  --------scans facade-|,,
    A-gz  -scans facade-------|,,,
24        (1.2)
```

The third extract ("Model") is from the museum corpus. In the transcript, A's perspective is displayed on the left (blue cursor) and P's on the right (green cursor). Two friends are at the Uniseum, the museum of the University of Freiburg/Br. The extract exemplifies an instance of unsuccessful gaze following and contains several repairs. As in extract 1, the participants are not in focused interaction. For various reasons, visual attention sharing is more difficult: the museum visitors are further apart than the participants at the market. They have established divergent foci of visual attention and are not available for mutual engagement. Instead of bodily adjustments, one of them has to walk over to the other to create a new interactional space. We join the participants when P summons A to look at an exhibit in the gynecological section while A is reading an explanation about

Extract 3: "Model" (Uniseum_01_00:32:26)

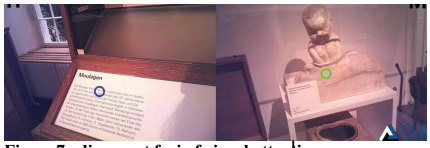


Figure 7 : divergent foci of visual attention

```
01  P-vb   °hh KUMma wie das AUSschaut;
           look PTCL what that looks like
    A-gz   -reading---------------------->

02         (2.3)
    A-gz   -reading----->

03  A-vb   WARte?
           wait
    A-gz   -reading-->
    P-gz   .|-A-|....>
```

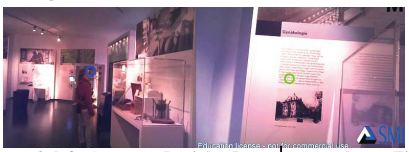


Figure 8, left: A-gaze to P; right: P-gaze on panel

Figure 9, left: gaze following by A; right: P-gaze on panel

```
04         (6.8)     (1.1)(0.2)(1.2)
    A-gz   -reading--|....|P-|~searching~>

05  A-vb   WAS,
           what
    A-gz   ~searching~>
    P-gz   .|model---->
```

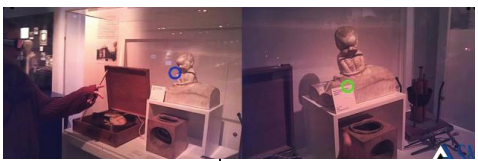


Figure 10, left: A-gaze at target; right: P-gaze at target

```
06         (0.8)     (0.2)
    P-ge           |-points->
    P-gz   -model----|-label-
    A-gz   ~searching|-model-

07  P-vb   gUck mal DEN hier;
           look at that one here
    P-ge   -points-------------|
    P-gz   -label--------------->
    A-gz   -model-------|-label->

08         (0.3)  (1.0)
    P-gz   -label|model----->
    A-gz   -label----------->

09  P-vb   das moDELL;
           the model
    P-gz   -model----->
    A-gz   -label----->

10         (1.0)
    A-gz   -label-|
    P-gz   -model->

11  P-vb   hhh° °h [hh

12  A-vb           [ja;
                   yes
    P-gz   -model----->
    A-gz   -model----->

13  P-vb   Irgendwie KRASS;
           somehow weird
    P-gz   -model---------->
    A-gz   -model---------->
14         (2.0)
15         ((laughing together))
```

moulage. The gaze following occurs during the pause in l. 4. The analysis focuses on the repair sequence initiated by A (l. 5) after A has turned to P and tried to follow P's line of regard.

The participants are several meters apart and have established divergent foci of attention (**Figure 7**, left and right) when P summons A to share an exhibit with her. It is a model of the expulsion phase of childbirth, as P has learnt from reading the label. When P summons A, she refers to the model with a demonstrative (l. 1: das/"that").

Instead of shifting her gaze to P, A asks her to wait (l. 3) and finishes reading. Subsequently, she turns round, walks to P (**Figure 8**, left), and follows P's line of regard (**Figure 9**, left) without, however, being able to narrow down the domain of scrutiny (job 6) and to identify the target (job 7). The search is documented in the eye tracking data by scanning eye movements across text panels and exhibits. A initiates repair (l. 5: WAS/"what"): the interrogative marks the demonstrative in P's summons (l. 1: das/"that") as the repairable (job 4).

P repairs (l. 7) by referring to the exhibit with a demonstrative (DEN/"that"), a proximal local deictic (hier/"here"), and a pointing gesture (jobs 4 and 5). The gesture precedes speech and directs A's gaze to the domain of scrutiny and target (jobs 6 and 7) before the utterance begins (**Figure 10**, left). P's utterance is designed to help A identify the target referent (jobs 7 and 8) at the moment in which her gaze arrives on site. A, however, does not respond. In contrast to extracts 1 and 2 in which displays of understanding (job 10) indicated successful reference, A's silence (l. 8) in extract 3 constitutes a noticeable absence.

The pause that ensues is taken by P as an indication that the problem persists. This is documented in P's subsequent turn: P repairs the deictic reference in her previous turn (l. 7: DEN hier/"that one here") with a coreferential nominal phrase (l. 9: das moDELL/"the model"). Sequentially, a response is expected, which displays A's understanding (job 10) and brings the repair sequence to a close. However, A still remains silent (l. 8). Why does she still withhold a response?

Note that in order to close the sequence as a whole, it does not suffice to display successful reference. From an interactional perspective, more is at stake. A also has to respond to P's implicit invitation to assess the referent[3]. The referential problem is thus inextricably linked to the problem of an expectable but missing assessment that A may find difficult to offer. Assessments are acts of participation in social activities (Pomerantz, 1984). They require and claim knowledge of the assessed referent. Refusing to assess a referent may index insufficient knowledge or trouble with the referent.

After another pause (l. 10) and an intake of breath that projects further talk by P (l. 11), A finally responds with a minimal acknowledgment token (l. 12: ja/"yes"). However, she does not engage with the exhibit any further. While the problem of localizing and identifying the referent has now been resolved, the interactional problem of assessing the exhibit persists. A's refusal to assess it is even more evident now. P orients to this and no longer pursues an assessment from A. Instead, P now offers an assessment herself (l. 13: Irgendwie KRASS/"somehow weird"). A, however, withholds a second assessment (Pomerantz, 1984). The sequence comes to a close as the participants, after another pause (l. 14), laugh their embarrassment at the gynecological model off (l. 15), and move on.

The extract shows that demonstrative reference is part and parcel of social actions such as assessments. Understanding the function of demonstratives, pointing, and gaze following in everyday activities is thus inexorably linked to understanding how their actual use is shaped by and continuously adapted to the unfolding interactional context.

To sum up, extracts 1 and 2 have exemplified instances in which speaker gaze following constitutes an important practice of establishing joint attention. Significantly, gaze following occurred early, i.e., the gaze follower, A, anticipated the domain of scrutiny (job 6) before it was indexed by demonstratives and pointing gestures in P's utterance (jobs 4 and 5). In both instances, verbal responses (job 10) documented successful reference by A in second position, i.e., after turn-taking. Extract 3, in contrast, exhibited an extended repair sequence. After a request to wait and an unsuccessful attempt at gaze following, A initiated repair (l. 5) with an interrogative (l. 5: *was*/"what"). Beyond verbal repair markers, delays or silences may also indicate referential problems. However, non-responding addressees may be silent for reasons other than unsuccessful reference. In extract 3, A's silence implemented her refusal to assess the referent. This led to a further repair by P until the two components of the expected response, acknowledging the referent and delivering an assessment, were untied. Whereas the referent was finally acknowledged by A, she never delivered an assessment. Extract 3 differs from extracts 1 and 2 in another respect: Not only is speaker gaze following unsuccessful, addressee gaze monitoring (see next section) is also absent. After briefly shifting her gaze to A at the very beginning, P never shifted her gaze to A again subsequently. The embarrassment at the gynecological model that prevented A from responding also forestalled mutual gaze between the participants, and it prevented P from addressee monitoring throughout the sequence. Consequently, P lacked important cues regarding A's participation and the (non)emergence of joint attention.

Deictic Reference, Addressee Gaze Monitoring, and Joint Attention

Verbal responses display addressees' understanding (job 10) of speakers' prior actions after turn-taking has occurred. In demonstrative reference, A's gaze orientation constitutes early evidence for P whether joint visual attention is emerging. This section investigates addressee gaze monitoring as a meta-perceptive practice of P to maintain intersubjectivity continuously in the course of reference. This minimizes the occurrence of extended repair sequences such as in extract 3.

We return to the balcony sequence (Extract 4: "Balcony Revisited") and analyze the temporal placement of P's gaze shift to A with respect to the sequential ordering of the participants' jobs. The example demonstrates that by shifting

[3]I would like to thank reviewer 2 for pointing out the additional trouble source of an expectable but missing assessment.

Extract 4: "Balcony Revised"

19 A-vb weil ähm: [dAdurch-
because uhm in this way

Figure 11, left: A-gaze monitoring by P; right: A-gaze at target

20 P-vb [ja_da könnte ma noch_ma schö:ne balkOne hier **HINb**auen;
PTCL you could PTCL build a nice balcony here
P-ge .....lifts arm..|--|,,,??
P-gz ..|--scans facade-----------------------------|,,.....|--As eyes-|
A-gz ----|,,,,,,,,,..|--Ps eyes-----|,,.........|--facade-------------->
21 (0.7)
P-gz ,,,..|-facade->
A-gz ------------->

Figure 12: simultaneous joint attention

22 A-vb <<creaky>**ja:_a**,>
yeah
P-gz -facade---------->
A-gz -facade---------->
23 A-vb das wÄr nicht SCHLECHT;
that would not be bad
P-gz --------scans facade-|,,
A-gz -scans facade-------|,,,
24 (1.2)

gaze from the target to A while P's initiating action is still underway, P gains evidence on how A is responding before the initiating action is closed. We join the participants when P interrupts A for the second time (l. 20) and focus on P's gaze shifts (green cursor) between the facade and A (blue cursor).

In the course of his utterance (l. 20), P continues to scan the facade. Toward the end, he shifts his gaze to A (**Figure 11**, left) who has stopped talking (l. 19). By looking at her eyes (**Figure 11**, left), P can see and infer that A is now looking at the facade as well (**Figure 11**, right). A moment of perceiving A's perception occurs (job 9). Shifting gaze to A and checking her visual orientation is an observable gaze practice in deictic reference (Stukenbrock, 2009, 2015, 2018c). In this extract, it occurs at a specific position within the sequential order of jobs, i.e., immediately after P's referential act and before A's response. Addressee gaze monitoring is used by P in a position-sensitive way, it is temporally coordinated with the jobs that A is expected to fulfill next. As will be shown below (extracts 5 and 6), this is not the only position for addressee gaze monitoring in deictic reference.

During the pause in l. 21, P shifts his gaze back to the facade. When A utters a response token (l. 22: ja:_a/"yes") and aligns with P's assessment (l. 23: das wÄr nicht SCHLECHT/"that would not be bad"), both are looking at the facade at the same time (**Figure 12**, left and right). A phase of simultaneous joint attention is thus established, one that is verbally confirmed (job 10) and thus mutually known by the participants. The temporal ordering reveals that A-gaze monitoring before the end of the utterance allows P to gather information about A's compliance and visual co-orientation before turn-taking occurs and a verbal response (job 10) is due. A-gaze monitoring is a practice of assessing online the outcome of initiating joint attention.

In the next example (Extract 5: "Carnival Game"), P engages in addressee gaze monitoring twice, before (l. 1, **Figure 13**) and after (l. 3, **Figure 14**) demonstrative reference (l. 2). In the transcript, A's perspective is on the left (green cursor) and P's perspective on the right (red cursor). The data are from recordings made in the Swiss Museum of Games. The participants, two friends who speak Swiss German, take a tour through the museum. In contrast to extract 4, addressee gaze monitoring is facilitated by the spatial configuration of the museum visitors. In contrast to the friends who return from the market and are walking while talking, the museum visitors have stopped in front of a showcase with games. They have established an interactional space that gives them access to the exhibits and to each other's faces. The following

Extract 5: "Carnival Game" (SM02_00:03:32)

Figure 13, left: A scanning exhibits; right: A-gaze monitoring by P before deictic reference

```
01  P-vb  ((laughs_____________)) (.)
    P-gz  |-game1-|......|-A--|......>
    A-gz  ........|game1|.|game2----->

02  P-vb  was au immer dAs ISCH,
          whatever    that is
    P-gz  ...|-game1----------|
    A-gz  -game2--------------|
```

Figure 14, left: A-gaze at target; right: A-gaze monitoring by P after deictic reference

```
03  P-vb  ich würd_s <<laughingly> gärn HA,> (.)
          I would              like to have it
    P-gz  ........................|A---|......>
    A-gz  ..|game1------------------------------>

04  P-vb  ((laughs)) [((laughs))]
    P-gz  ...|-game1-------------->
    A-gz  -game1------------------>

05  A-vb             [en BALL   ] [drischüsse;      ]
                      a ball shoot in it
06  P-vb                          [irgend so ne JOOR]marktspiel,
                                   some kind of carnival game
    P-gz             -game1-----------------------------------|
    A-gz             -game1-------------------------------|
```

analysis highlights the temporal and contextual sensitivity of A-gaze monitoring and its function in establishing joint attention.

Amidst the games in the showcase, P notices an unidentifiable exhibit, displays her surprise by laughter (l. 1), and in the course of laughing shifts her gaze to A. When P gazes at A (**Figure 13**, right), A is looking at a different exhibit (**Figure 13**, left). Subsequently, P formulates her stance toward the target exhibit (l. 2–3), referring to it with a demonstrative and a pointing gesture (jobs 4 and 5). At the end of P's deictic act, A shifts his gaze to the target (jobs 6 and 7) and keeps looking at it through the second half of P's turn (l. 3). When P shifts her gaze to A toward the end of her utterance (job 9), she can infer from his gaze orientation (**Figure 14**, right) that he is now attending to the same exhibit (**Figure 14**, left). Joint attention and successful reference (job 8) are further displayed (job 10) in A's comment (l. 5), which overlaps with P's laughter (l. 4) and talk (l. 6).

The sequential positioning of A-gaze monitoring differs from that in extract 4 in contextually shaped ways. The first A-gaze monitoring occurs concurrently with P's laughter (l. 1) and serves to check A's attention and availability. In the sequential context, laughter functions as an attention-getting device designed to engage the coparticipant. It also functions as a preinvitation to

A to share the object of P's laughter. In order to do so, A has to identify the reason for P's laughter and orient to it. P's subsequent turn is designed to engage A further. The demonstrative reference is part and parcel of a riddle (l. 2: was auch immer dAs ISCH/"whatever that is"), which invites coparticipation in solving it. In the second part of her turn (l. 3), P playfully expresses a desire for the exhibit and monitors A's gaze for the second time.

To sum up, P constructs a three-step sequence designed to engage A in sharing the exhibit with her:

(1) P-laughter as an attention-getting device and preinvitation, first A-gaze monitoring to check A's interactional and perceptual availability (l. 1),
(2) Demonstrative reference as part of a riddle (l. 2),
(3) Playful expression of stance toward referent, second A-gaze monitoring (l. 3).

By perceiving A's perception (job 9) at this particular moment, P can infer that A has oriented his visual attention to the object. Subsequently, A's verbal response displays successful reference (job 10). He offers a candidate solution to the riddle by replacing the demonstrative in P's utterance with a lexical item marked by hedges (l. 6: irgend so ne JOORmarktspiel/"some kind of carnival game").

This section concludes with an analysis of A-gaze monitoring in a triadic encounter (Extract 6: "Gambling Table"). Two colleagues and friends, C and T, are at the Swiss Museum of Games. Before they visit the exposition, they meet with the head of the museum (A) to plan a conference at the museum's event hall. We join the participants when they change topic from organizational (l. 11–17) to spatial arrangements (l. 18–24). The deictic reference occurs in a question asked by C and addressed at A, the head of the museum (l. 18-21). C is the pointing participant. In the figures in transcript 6, C's perspective is displayed on the right (red cursor) and T's perspective on the left (green cursor). The head of the museum is not wearing eye trackers. The bottom of the split screen displays the recording of the external camera.

After organizational matters have been settled, the participants close the topic (l. 11–15). C and T engage in mutual gaze (l. 11), C then withdraws her gaze from T, vaguely orients to A and the surrounding space (l. 12). When she utters the final closing device (l. 15: voiLÀ/"right"), she has already shifted her gaze to the surroundings (**Figure 15**, right), projecting a future domain of pointing (job 2). In contrast, her colleague T is still looking at the head of the museum (**Figure 15**, left).

C keeps looking at the domain of pointing (job 2) as she utters the first deictic (l. 16: HIER/"here") and performs a concurrent pointing gesture (jobs 4 and 5). The preparation phase of the gesture precedes the beginning of the turn; its apex is synchronized with the articulation of the first deictic (**Figure 16**, right). T, who is not directly addressed by the speaker, follows her pointing act and shifts his gaze to the domain of scrutiny (**Figure 16**, left).

The addressee of the utterance is the head of the museum, A, who is not wearing eye trackers and currently not visible in the video recording. Note that, at this point (l. 16), the participant roles of the addressed and the unaddressed recipient are not yet evident. Only in the course of the next turn-constructional unit (l. 17), when C shifts her gaze back to the head of the museum (A) and selects her with the VOS pronoun (l. 17: SIE/you), is the participation framework of this utterance established and the type of social action (a request) projectable[4]. T, who does not yet know that he will not be the addressee, shifts his gaze to the domain of scrutiny (job 6). This is valid eye tracking evidence (**Figure 16**, left) for the gaze-summoning function of deictics in demonstrative reference. By gazing at the domain of scrutiny before the VOS pronoun is uttered, he follows C's referential act continuously and prepares for the role of potential next speaker.

With less technological precision, the head of the museum (A) can be observed looking at the domain of scrutiny as well when speaker C shifts her gaze to A (**Figure 17**, right). A-gaze monitoring (job 9) allows C to perceive the perception of A and check online whether the referential act be successful.

Significantly, C's first gaze shift to A (job 9) is temporally placed within the uncompleted utterance, after the second occurrence of the proximal deictic HIER/"here" (l. 17). In both instances, the deictic bears the focal accent and is accompanied by a gesture (**Figure 16**). In the course of the second part of her utterance, C withdraws her gaze from A (l. 17–18). Toward the end of her utterance, C shifts her gaze back to A (**Figure 18**, right) who makes a broken-off attempt to respond (l. 19: ab') and then confirms C's request (l. 19: ABsolut/"absolutely").

C's gaze shift to A toward the end of the utterance is consistent with findings on the function of gaze in turn-taking (Kendon, 1967; Auer, 2020). Whereas the first, turn-internal gaze shift to A is closely tied to the referential act and serves addressee gaze monitoring (job 9), C's final gaze shift to A is motivated by the end of her turn. C's turn implements a request that makes a type-conforming second action relevant (i.e., granting or declining the request). Gaze allocation to A selects her as next speaker and serves to mobilize A's response (Stivers and Rossano, 2010).

Interim Summary: The previous sections have shown that meta-perceptive gaze practices, i.e., speaker gaze following and addressee gaze monitoring, contribute to the successful establishment of joint attention in deictic reference. Addressees who follow speaker gaze early may anticipate the domain of scrutiny. Their gaze thus arrives on the scene before, or at the moment in which talk further elaborates target and referent (previous section). Likewise, speakers may use different sequential positions to shift gaze between target and addressee to monitor A's visual attention and adapt to it ongoingly (this section). These practices are deployed in context-sensitive ways; they reflect the context of their use and at the same time contribute to the emergence of that context. Absence of mutual monitoring may create, or sustain, problems that disturb the sequential order, threaten intersubjectivity, and lead to extended repair sequences (extract 3). However, absence of mutual monitoring does not necessarily cause problems, as will be shown in the last section.

[4]Concurrently, C transforms her pointing gesture into an open hand palm up shape (OHPU) (Kendon, 2004). A discussion of different gesture shapes is beyond the scope of this paper (see, however, Stukenbrock, 2015, p. 97–230).

Extract 6: "Gambling Table" (SM01_Saal_00:01:33:12)

```
11    C-vb    <<p>voiLÀ;>
                  right
      C-gz    --T------->
      T-gz    --C------->

12            =also (.) LÄUFT alles;
              PTCL everything is working
      C-gz    --|......|A|.............>
      T-gz    --|.......................>

13    C-vb    okAY? (.)
              okay
      C-gz    .......>
      T-gz    ..|A--->
```

Figure 15, right: C-gaze projecting a domain of pointing

```
14    T-vb    MAchen wir da[s;]
              let's do that
      T-gz    -A---------------|
15    C-vb                   [voi]LÀ; °hh (.)
                              right
      C-gz    ............|--DP----------->
      C-ge                             .....>
```

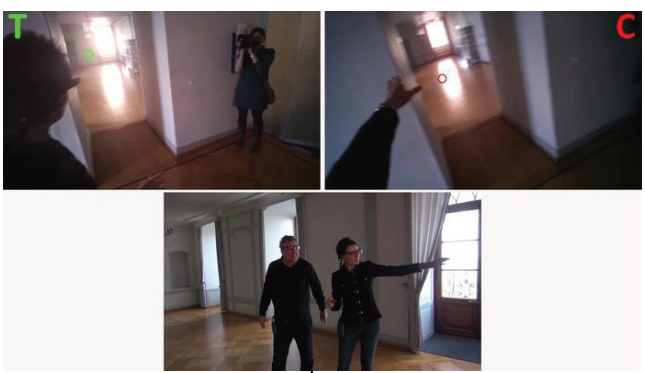


Figure 16, left: T-gaze at domain of scrutiny; right: C-gaze at target and pointing gesture

```
16    C-vb    und HIER,
              and here
      C-gz    --DP-----|
      C-ge    ...|-PG->
```

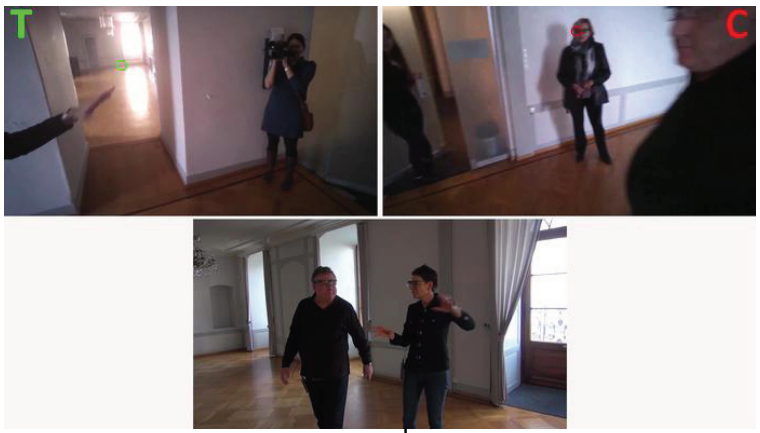


Figure 17, right:A-gaze monitoring by speaker C

```
17    C-vb    HIER können sie dAnn:- (.)
              here could  you then
      C-gz    ......|--A-----|.........>
      C-ge    --------|,,,,,,,.|-OHPA-->
```

Figure 18, right: A-gaze monitoring for turn-taking and response mobilization

```
18    C-vb    die' [die  ] SPIELtische auf[bauen;]
              set up the the gaming tables
19    A-vb         [(ab')]                [ABso]lut;
                     ab                    absolutely
      C-gz    ...........|--A------------------------>
      C-ge    -freezes OHPA-------------------------->
20    A-vb    =un_wie sIe [SEhen-]
               and as you can see
21    C-vb                [oKAY; ]
                           okay
      C-gz    --A----------------->
```

Absence of Mutual Gaze Monitoring and the Establishment of Joint Attention

The endogenous organization of a range of social activities emerges in ways that do not invite mutual gaze monitoring, contextual factors making them either unnecessary or preventing them. The examples in this section illustrate that particular spatial configurations and participation frameworks account for variations and non-occurrence of the gaze practices observed so far.

The following example (Extract 7: "Magic Robot") demonstrates a typical case in which contextual factors contribute to the absence of the gaze pattern described in the

Extract 7: "Magic Robot" (SM02_Rundgang_00:02:30)

```
01          (1.1)
    M-gz    -MR--|
02  M-vb    <pp, whispering>> magic RObot;>
    M-gz    --scanning exhibits----------->
03          (1.0) (0.4)
    M-gz    -----|-MR--
```

Figure 21: simultaneous joint attention

```
04  A-vb    mi vater het SAUviel spiel [bi sich dehei;
            my father has very many games at home
05  M-vb                               [lueg DAS han ich au schomol GSEH;
                                        look I have already seen this before
    M-gz    -MR------|,.|--MR-------------------------------------|~~~~~~~~>
    M-ge                                ..............|----PG to MR----,,,,,
    A-gz    --scanning exhibits-----------------|..........|-MR------------>
06          (0.3)
    M-ge    ,,,,,,,
    M-gz    ~~~~~~>
    A-gz    -MR--->
07  A-vb    magic RObot,
    A-gz    -MR--------|
    M-gz    ~~~~~~~~~~~>
08          (0.2)(0.15)(2.0)(0.15)
    A-gz    ~~~~~|-MR-------|~~~~~
    M-gz    ~|-MR----|~~~~~~~~~~~~
09  A-vb    das isch (.) das seit [mer jetzt irgendwie GAR nüt;=
            that is that doesn't tell me PTCL PTCL anything
10  M-vb                           [aber nur im (.)
                                    but only on
11  A-vb    [<<laughingly>=Aber hehe
                          but  hehe
12  M-vb    [im INternet;
             on the internet
13  A-vb    es seht nit SCHLÄCHT us;>
            it doesn't look too bad
```

section on "Deictic Reference, Addressee Gaze Monitoring, and Joint Attention." Two friends, M and A, are at the beginning of a tour through the Swiss Museum of Games. M is the pointing participant. In the transcript, M's perspective is displayed on the left (green cursor), A's perspective is on the right (red cursor). We join them on their way along a corridor with display cases showing a large array of games. They take small steps from case to case, pausing from time to time. The spatial configuration between mobile and stationary phases differs in a slight, but significant way. Whereas the participants' bodies move into an oblique front-to-back orientation during mobile phases, with A taking the lead (**Figure 19**), they get into a side-by-side configuration and establish a lateral interactional space (**Figure 20**) in stationary phases.

The side-by-side configuration, while displaying togetherness, co-orientation, and readiness for attention-sharing, allows them on the one hand to face out of the interactional space centrifugally to establish divergent foci of attention, and, on the other hand, to face inside and arrange themselves "in such a way that their individual transactional segments overlap to create a joint transactional space" (Kendon, 1990, p. 211). The spatial overlap between the participants' transactional

FIGURE 19 | Front-to-back configuration.

FIGURE 20 | Side-by-side configuration, facing inwards.

segments, termed "o-space" (Kendon, 1990, p. 211), depends on how far they orient centripetally, i.e., toward the space projected by the coparticipant's body. The referential act (l. 5) occurs after the participants have stopped and A has started talking (l. 4).

M does not engage in addressee monitoring either before or after the referential act. For the following reasons, it can be dispensed with in the local context: First, the participants are in close side-by-side configuration, slightly turned inwards with overlapping transactional segments. Peripheral vision thus gives them mutual access to one another. Second, focused interaction (job 1) has been re-established verbally by A; M thus knows that A is co-oriented with him. Third, the fact that A engages in displaced talk about her father (l. 4) does not create a visual conflict of interest with the situated noticing initiated by M even though the latter is interruptive. Fourth, the temporal design of the noticing as cutting into A's talk functions as an attention-getting device. Together with the imperative (l. 5: LUEG/"look") and the demonstrative (l. 5: DAS/"this"), it serves to underpin the primacy of situated vs. displaced speech and legitimizes the interruption. A aligns and shifts her gaze to the target (jobs 6 and 7) shortly after M's gesture becomes visible to her (l. 5, **Figure 21**, right). M and A briefly look at the target simultaneously (**Figure 21**, left and right)[5] before M returns to scanning the other exhibits while A continues looking at the target (l. 5–7). When A displays her understanding (job 10) by denominating the referent (job 8) with a proper name (l. 7: magic RObot), M shifts his gaze back to the game, thus creating another moment of joint attention. Visual attention sharing is followed by further talk about the referent. Thus, the establishment of joint attention is mutually known and integrated in the participants' common ground.

None of the meta-perceptive gaze practices described above contributes to visual attention sharing. Instead, a high degree of implicitness is involved. The participants rely on their close side-by-side position, a tacit agreement about visual co-orientation, and – instead of mutual monitoring – on reciprocal inferences about each other's visual perception. Significantly, part of the inferential work done by A is anchored in overhearing M talking to himself shortly before the verbal exchange starts. While still on the move, M has whispered "magic RObot" to himself (l. 2), indexing an individual discovery he will present as a noticeable to be shared with his coparticipant a few moments later (l. 5). Overhearing a coparticipant's subdued self-talk may enhance common ground in subtle ways. It resembles acts of unperspicuously observing the other's visual perception. In more general terms, the overhearing episode exemplifies how participants' interactional microhistories create intersubjectivity and contribute to the common ground on various levels in explicit and implicit ways.

The last example (Extract 8: "Flipper") from the Swiss Museum of Games represents another instance in which addressee gaze monitoring is dispensed with. P's perspective is on the right (red cursor), and A's perspective is on the left (green cursor). The spatial configuration is the same as in the previous example; it facilitates mutual monitoring and attention sharing by overlapping segments of the visual field. In contrast to the previous extract, the sequential implementation of the referential action is different. The analysis highlights that, beyond particular bodily configurations, the temporal design and type of social action also contribute to a context in which addressee monitoring is dispensed with. In contrast to the attention-getting devices and the type of social action performed in the previous example (interruption, verb of perception in the imperative, noticing), P offers an assessment. Concurrently, P uses a demonstrative and a pointing gesture to direct A's visual attention to the target game. When A's gaze arrives at the target (**Figure 22**, left), P briefly looks elsewhere (**Figure 22**, right) before shifting her gaze back to the game. An extended phase of simultaneous joint attention ensues (l. 3–5). The referential act is part of an assessment (l. 1: witzig/"funny") followed by giggling (l. 2). A responds minimally by chuckling (l. 3) and denominating the referent (l. 4: FLIPper), thus displaying successful identification of target and referent. Again, there is no A-gaze monitoring before or after the referential act. Instead, the participants rely implicitly on visual co-orientation and interpersonal coordination; mutual monitoring is dispensed with.

To sum up, the context-related factors that contribute to the endogenous organization of deictic reference and joint attention without mutual gaze monitoring are particular bodily configurations in (semi-)mobile activities. While "an eye-to-eye ecological huddle" (Goffman, 1963, p. 95) is not invited by the activity, participants display to each other bodily and/or verbally that they are together, closely attending to the activity at hand. Thus, individual perceptions can be transformed into shared perception with a minimal array of resources and in sequentially reduced formats.

[5]For reasons of space, the external camera perspective has been cut out in **Figure 21**.

Extract 8: "Flipper" (SM02_Rundgang_00:03:18)

```
01          (2.0)
      P-gz    -game -----------------      >
```

Figure 22, left: A-gaze at target; right: P-gaze not at target

```
02  P  -vb  dAs seht au WIT      zig   us,
             that looks also funny
      P-gz    -game -----  |~~~~~~~~~~    ~|
      P-ge  ............|      -point  -|,,
      A-gz  ~~~~~|.......|       -game  -|~~
03  P  -vb  ((giggles))
      P-gz    -game ------  >
      A-gz    -game ------  >
04  A  -vb  ((chuckles subduedly))
      A-gz    -game -----------------      >
      P-gz    -game -----------------      >
05  A  -vb  <<pp, whispering>FLIPper;>
      A-gz    -game ---------------------
      P-gz    -game --------------     |~~~~~~
```

DISCUSSION

Based on video data recorded with an external camera and mobile eye tracking glasses worn by participants in naturally occurring social activities, this paper investigated the function of gaze practices deployed concurrently with deictics and embodied pointing to establish joint attention as a mutually known interactional achievement. The focus of the analyses was on meta-perceptive gaze practices that participants themselves oriented to in the course of demonstrative reference, notably, *speaker gaze following* and *addressee gaze monitoring*. The analysis drew on a model of deictic reference which specified the jobs participants have to fulfill sequentially according to their roles as referring participant (P) and addressee (A) in order to establish joint attention and mutual understanding. The jobs were conceived of as participants' routine solutions to the problem of reference and attention sharing in everyday life.

The methodological challenge of this study consisted of acquiring precise and ecologically valid eye gaze in order to meet with the conversation analytic requirement of preserving the endogenous order of social activities within the context of their occurrence (Mondada, 2013). This was achieved by taking eye tracking technology out of the laboratory to record participants "in the wild." Mobile eye tracking technology delivered detailed eye gaze data on the participants' gaze practices without restricting their bodily freedom. Sequential analysis uncovered the participants' meta-perceptive gaze practices within two activity contexts, shopping at a market and visiting a museum. The analysis revealed that meta-perceptive gaze practices are context dependent, positionally sensitive, tied to the participant roles of P and A, and temporally fine-tuned to the stream of verbal and embodied conduct. It was shown that the temporal placement of these gaze practices with respect to the jobs that participants fulfill in the course of demonstrative

reference shape the functions that meta-perceptive gaze acquires in that process.

In the eye tracking data, gaze following appeared as an act of looking at another person's gaze and following it to a target within a domain of scrutiny. As such, it has three elements: (1) a starting point, the other's eyes; (2) an end point, the presumed focus of the other's visual attention; and (3) the trajectory from starting to end point, or from the other's eyes to the target. This description is in line with conceptualizations of gaze following in developmental studies (Scaife and Bruner, 1975; Flom et al., 2007; Brooks and Meltzoff, 2008, 2014). It lacks, however, an account of the temporality, interactivity, and intersubjectivity of gaze following and joint attention within ordinary activities. The sequential analysis (section "Deictic Reference, Speaker Gaze Following, and Joint Attention") revealed that, in naturally occurring interaction, gaze following constitutes a complex and heterogeneous phenomenon. It is not initiated by an experimenter who shifts gaze for a predetermined gaze follower to follow. Instead, it emerges online as participants jointly engage with the world, notice things, and notice that their coparticipants notice things. Instead of extrapolating a vector from the other's gaze, as the metaphor "to follow a person's the line of regard" suggests, it is a social, interpretive act based on the participants' mutual assumptions, their involvement in the ongoing activity, and the interactional microcontext. It was argued that the concept of gaze following needs refinement when applied to the investigation of ordinary activities. Gaze following was therefore defined as an interactional gaze practice of tapping into a coparticipant's gaze and exploiting it as a resource to gather information on *where*, *how*, and for *how long* he or she is looking, and to infer *what* he or she is looking at, and *why*. The inferences that gaze followers drew on their coparticipants' acts of seeing were socially displayed in how they designed their own next actions and documented understanding of the speaker's prior turn.

In contrast to speaker gaze following, addressee gaze monitoring does not induce a pointing participant to follow the addressee's line of regard. Although P may shift gaze from A back to the target (e.g., extracts 1 and 2), or look elsewhere (e.g., extract 5), gaze shifts from A to the target do not constitute instances of gaze following, since P, instead of constituting a new focus of attention, only revisits a target previously established by him- or herself. The sequential analysis (section "Deictic Reference, Addressee Gaze Monitoring, and Joint Attention") documented that addressee gaze monitoring occurs in the course of, or immediately at the end of demonstrative reference, i.e., before speaker change takes place. It was argued that addressee gaze monitoring is an interactional resource for P to gather real-time evidence on whether joint attention is emerging, and to incrementally add material when they anticipate that intersubjectivity be threatened. Although addressee monitoring is an important means to maintain intersubjectivity in the course of a demonstrative act, it can also be dispensed with (section "Absence of Mutual Gaze Monitoring and Joint Attention"). Mobile settings often complicate visual access to the other's eyes. In front-to-back and side-by-side configurations, participants often refrain from mutual gaze monitoring. This suggests that the accomplishment of joint attention in mobile, spatially fluid configurations brings about variations in the sequential format of jobs that participants have to accomplish and leads to an absence of gaze practices regularly observable in face-to-face and L-configurations.

This paper has presented qualitative analyses of meta-perceptive practices in two settings. In order to fully understand gaze following and gaze monitoring in naturally occurring interaction, further studies should investigate the occurrence and non-occurrence of these practices in a range of different settings, social activities, and participation frameworks, and take into consideration alternative practices such as eye–hand coordination (Yu and Smith, 2013) which help understand the affordances and constraints that account for participants' local preferences. Reliable eye tracking data of activities in their natural habitat are needed to build collections of cases that are large enough to uncover the systematicity of context-dependent variations and serve to further develop and refine the interactional model of deictic reference and joint attention. In its current design, the framework offers a high degree of granularity, thus enabling detailed analyses of the interactional jobs that participants need to accomplish in order to establish deictic reference. By accounting for context-dependent variations, it allows for the description of more and less elaborate sequences and formats as well as for a distinction between jobs (e.g., directing the coparticipant's attention to an object) and resources used to accomplish those jobs (manual pointing, whole body or eye gaze orientation, etc.). In future studies, the usability of the model could be further verified with respect to child development research, and its adaptability to the study of reference and joint attention in atypical interaction, technologically mediated communication, and human robot interaction could be assessed.

DATA AVAILABILITY STATEMENT

The video and eye tracking recordings for this study are not publicly available because of restricted informed consent of the participants.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AS developed the theoretical framework, collected the data with her team, and carried out the empirical analysis.

FUNDING

This research was funded by the Swiss National Science Foundation (SNSF) as part of the project 179108 "Deixis and Joint Attention: Vision in Interaction" (DeJA-VI).

ACKNOWLEDGMENTS

I would like to thank the participants of the eye tracking recordings and the members of my project team, in particular Stefanie Lazaro, Letizia Manco, and Sonja Salerno, for data collection and preparation. I am also indebted to the Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg/Br., for support of my previous research on deixis and mobile eye tracking.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01779/full#supplementary-material

REFERENCES

Auer, P. (2020). "Die Struktur von Redebeiträgen und die Organisation des Sprecherwechsels," in *Einführung in die Konversationsanalyse*, eds K. Birkner, P. Auer, A. Bauer, and H. Kotthoff (Berlin: De Gruyter), 115–245.

Brooks, R., and Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *J. Child Lang.* 35, 207–220. doi: 10.1017/s030500090700829x

Brooks, R., and Meltzoff, A. N. (2014). "Gaze following: a mechanism for building social connections between infants and adults," in *Mechanisms of Social Connection. From Brain to Group*, eds M. Mikulincer and P. R. Shaver (Washington DC: American Psychological Association), 167–183. doi: 10.1037/14250-010

Bühler K. (1965 [1934]). *Sprachtheorie. Die Darstellungsfunktion der Sprache*, 2nd Edn. Stuttgart: Gustav Fischer Verlag.

Butterworth, G. (2003). "Pointing is the royal road to language for babies," in *Pointing. Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Erlbaum), 9–33.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 4, 1–166. doi: 10.1515/9783110802047-fm

Clark, A. T., and Gergle, D. (2010). "Effects of shifting spatial context on referential form," in *Proceedings of the 20th Annual Meeting of the Society for Text and Discourse*, Chicago, IL. doi: 10.17250/khisli.36..201909.001

Clark, E. V. (1978). "From gesture to word: on the natural history of deixis in language acquisition," in *Human Growth and Development*, eds J. Bruner and A. Garton (Oxford: Oxford University Press), 85–120.

Clark, H. H. (1996). *Using Language*. Cambridge, MA: Cambridge University Press.

Clark, H. H., and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *J. Mem. Lang.* 50, 62–81. doi: 10.1016/j.jml.2003.08.004

Clark, H. H., and Marshall, C. R. (1981). "Definite reference and mutual knowledge," in *Elements of Discourse Understanding*, eds A. K. Joshi, B. L. Webber, and I. A. Sag (Cambridge, MA: Cambridge University Press), 10–63.

Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *J. Verbal Learn. Verbal Behav.* 22, 245–258. doi: 10.1016/s0022-5371(83)90189-5

Cooperrider, K., and Núñez, R. (2012). Nose-pointing: notes on a facial gesture of Papua New Guinea. *Gesture* 12, 103–129. doi: 10.1075/gest.12.2.01coo

Couper-Kuhlen, E., and Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge, MA: Cambridge University Press.

De Stefani, E. (2010). "Reference as an interactively and multimodally accomplished practice. Organizing spatial reorientation in guided tours," in *Spoken Communication*, eds M. Pettorino, A. Giannini, I. Chiari, and F. Dovetto (Newcastle upon Tyne: Cambridge Scholars Publishing), 137–170.

De Stefani, E., and Gazin, A.-D. (2014). Instructional sequences in driving lessons: mobile participants and the temporal and sequential organization of actions. *J. Pragmat.* 65, 63–79. doi: 10.1016/j.pragma.2013.08.020

De Stefani, E., and Mondada, L. (2014). Reorganizing mobile formations: when "guided" participants initiate reorientations in guided tours. *Space Cult.* 17, 157–175. doi: 10.1177/1206331213508504

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Enfield, N. (2001). Lip-pointing. *Gesture* 1, 185–212. doi: 10.1075/gest.1.2.06enf

Eriksson, M. (2009). Referring as interaction: on the interplay between linguistic and bodily practices. *J. Pragmat.* 41, 240–262. doi: 10.1016/j.pragma.2008.10.011

Fillmore C (1997 [1971]). *Lectures on Deixis*. Stanford, CA: University of Chicago Press.

Fillmore, C. J. (1982). "Towards a descriptive framework for spatial deixis," in *Speech, Place, and Action. Studies in Deixis and Related Topics*, eds R. J. Jarvella and W. Klein (Chichester: Wiley), 31–60.

Flom, R., Lee, K., and Muir, D. (eds) (2007). *Gaze-Following. Its Development and Significance*. Mahwah, NJ: Lawrence Erlbaum.

Franco, F., and Butterworth, G. (1996). Pointing and social awareness: declaring and requesting in the second year. *J. Child Lang.* 23, 307–336. doi: 10.1017/s0305000900008813

Gergle, D., and Clark, A. T. (2011). "See what I'm saying? Using dyadic mobile eye-tracking to study collaborative reference," in *Proceedings of Conference on Computer Supported Cooperative Work*, Hangzhou, 435–444.

Goffman, E. (1963). *Behavior in Public Places. Notes on the Social Organization of Gatherings*. New York, NY: The Free Press.

Goffman, E. (1981). *Forms of Talk*. Philadelphia, PA: University of Pennsylvania Press.

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociol. Inq.* 50, 272–302. doi: 10.1111/j.1475-682x.1980.tb00023.x

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. London: Academic Press.

Goodwin, C. (1994). Professional vision. *Am. Anthropol.* 96, 606–633.

Goodwin, C. (1996). "Transparent Vision," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge, MA: Cambridge University Press).

Goodwin, C. (2000). Pointing and the collaborative construction of meaning in Aphasia. *Texas Linguist. Forum* 43, 67–76.

Goodwin, C. (2003). "Pointing as situated practice," in *Pointing. Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Erlbaum), 217–241.

Goodwin, C. (2007). "Environmentally coupled gestures," in *Gesture and the Dynamic Dimension of Language. Essays in Honor of David McNeill Gesture Studies*, eds S. D. Duncan, J. Cassell, and E. T. Levy (Amsterdam: Philadelphia), 195–212. doi: 10.1075/gs.1.18goo

Haddington, P., Mondada, L., and Nevile, M. (eds) (2013). *Interaction and Mobility. Language and the Body in Motion*. Berlin: De Gruyter.

Hanna, J. E., and Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *J. Mem. Lang.* 57, 596–615. doi: 10.1016/j.jml.2007.01.008

Hausendorf, H. (2003). "Deixis and speech situation revisited. The mechanism of perceived perception," in *Deictic Conceptualisation of Space, Time and Person*, ed. F. Lenz (Amsterdam: John Benjamins), 249–269. doi: 10.1075/pbns.112.13hau

Heath, C. (1985). The consultation's end: the coordination of speech and body movement. *Int. J. Sociol. Lang.* 51, 27–42.

Heller, V. (2019). Embodied displacements in young children's storytelling: layering of spaces, voices, bodies. *Res. Children Soc. Interact.* 3, 168–195. doi: 10.1558/rcsi.37311

Heller, V., and Rohlfing, K. (2017). Reference as an interactive achievement: sequential and longitudinal analyses of labeling interactions in shared book reading and free play. *Front. Psychol.* 8:139. doi: 10.3389/fpsyg.2017.00139

Heritage, J. (2007). "Intersubjectivity and progressivity in person (and place) reference," in *Person Reference in Interaction. Linguistic, Cultural, and Social Perspectives*, eds N. Enfield and T. Stivers (Cambridge, MA: Cambridge University Press), 255–280. doi: 10.1017/cbo9780511486746.012

Himmelmann, N. P. (1997). *Deiktikon, Artikel, Nominalphrase. Zur Emergenz Syntaktischer Struktur*. Tübingen: Niemeyer.

Hindmarsh, J., and Heath, C. (2000). Embodied reference: a study on deixis in workplace interaction. *J. Pragmat.* 32, 1855–1878. doi: 10.1016/s0378-2166(99)00122-8

Jermann, P., and Nüssli, M.-A. (2012). "Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, Seattle: ACM Press, 1125.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 1, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendon, A. (1988). "How gesture can become like words," in *Cross-Cultural Perspective in Non-Verbal Communication*, ed. F. Poyatos (Toronto: C. J. Hogrefe), 131–141.

Kendon, A. (1990). *Conducting Interaction. Patterns of Behavior in Focused Encounters*. Cambridge, MA: Cambridge University Press.

Kendon, A. (2004). *Gesture. Visible Action as Utterance*. Cambridge, MA: Cambridge University Press.

Kendon, A., and Versante, L. (2003). "Pointing by hand in neapolitan," in *Pointing. Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Erlbaum), 109–138.

Kita, S. (ed.) (2003). *Pointing. Where Language, Culture, and Cognition Meet*. Mahwah, NJ: Lawrence Erlbaum.

Land, M. F. (2006). Eye movement and the control of actions in everyday life. *Prog. Retinal Eye Res.* 25, 296–324. doi: 10.1016/j.preteyeres.2006.01.002

Levinson, S. C. (2004). "Deixis," in *The Handbook of Pragmatics Blackwell Handbooks in Linguistics*, eds L. R. Horn and G. L. Ward (Malden, MA: Blackwell), 97–121.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., and Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Dev. Sci.* 7, 297–307. doi: 10.1111/j.1467-7687.2004.00349.x

Louwerse, M. M., and Bangerter, A. (2005). Focusing attention with deictic gestures and linguistic expressions. *Proc. Annu. Meet. Cogn. Sci. Soc.* 27, 1331–1336.

Louwerse, M. M., and Bangerter, A. (2010). Effects of ambiguous gestures and language on the time course of reference resolution. *Cogn. Sci.* 34, 1517–1529. doi: 10.1111/j.1551-6709.2010.01135.x

Luhmann, N. (1984). *Soziale Systeme. Grundriß Einer Allgemeinen Theorie*. Frankfurt am Main: Suhrkamp.

McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (ed.) (2000). *Language and Gesture*. Cambridge, MA: Cambridge University Press.

Mondada, L. (2005). La constitution de l'origo déictique comme travail interactionnel des participants: une approche praxéologique de la spatialité. *Intellectica* 41, 75–100. doi: 10.3406/intel.2005.1722

Mondada, L. (2007). "Interaktionsraum und Koordinierung," in *Koordination.Analysen zur Multimodalen Interaktion Studien zur Deutschen Sprache*, ed. R. Schmitt (Tübingen: Narr), 55–93.

Mondada, L. (2012). "Deixis: an integrated interactional multimodal analysis," in *Interaction and Usage-Based Grammar Theories. What about Prosody and Visual Signals?*, eds P. Bergmann and J. Brenning (Berlin: De Gruyter), 173–206.

Mondada, L. (2013). "The conversation analytic approach to data collection," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Chichester: Wiley), 32–56. doi: 10.1002/9781118325001.ch3

Mondada, L. (2014). "Pointing, talk and the bodies: reference and joint attention as embodied interactional achievements," in *From Gesture in Conversation to Visible Utterance in Action*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins), 95–124. doi: 10.1075/z.188.06mon

Mondada, L. (2015). "Multimodal completions," in *Temporality in Interaction*, eds A. Deppermann and S. Günthner (Amsterdam: John Benjamins), 267–307. doi: 10.1075/slsi.27.09mon

Mondada, L. (2019). *Conventions for Multimodal Transcription*. Available online at: https://www.lorenzamondada.net/multimodal-transcription (accessed February 12, 2020).

Moore, C., and Dunham, P. J. (eds) (1995). *Joint Attention. Its Origins and Role in Development*. Hillsdale, NJ: Erlbaum.

Pomerantz, A. (1984). "Agreeing and disagreeing with assessments," in *Structures of Social Action.Studies in Conversation Analysis Studies in Emotion and Social Interaction*, eds J. M. Atkinson and J. Heritage (Cambridge, MA: Cambridge University Press), 57–101. doi: 10.1017/cbo9780511665868.008

Povinelli, D. J., and Davis, D. R. (1994). Differences between chimpanzees (Pan troglodytes) and Humans (Homo sapiens) in the resting state of the index finger: implications for pointing. *J. Comp. Psychol.* 108, 134–139. doi: 10.1037/0735-7036.108.2.134

Rossano, F. (2012). *Gaze Behavior in Face-to-Face Interaction*. Doctoral dissertation. Available online at: http://hdl.handle.net/2066/99151 (Accessed April 16, 2020).

Rossano, F. (2013). "Gaze in interaction," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden, MA: Wiley-Blackwell), 308–329.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Scaife, M., and Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature* 253, 265–266. doi: 10.1038/253265a0

Schegloff, E. A. (2007). *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Cambridge, MA: Cambridge University Press.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). the preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382. doi: 10.2307/413107

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., and Birkner, K. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung* 10, 353–402.

Selting, M., and Couper-Kuhlen, E. (eds) (2001). *Studies in Interactional Linguistics. Amsterdam*. Philadelphia: John Benjamins.

Sherzer, J. (1973). Verbal and nonverbal deixis: the pointed lip gesture among the San Blas Cuna. *Lang. Soc.* 2, 117–131. doi: 10.1017/s0047404500000087

Sidnell, J., and Stivers, T. (eds) (2013). *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31.

Streeck, J. (1988). The significance of gesture: how is it established. *IPrA Pap. Pragmat.* 2, 60–83. doi: 10.1075/iprapip.2.1-2.03str

Streeck, J. (1993). Gesture as communication I: its coordination with gaze and speech. *Commun. Monogr.* 60, 275–299. doi: 10.1080/03637759309376314

Streeck, J. (2002). Grammars, words, and embodied meanings. *J. Commun.* 52, 581–596. doi: 10.1111/j.1460-2466.2002.tb02563.x

Streeck, J. (2009). Forward-gesturing. *Discourse Process.* 46, 161–179. doi: 10.1080/01638530902728793

Stukenbrock, A. (2008). "Wo ist der Hauptschmerz?" - Zeigen am eigenen Körper in der medizinischen Kommunikation. *Gesprächsforschung* 9, 1–33. doi: 10.1515/9783486849028-002

Stukenbrock, A. (2009). Referenz durch Zeigen: zur Theorie der Deixis. *Deutsche Sprache* 37, 289–315.

Stukenbrock, A. (2010). Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks "so.". *InLiSt Interact. Linguist. Struct.* 47, 1–23.

Stukenbrock, A. (2011). "Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks 'so,'" in *Beschreibungen für Gesprochenes Deutsch auf dem Prüfstand. Analysen und Perspektiven*, eds N. Dittmar and N. Bahlo (Frankfurt am Main: Peter Lang), 165–193.

Stukenbrock, A. (2015). *Deixis in der Face-to-Face-Interaktion*. Berlin: De Gruyter.

Stukenbrock, A. (2016). "Deiktische Praktiken: zwischen Interaktion und Grammatik," in *Sprachliche und kommunikative Praktiken*, eds A. Deppermann, H. Feilke, and A. Linke (Berlin: De Gruyter), 81–126.

Stukenbrock, A. (2018a). Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis: mobile Eye Tracking-Analysen zur Herstellung von joint attention. *Gesprächsforschung* 19, 132–168.

Stukenbrock, A. (2018b). "Forward-looking: where do we go with multimodal projections?," in *Modalities and Temporalities: Convergences and Divergences of Bodily Resources in Interaction*, eds A. Deppermann and J. Streeck (Amsterdam: John Benjamins), 31–68. doi: 10.1075/pbns.293.01stu

Stukenbrock, A. (2018c). "Mobile dual eye-tracking in face-to-face interaction: The case of deixis and joint attention," in *Advances in Interaction Studies*, eds G. Brône and B. Oben (Amsterdam: John Benjamins), 265–302. doi: 10.1075/ais.10.11stu

Stukenbrock, A., and Dao, A. N. (2019). "Joint attention in passing. What dual mobile eye tracking reveals about gaze in coordinating embodied activities at the market," in *Embodied Activities in Face-to-Face and Mediated Settings*, eds E. Reber and C. Gerhardt (Basingstoke: Palgrave Macmillan), 177–213.

Tomasello, M., and Call, J. (2019). Thirty years of great ape gestures. *Anim. Cogn.* 22, 461–469. doi: 10.1007/s10071-018-1167-1

Tomasello, M., Hare, B., Lehmann, H., and Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *J. Hum. Evol.* 52, 314–320. doi: 10.1016/j.jhevol.2006.10.001

Quine, W. O. V. (1960). *Word and Object*. Cambridge, MA: Massachusetts Institute of Technology Press.

Wilkins, D. (2003). "Why pointing with the index finger is not a universal (in sociocultural and semiotic terms)," in *Pointing. Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Erlbaum), 171–215.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genoa, 1556–1559.

Yu, C., and Smith, L. B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One* 8:e79659. doi: 10.1371/journal.pone.0079659 doi: 10.1371/journal.pone.0079659

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 06 November 2020
doi: 10.3389/fpsyg.2020.583763

Contribution of the Semiological Approach to Deixis–Anaphora in Sign Language: The Key Role of Eye-Gaze

Brigitte Garcia and Marie-Anne Sallandre**

Structures Formelles du Langage Laboratory, UMR 7023, Centre National de la Recherche Scientifique, University of Paris 8 – University Paris Lumières, Paris, France

We address the issue of deixis–anaphora in sign language (SL) discourse, focusing on the role of eye-gaze. According to the Semiological Approach, SL structuring stems from a maximum exploitation of the visuo-gestural modality, which results in two modes of meaning production, depending on the signer's semiotic intent. Involving both non-manual and manual parameters, the first mode, expressing the intent to *say while showing*, uses constructions based on structures, the termed "transfer structures." The second one, expressing the intent to *say without showing*, involves lexical, pointing and fingerspelling units. In order to situate our descriptive concepts with respect to those used by SL linguists who, like us, adopt a cognitive–functionalist perspective, we expose a specific theoretical foundation of our approach, the "enunciation theories." The concept of "enunciation" is decisive for understanding the role of eye-gaze, as being at the foundation of deixis and the key vector of referential creation and tracking in SL discourse. "Enunciation" entails the opposition between "Enunciation" and "Utterance" Domains. The first links, as co-enunciators, the signer/speaker and his/her addressee, establishing them by the very "act of enunciation" as 1st and 2nd person. The second is internal to the discourse produced. Grounding on corpora of narratives in several SLs (some with no historical link), we illustrate this crucial role of eye-gaze and the diversity of functions it fulfills. Our analyses, carried out in this perspective, attest to the multiple structural similarities between SLs, particularly with regard to transfer structures. This result strongly supports the typological hypothesis underlying our approach, namely, that these structures are common to all SLs. We thus show that an enunciative analysis, based on the key role of eye-gaze in these visual languages that are SLs, is able to give the simplest account of their own linguistic economy and, in particular, of deixis–anaphora in these languages.

Keywords: sign language, reference, deixis–anaphora, eye-gaze, enunciation, corpus, typology, highly iconic constructions

OPEN ACCESS

Edited by:
Olga Capirci,
Italian National Research Council, Italy

Reviewed by:
Sherman Wilcox,
University of New Mexico, United States
Sabina Fontana,
University of Catania, Italy

****Correspondence:***
Brigitte Garcia
brigitte.garcia@univ-paris8.fr;
bridge.garcia86@gmail.com
Marie-Anne Sallandre
msallandre@univ-paris8.fr

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *15 July 2020*
Accepted: *10 September 2020*
Published: *06 November 2020*

Citation:
Garcia B and Sallandre M-A (2020) Contribution of the Semiological Approach to Deixis–Anaphora in Sign Language: The Key Role of Eye-Gaze. Front. Psychol. 11:583763. doi: 10.3389/fpsyg.2020.583763

INTRODUCTION

Following Apothéloz and Pekarek Doehler (2003, 110), reference can be defined as "the relationship that language maintains with its external environment (whether it is called 'mental representation,' 'world,' or 'reality')", and the action of referring as drawing attention to an entity—of which the deixis is the vector par excellence. Here we address the issue of deixis and anaphora expression in

sign language (SL) discourse and the shared attention processes that underlie it (e.g., Lyons, 1977; Cornish, 2014; Sidnell and Enfield, 2016).

What is at stake is precisely to highlight the central and specific role of the interlocutors' eye-gaze (both the signer's and his/her addressee's) at the very basis of the deixis in these face-to-face languages. To explain the link between eye-gaze and deixis–anaphora in SL—a link that is, in our view, specific to SL—we must first describe our theoretical framework, known as the Semiological Approach (e.g., Cuxac, 1985, 1999, 2000; Fusellier-Souza, 2006, 2012; Cuxac and Sallandre, 2007; Cuxac and Antinoro-Pizzuto, 2010; Garcia and Sallandre, 2014). We focus in particular on our original modeling of highly iconic constructions, typically described in SL literature as "non-conventional" (encompassing terms such as *classifiers predicates/constructions*, *productive signs* or *depicting signs* on the one hand, and *role shifts*, *surrogates*, *enactments*, *constructed actions or dialogues* on the other), which contrast with "conventional units" (that is, lexical units, fingerspelling, and mouthing).

After a long domination of formalist approaches[1], the study of SL linguistics slowly began to diversify theoretically primarily in the 1990s, and more so in the 2000s (e.g., Stokoe, 1991; Engberg-Pedersen, 1993, 2003; Armstrong et al., 1995; Liddell, 1995, 2003; Wilcox and Wilcox, 1995; Johnston and Schembri, 1999, 2007; Pizzuto and Volterra, 2000; Woll and Sutton-Spence, 2005; Pizzuto et al., 2007). A growing number of SL linguists are adopting, as we are, a cognitive–functionalist perspective. In this context, following the work on ASL by Winston (1991, 1995), Metzger (1995), and especially by Liddell (2003), the non-conventional constructions mentioned, which are highly iconic and therefore for a long time kept on the margins of SL modeling, have aroused a strong revival of interest. Nowadays, these constructions are the object of numerous studies, especially with respect to their role in the expression of reference and referential cohesion. Yet, as shown below, highly iconic constructions have been at the heart of the Semiological Approach from its inception, where they were described early on as part of the set known as Transfer Structures (Cuxac, 1985, 1999). They have been shown to play a central role in doing reference and for referential cohesion, in LSF (Cuxac, 1999, 2000; Sallandre, 2003; Jacob, 2007; Garcia and Sallandre, 2014), in LIS (Pizzuto, 2007), and in other SLs, considered from a comparative typological perspective (Pizzuto et al., 2008; Sallandre, 2014; Sallandre et al., 2016).

However, a difficulty is that, beyond a number of proximities, there are significant discrepancies between our respective ways of segmenting and analyzing these non-conventional constructions, i.e., between our "transfer units" and particularly the constructions that are described as "depicting signs" on the one hand, and "constructed actions/dialogues" (CA/CD) on the other. A theoretical dimension of our approach helps to explain these specificities, namely, the fact that we have opted, from the outset, for "enunciation theories," a conception of language that developed in Europe in the 1960s and 1970s—and more particularly in France (e.g., Jakobson, 1957; Benveniste, 1966, 1970, 1974; Ducrot, 1972, 1980; Lyons, 1977; Culioli, 1990, 1995). We will therefore recall the main lines of this conception.

We begin with an overview presenting some of the central aspects in the study of reference both in spoken language (SpL) and in SL. We then present our theoretical framework, the Semiological Approach. We next underline and illustrate its typological scope through twelve SLs. Finally, we expose the contribution of the European "enunciation theories" which are inherent to our approach and we deem particularly appropriate for understanding the key role played by the eye-gaze in the expression of SL deixis and more generally in the creation of reference and maintenance of referential cohesion in SL.

REFERENCE, DEIXIS, AND ANAPHORA: BACKGROUND

The linguistic literature on reference and on the resources used by languages to introduce, maintain, and reintroduce an entity is obviously considerable, both for SpL and SL, and cannot be fully presented here. For a long time, however, the work has only concerned SpLs, which moreover were seen as monomodal and, initially at least, in their written form.

Reference in Spoken Language

A significant part of the discussion on reference in SpL has focused on the respective limits of deixis and anaphora and whether the distinction between these two major sources of reference is relevant (for an overview, see for example Apothéloz and Pekarek Doehler, 2003; Lombardi Vallauri, 2007), the deixis having first been thought to refer to an entity in the extralinguistic context (exophoric reference) while the anaphora would refer to an entity already introduced in the text/discourse (endophoric reference), referred to as its antecedent. Many authors (among them already Lyons, 1977) have in fact shown that in many cases it is difficult to identify such an antecedent and even argued that the addressee can reach the intended anaphoric interpretation without having to identify an antecedent (e.g., Cornish, 1990; Apothéloz, 1995; Croft, 2003). Rather, the addressee should rely on the representation he developed of the referent, which is often difficult to classify as based on prior discourse (the text) or on external context (exophoric). After Fillmore (1975), Levinson (1983) proposed a distinction between two deictic uses: *gestural deixis*, whose interpretation necessarily requires the establishment of physical links with the communicative situation, and *symbolic deixis*, whose interpretation requires only "the knowledge of the basic spatio-temporal parameters of the speech event, of the participants' roles, their social relationships, and some notions about the preceding discourse" (Lombardi Vallauri, 2007, 312). However, it often proves irrelevant to distinguish the (symbolic) deixis from the anaphora on the basis that deixis would refer to the introduction of the referent, insofar as the latter can be salient in the universe of discourse even if it is not physically present in the situation. In other words, attention is drawn to the difficulty of dissociating extralinguistic context and

[1] For a historical perspective of SL linguistics, see in particular Newport and Supalla (2000), Vermeerbergen (2006), and, for a French perspective, Garcia (2010), Garcia under review.

"text"—together contributing to the "universe of discourse"—and, thus, of distinguishing deixis and anaphora. This would explain why languages use the same resources in both processes, as mentioned by Lyons (1977), Cornish (1990), and Apothéloz (1995). This is the conclusion reached by Lombardi Vallauri (2007, 334), which we will adopt here, motivating our choice to use the undifferentiated term *deixis–anaphora*.

Models have thus moved from the traditional textual approach focusing on the notion of antecedent, to cognitive approaches opening to the concepts of degrees of salience, informational, memorial, and inferential mechanisms, with increasing consideration of mental representations constructed by and from discourse. According to this cognitive–informational paradigm, anchored in a functional conception of language, the mental representations of the interlocutors are dynamic, resulting from the ongoing discourse, the context, and the shared knowledge. The focus is therefore on the search for what conditions the choice of one or another referential expression, this choice being rather considered as relative to the cognitive status of the referents (to their activation status). This concept has led to various models, notably Accessibility Theory, developed by Ariel (1988, 1990). According to Ariel, the speaker chooses a referential expression depending on the degree of accessibility (cognitive or memory status) he/she assumes the addressee attributes to the referent. The lower the referent's accessibility, the higher the informational content provided by the chosen referential expression (and thus its "phonological weight"), and vice versa. Thus, markers of high accessibility include clitic pronouns, unstressed pronouns, and zero forms, while definite NPs and proper names, with descriptive content and higher phonological weight, are analyzed as markers of referents with low accessibility.

These cognitive–informational approaches are highly relevant in the study of reference today, regardless of language type. The main criticism against them involves the nature of the data, namely, monolog (rather than interactional) sequences, tending to narrative and written texts. Various authors (mostly in SpL linguistics) have shown that with the addition of data from ecological corpora of face-to-face interactions, factors such as cognitive accessibility of the referent and their attentional status are no longer sufficient to account for the choice of one referential expression over another (for an overview of this issue, see Apothéloz and Pekarek Doehler, 2003). We will not elaborate on this point, for lack of space, and given that the data that SL linguists are working on (including ourselves) are mainly monologs and narratives as well. However, we raise this issue as a point of consideration, all the more so, given that SLs are quintessentially face-to-face languages.

Reference in Sign Language

For SLs, which were introduced into the linguistic discipline much more recently, the study of reference followed the epistemological evolution described in the introduction. Initially seeking to find similarities with SpL, SL studies most often described processes of a nominal (lexical) and pronominal nature—the latter corresponding to a series of visual indexes considered as analogous to SpL pronouns—but also spatial modifications of certain predicates, which were likened to verbal person inflection.

Regarding the reference to an entity present in the situation, the manual pointing (closed fist, index finger fully or partially extended) is very early analyzed in the SLs studied as the main means of creating a reference in a personal (1st, 2nd, or 3rd person) or demonstrative pronoun function (e.g., Friedman, 1975; Deuchar, 1984; Johnston, 1991; Engberg-Pedersen, 1993; Sutton-Spence and Woll, 1999; Cuxac, 2000). Another common description concerns transitive predicates characterized by a directional movement between their two points of articulation (*directional/agreement verbs*). The distinction of persons can be signaled by these spatial modifications, marking a location close to the signer for the 1st person, close to the addressee for the 2nd person, or close to another referent present for the 3rd person. Concerning the reference to entities and events not present in the situation, several processes have been identified too across various SLs: the lexical sign referring to the entity is associated with a point in the signing space (termed "locus") either by modification of its own location (direct spatialization of the sign) or through the creation of this locus via manual pointing, often accompanied by a look at the locus. The locus therefore marks the entity and can be reactivated by the same processes (spatialization of the lexical sign, manual pointing, and possibly the eye-gaze).

However, alongside these nominal and pronominal processes and the use of space, SL linguists soon noted that "non-conventional constructions" also played a role for doing reference (e.g., from various perspectives, Kegl, 1976; Wilbur, 1979; Pizzuto, 1986; Collins-Ahlgren, 1990; Brennan, 1992). The first major type of such constructions was described under the initial designation *classifier constructions* or *classifiers predicates* (Supalla, 1978). The abundant literature on these constructions begins with Frishberg (1975) and Kegl and Wilbur (1976), who mention manual handshapes that vary depending on salient properties of the referents. The study of "handshapes classifiers" is extended to the study of the complex manual constructions of which they are part and which are described as differing from conventional (lexical) signs: the handshape is supposed to refer iconically to a particular class of referents, and the movement represents iconically either the displacement and/or location of the entity, or the way in which the entity is handled, or the size or contours of the shape, by tracing. Those manual constructions are subsequently the object of multiple debates under a wide range of terms impossible to reproduce here (for an overview, see Schembri, 2003).

During the 1980s, another type of construction was described in which the signer takes on the role of one of the discourse entities and which is defined as the privileged means of expression of reported discourse. This phenomenon was first termed "role shifting" following Mandel (1977), then more commonly known as "role shift." Role shift, described as marked mainly by a movement of the shoulders and by facial expression, is primarily analyzed, within formal approaches, as allowing the expression of a change in point of view, implying changes to the frame of reference, and thus a rearrangement of the loci associated with the referents (e.g., Padden, 1986, 1990; Lillo-Martin and Klima, 1990). With the exception of Mandel and

DeMatteo (in Friedman, 1977), the consequences of this "shift" on the referential framework constitute all that is said of these constructions in the literature until the early 1990s. However, starting in the 1990s, some authors, working on corpora of narratives and generally hostile to formalist approaches, broadened the scope of thinking by stressing that the role shift also makes it possible to report not only dialogues but referred entity's actions, states or thoughts (Smith et al., 1988; Ahlgren, 1990; Meier, 1990). Winston (1991, 1995) proposes a new term, *constructed actions* (CA), for these constructions in which the narrator reproduces the actions of one of the protagonists of the utterance (or of himself at an earlier point), because, she says, this is not simply a copy but a selective reproduction of the reported action by the narrator.

Reference and Multimodality of Human Language

A notable development in the study of reference both in SL and in SpL occurred at the end of the 1990s, in parallel with the development of models of cognitive accessibility and, more broadly, the rise of cognitive grammars and usage-based grammars. This evolution is fundamentally linked to the introduction of studies on gesture to the linguistic discipline, specifically, the work of Kendon (1988) and McNeill (1992). These authors renewed the field by advocating for a broader conception of human language, whereby any language, SpL or SL, should be seen as a multimodal and multi-semiotic integration.

With regard to SL linguistics, Liddell (1995, 2003) has been the figurehead for this multimodality paradigm. His new descriptive concepts had a significant impact on how these "non-conventional constructions" were going to be taken into account in the analysis of reference in SL. The key point for Liddell is that, whatever the language, SpL or SL, human communication, when considered in the face-to-face interaction, is not confined to the "symbolic" but resorts to other semiotics, such as indexicality and iconicity. It is therefore a question of integrating as such the at least partially "gradient,"[2] i.e., according to Liddell, "non-grammatically specified" character of certain categories of SL signs, which, he argued, have remained problematic in the literature.

The first of these are the pointing units referred to as "pronouns," and the so-called agreement verbs (which he renames "indicating verbs"). Relying on *conceptual integration theory* (Fauconnier and Turner, 1996), Liddell analyzes both types as *conceptual blends* of linguistic components (i.e., integrated into the ASL lexicon) and a *gestural* component (i.e., "non-grammatical"). More generally, the set of "directionality" phenomena must, according to him, be understood as pure pointing *gestures* directed toward those spatially grounded conceptual entities that are always "referents," whether they are physically present in the signing space or discourse constructs. He then proposes the same analysis for classifier constructions, which he renames "depicting verbs." In his view, these constructions are also a mix of two types of components, a lexically specified component (mainly the handshape) and a gestural component (movement and location). Thus, the unifying property of depicting and indicating verbs would be that both are "directional," i.e., indexical, the cause of this specificity being the possibility for the articulators used in SL to be oriented in the signing space while conveying a symbolic content. The only difference between the two categories would be that "(. . .) the directionality of depicting verbs depicts topographical locative information while the directionality of indicating verbs identifies entities" (Liddell, 2003, 268). According to Liddell, depicting verbs constitute a long but finite list of manual constructions. While a full inventory remains to be achieved, it would be possible to describe them as a "large semi-productive derivational system" (Liddell, 2003, 274) based on verbal roots. Finally, following Winston (1991) and Metzger (1995), Liddell extends his application of conceptual blend theory to CA. He characterizes these as a specific type of "blend," noted for the fact that the signer is a part of it, thus creating what he terms a "surrogate blend." Like indicating verbs and depicting verbs, any part of the CA that does not involve grammatically specified signs, i.e., for him, anything that involves "gradience," is considered as gestural.

Ferrara and Hodge (2018), following Liddell and building on Enfield (2009, 2013) concept of *composite utterance*, take up Clark (1996) tripartition and propose that any language production (in SpL or SL) can be analyzed according to three "methods of signaling" (*describing*, *indicating*, and *depicting*), which can be used separately or jointly. For Ferrara and Hodge, this distinction intersects with another one that seems increasingly more widely accepted, especially among authors adopting Liddell's approach (e.g., Johnston and Schembri, 1999, 2007; Johnston, 2008, 2012, 2019; Cormier et al., 2013; Hodge et al., 2019). The distinction incorporates the types of signs in a continuum from lexical to non-lexical: fully lexical (highly conventionalized) signs; partly lexical signs—which include pointing signs and indicating verbs (cf. agreement verbs), both characterized by their indicative dimension, as well as depicting signs (cf. classifiers constructions) which combine indicating and depicting; and finally non-lexical signs, the "enactments" or "constructed actions/dialogues" (cf. role shifts). The latter "do not have properties of conventionalized symbolism, i.e., meanings that are additional or predictable from the value of their form given a particular context" (Hodge et al., 2019, 36). The recent study by Hodge et al. (2019) on reference in Auslan (Australian SL) adopts this theoretical framework. The authors set out from what they see as a consensus in the literature: that in accordance with the predictions of accessibility theory, both signers and speakers would choose the most informative and phonologically heavy expressions (particularly fully lexical noun phrases) to introduce new referents and, conversely, would favor high accessibility markers (pronouns or zero anaphora) for referents with a high degree of conceptual discourse salience. The study aims to determine to what extent other factors influence the choice of referential expressions, as in particular motivated use of space, animacy, and semiotic form.

The authors statistically analyze tokens of referring expressions in a large corpus of narratives in Auslan. Their

[2] We fully agree with one of our reviewers that Liddell's use of the term "gradient" is somewhat confusing and inconsistent with Langacker's understanding of the term. For Langacker indeed, "gradience" is everywhere in language and, far from being opposed to "symbolic," it is inherent to it (on this point, see in particular Occhino and Wilcox, 2017).

results confirm the role of activation status in the choice of referring expression. New referents are phonologically heavier (according to the authors' definition of "phonological weight," i.e., combining more diverse semiotics); above all, they use relatively more conventional forms (lexical signs, fingerspelling, and mouthing) than with reintroduced or maintained referents, which involve fewer semiotics and less conventional forms (depicting signs and surrogates/enactments/CA). In addition to activation status, animacy has a significant effect on the number and on the nature of the strategies chosen by signers for each referring expression. Human referents require the fewest semiotics overall; however, animate referents (humans and animals) tend to be phonologically heavier than inanimates when reintroduced. Finally, according to the authors, these various results call into question the assertion that the newer a referent (i.e., the less cognitively accessible), the more informative its expression would be. In fact, non-conventional semiotic strategies such as depicting signs, visible and invisible surrogates, those tend to be used more frequently for reintroduction and maintained reference, are particularly rich in information.

Our overview of the literature is far from being exhaustive. However, we have intentionally excluded previous work (mentioned in the introduction) on reference in LSF, LIS (Italian SL), ASL, and other SLs. These studies, carried out within the Semiological Approach framework, had produced results somewhat similar to those achieved by Hodge et al. (2019) for Auslan. In order to account for this more precisely, we must first present our theoretical framework, which we do in the next section.

THE SEMIOLOGICAL APPROACH TO SIGN LANGUAGE

Our conception of SL, designated in recent years as the Semiological Approach, was used to describe LSF (e.g., Cuxac, 1985, 1999, 2000; Cuxac and Sallandre, 2007; Sallandre, 2007; Cuxac and Antinoro-Pizzuto, 2010; Garcia and Sallandre, 2014; Garcia et al., 2018), but other SLs as well: see Pizzuto et al., 2008; Sallandre et al., 2016 for crosslinguistic comparison of various institutional SLs, but also Fusellier-Souza, 2004, 2006, 2012; Martinod, 2019; Martinod et al., 2020, for the description of family and micro-community SLs.

This model was progressively developed from the early 1980s on the basis of close, frame-by-frame, analysis of long discourse corpora, recorded *in situ* (Cuxac, 1985, 1993, 1999). The methodological decision to work on corpora[3], setting out from a functional and therefore semantically centered perspective (a top-down approach), was original at the time (and remained so until the 1990s), as research on other SLs had long been focused primarily on elicited data such as decontextualized sentences. Cuxac's preliminary description, setting out from meaning and systematically seeking what conveys this meaning, takes into account from the very beginning all articulators, both manual and non-manual, focusing in particular on the role of eye-gaze (soon established as central). Very early on, he hypothesized that the modality has a strong impact on the structural and typological characteristics of SLs and that the close similarities between them are significant in this respect.

Like Jouison (1986/1995), the other pioneering LSF linguist, Cuxac soon highlighted the high frequency of highly iconic constructions, which could not be analyzed in the terms of lexical signs. Although involving the same types of manual components as lexical units, these constructions did not meet the criteria then used to define "verbal," i.e., they were iconic, their meaning varying continuously as their form changed. Focusing on these constructions, Cuxac succeeded in establishing that they stem from a few linguistic structures (or "patterns"), which he calls "transfer structures" (Cuxac, 1985). These structures indeed make it possible to account for the multitude of highly iconic constructions observed in discourse, therefore termed as "transfer units." The three main transfer structures are the following: the "Size and Shape Transfer" (SST), which allows to show the shape and/or the size of an entity (**Figure 1**); the "Situational Transfer" (ST), which is to show an actant (dominant hand) moving with respect to a stable locative (typically the non-dominant hand), the scene being represented as a global view, from a distance (external point of view) (**Figure 2**); and the "Personal Transfer" (PT), by which the signer literally takes on the role of the entity he refers to, and thus shows, as in a close-up shot, the actions it performs or suffers (internal point of view) (**Figure 3**)[4]. Any transfer unit built from these structures, simultaneously involves all parameters, manual and non-manual. These constructions are *verbal* (that is, linguistic) precisely because they are based on structures, that is, they are composed of constrained elements that fit into paradigms. Another key point which we return to below is that transfer structures share a formal feature, the breaking of eye-gaze toward the addressee.

The characteristic of transfer units is that their global meaning, with a specifying value, comes down to the sum of the (iconic) meaning of their components. However, it must be emphasized that, while it is true that the precise meaning of a transfer unit depends on the context, as frequently noted in the literature, these units have in themselves a highly generic semantic value. Thus, the situational transfer unit in **Figure 2** (right image) shows a "shape with double salience moving along an arched path toward a horizontal flat shape." The transfer structures actually reveal a specific way of saying, their mode of meaning production being, directly, iconicity.[5]

It must be underlined that what is described as "classifier construction"/"depicting sign" in literature only matches the manual component of our transfer units. For Liddell (2003) and

[3]At that period, this could not have been a corpus in the modern sense of the term, i.e., *machine-readable*, as defined by Johnston (2008). Yet, given that discourse data were mainly collected in an ecological manner and transcribed frame by frame following an explicit procedure (see Cuxac, 2000; †Antinoro Pizzuto and Garcia, under review), it is as close as one can get. The important point here is that the description and theory that followed were authentically *corpus-driven* (vs. *corpus-based*, *cf.* Tognini-Bonelli, 2001).

[4]For a detailed description of what these three main transfer structures are precisely made of, the reader may refer to Cuxac and Sallandre (2007) and Garcia and Sallandre (2014).

[5]We return to this term later (see note 11 below).

FIGURE 1 | Transfer of size and shape "thin, vertical, elongated shapes" for the referent "fence" in LSF and Libras.

FIGURE 2 | Situational transfer "jump over the fence" in LSF and Libras.

FIGURE 3 | Personal transfer "the horse galloping" in LSF, LIS, and NS.

the SL linguists who follow his theoretical framework (as Hodge et al., 2019), "depicting verbs/signs" are thus merely manual signs. For us, these manual elements are only one *component* of many in much bigger structures, the transfer structures, which incorporate all parameters, manual and non-manual alike. It should be stressed that a notable difference is the attention we pay to non-manual parameters in the identification of structures and unit types and thus in the choice of linguistic tags. The handshape present within the transfer units is called "proform" in the Semiological Approach. Far from categorizing the referent, proforms constitute a closed list of handshapes that aim to show some aspect of the referent[6] (Cuxac, 2000).

[6]See examples below (**Figures 2**, **3**).

Transfer units are extremely frequent in discourse, representing up to 70% of occurrences in narratives and up to 30% in other genres, prescriptive and argumentative (*cf.* Sallandre, 2003; Sallandre et al., 2019). They intertwine with the other main type of units, the lexical units, but also with fingerspelling and pointing units. The lexical units have the same mode of meaning production as the SpL words, that is, pure convention. The lexical meaning being mainly carried by the manual components, these units, having a conventional global meaning and a generic value, are instantiated in discourse by a pluri-linear organization of non-manual, semantically specialized parameters: the gaze (rector of interaction and activator of deixis, see section "Enunciation and Deixis-Anaphora: Key Role of Eye-Gaze"), facial expression (carrying aspectual and modal values), facial movements (ensuring a phatic function), and body movements (marking phrases, coordination, and thematic organization). The termed "standard" structures, which involve lexical, pointing, and fingerspelling units, are mainly characterized by the use of signing space in order to create references and to express various semantic relations between referents. Being part of the classical mode of "saying," the lexical units and the standard structures in which they are employed were essentially the focus of linguistic research on other SLs in the first three decades (1960–1990). Although non-conventional units have been widely studied since the late 1990s, the core of the "grammatical" system is still considered to be the lexical units and the structures that employ them (e.g., Goldin-Meadow and Brentari, 2017; Fenlon et al., 2018).

From our point of view, the space used in the transfer units is itself iconic. It is an imagistic iconicity, showing the deployment of a shape in size and shape transfer, the moving of the actant in situational transfer, the space of action of the transferred entity in personal transfer. However, the space in which the lexical units are used is also iconic. Yet, it is a different type of iconicity, a diagrammatic one (in the Peircean sense)[7]. Transfer units most often have the same format as lexical units (they coincide overwhelmingly with a "minimum unit of realization") and involve the same types of parametric components[8]. They can also combine with each other, with a lexical unit, or with a manual pointing, depending on regular patterns that result in greater structural complexity. We precise and illustrate these points and the most frequent of these combinations in the next section.

However, a central aspect of the Semiological Approach still needs to be clarified, since it gives the model its explanatory dimension. We indeed hypothesize that transfer structures would be found across all SLs around the world. This hypothesis was first supported by the analysis of *homesigns*, family sign languages developed by deaf children isolated in hearing environments

[7]The loci are therefore not intended to reflect the actual absolute positions of the entities they represent, but rather the (person, space, time) relationships between these entities.

[8]Cuxac (2013, 78) refers to these two main types of SL discourse units as "multi-track body matrix with relevant cells that are more or less filled in (eye-gaze, body posture, facial expression, manual parameters) depending on the structure achieved" [our translation; French: "(. . .) une matrice corporelle multipiste à cases pertinentes plus ou moins remplies (regard, posture, MF [mimique faciale], paramètres manuels) en fonction de la structure réalisée »].

(Goldin-Meadow, 1991, 2003), and the description of SLs developed in ontogenesis by deaf adults in Brazil, without contact with any institutional SL (see Fusellier-Souza, 2004, 2006, 2012, and more recently, Martinod, 2019; Martinod et al., 2020)[9]. In these SLs, created at the initiative of the deaf themselves and developed over their lifetime, these studies have found the same transfer structures. They coexist alongside standard structures, just like in institutional macro-community SLs. These observations form the basis for our proposed scenario for SL semiogenesis. Starting from an initial mode of saying based (as *in spite of* signers themselves) on an iconicization of their perceptivo-practical experience, transfer structures would progressively emerge at a certain stage from the repeated use by the signers of a deliberate intent to do as much iconic as possible[10] in order to make themselves better understood[11]. In this context, a structural bifurcation would gradually occur, opening the way, parallel to the first one, to another mode of saying, in a generic way, with no intention of being iconic (that is, with no illustrative intent). This would result in the emergence and multiplication of lexicalized units, born from the routinization (entrenchment) of transfer units having lost their illustrative scope. However, the centrality of transfer structures is not just a diachronic or historical phenomenon. Rather, it accounts for the current discursive dynamics in SLs. The dynamics is grounded on the functional complementarity of the two available modes of saying, depending on the signer's intent: saying while showing and saying without showing.[12] The iconicity attested in many lexical signs, where it is not, however, the mode of meaning production, is not a pure etymological remnant, doomed to disappear. Instead, this "dormant iconicity" is functionalized: it only allows the economic integration of the two main types of structures and units into SL discourse (see next section, **Figure 4**).

Ultimately, deafness would be the root of SL structuring, in two respects: first, from a genetic perspective, as 95% of deaf children are born from hearing parents,[13] the scenario proposed above for the semiogenesis of SL is very likely as old as deafness itself; secondly, from a communicative perspective, the need to maximize the communicative potential of the sole modality available, the visuo-gestural modality. From then on, the Semiological Approach is grounded on a preliminary "semiology of the channel," the latter being understood as a modality that generates constraints but also carries its own potential. For the visuo-gestural channel, that means, first, the possibility to maintain the figurative, allowing one to reproduce *as closely as possible* the universe of mental imagery, as this is the very channel by which we experience the world, but also, the dual possibility (linked to the nature and visibility of body articulators) to reflect semantic relations in spatial terms, and to fully exploit simultaneity. The Semiological Approach thus opens up an epistemological reversal, inviting us to contemplate the *forms* used in SpL similarly, removed from their habitual privileged position in general linguistics, and instead as particular reactions to *constraints* imposed by the audio-vocal channel[14], but also having the option, in hearing communication, to employ two modalities jointly[15]. This explains the name of the Semiological Approach, which aims, therefore, to *describe all human language*. The Semiological Approach models SL as a *type* of language, because it is rooted, on the one hand, on the incidence of deafness, and on the other, on the hypothesis of a close link between linguistic structure and mental imagery (that itself stems from experiential interactions with the world)[16]. The following section is intended to support the typological contribution of our approach.

TYPOLOGICAL SCOPE OF THE SEMIOLOGICAL APPROACH: REFERENCE TO ENTITIES IN DIFFERENT SIGN LANGUAGES

The examples we present in this section are drawn from two sets of data: data in three SLs collected and analyzed by Pizzuto et al. (2008) and those from a large corpus of narratives in eleven SLs presented in Sallandre (2014, 2020), Sallandre et al. (2016).

Pizzuto et al. (2008) is the first crosslinguistic SL study on doing reference in SL carried out within our theoretical framework. The corpus is made of narratives from three signers, in LIS, ASL, and LSF, elicited from two stimuli: for LIS and ASL, the story-board *Frog, where are you?* (Mayer, 1969), and, LSF, the story-board *The Horse*. This study has shown strong similarities between the three SLs, LIS, LSF and ASL. Their results show that lexical units are favored for introducing animate

[9]Similarly, see the description of a micro-community SL in Senegal, Jirou-Sylla (2008).

[10]This hypothesis of an "intent" to be "as iconic as possible" is a key one for the Semiological Approach, both at the diachronic and ontogenetic levels and in the synchronic use of language. As an explanatory hypothesis, it is of the same nature as, among other examples, the one that underlies Ariel's Accessibility Theory, for which the speaker/signer would choose his/her referential expressions according to the degree of accessibility of this referent that he assumes in his interlocutor.

[11]If so, the transfer structures (and their components) are what gives us the closest insight into what "iconic" means. They are indeed considered as the product of one human being's efforts to make himself understood by another human being by making maximum use of the resources of a channel that also happens to be the very channel through which they both experience and conceptualize the world.

[12]For examples of such functional complementarity, see in particular Garcia and Sallandre (2014).

[13]The atypical language acquisition of deaf children from hearing families is thus, from the very beginning, a corner stone of our approach (see notably Cuxac et al., 1999).

[14]We sincerely thank the reviewer who drew our attention to the astonishing similarity with the words of the great Hockett, of which we give only snippets: "The difference of dimensionality means that signages can be iconic to an extent to which [spoken] languages cannot. [. . .], so that it is perhaps more revealing *to put the difference the other way around, as a limitation of spoken languages*. Indeed, *the dimensionality of signing is that of life itself* [. . .]" (Hockett, 1978, 274–275, quoted by Wilcox, 1996, 184), *emphasis added*. In the same vein, we fully subscribe to the model recently developed by Occhino (2016, 2017); she writes for instance: "By beginning with signed languages, and expanding our analysis to spoken languages (*the opposite of the typical direction of linguistic theory building*), we gain keen insight into the nature of schematization and emergence of structure, *which is obfuscated by the opaqueness of the articulatory mechanism of spoken languages*" (Occhino, 2017, 94, *emphasis added*).

[15]Herein lies the explanation to the differences between coverbal gesture in hearing contexts and SL transfer structures, beyond any similarity. On this point, see Cuxac (2008).

[16]That is why we consider iconicity not only as a "pervasive aspect" in SL but as their "organizing principle."

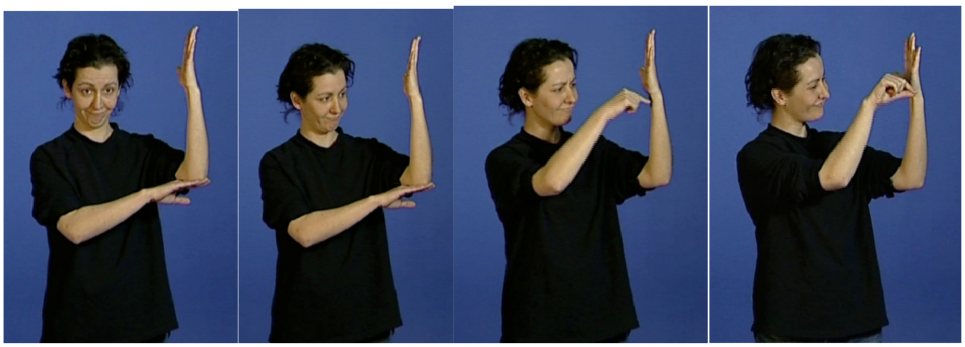


FIGURE 4 | Referential frame switching through the gaze; the tree sequence (corpus LS-COLIN, Cuxac et al., 2002).

or inanimate referents (50%–83%), while transfer units are the preferred method for maintaining and reintroducing referents (76%–95%). More specifically, personal transfers are used mostly for animate referents while transfers of size and shape and situational transfers are preferred for inanimate referents; double transfers are used to reintroduce referents of both types but never used to introduce new referents. Finally, a small proportion of anaphoric reference is marked through manual pointing signs (3%–7%). These initial results should be compared with the more recent results in Hodge et al. (2019), within a different theoretical framework.

Following this work, Sallandre (2014, 2020), Sallandre et al. (2016) compared reference to animates in eleven SLs, focusing in particular on how personal transfer units interact with lexical and pointing units to introduce and maintain reference. The SLs studied, as illustrated below, are LSF (French SL) LIS (Italian SL), LSR (Romanian SL), DGS (German SL), VGT (Flemish SL), PJM (Polish SL), SASL (South African SL), NS (Nihon Shuwa, Japanese SL), Libras (Brazilian SL), LCSh (Chilean SL), and LSM (Mauritian SL)[17]. These are both European and non-European SLs with diverse institutional statuses. The data were collected by us or by our colleagues in the various countries. The same narrative, *the Horse*, was produced by five deaf signers in each language; the productions were then annotated using the ELAN software (Sloetjes and Wittenburg, 2018). The same template was used to annotate all productions and is relatively synthetic, using the following fields (called *tiers* or *actors*): The tier *unit of meaning* proposes a translation of the minimal units of realization in the written SpL of the relevant country (e.g., Italian, for the LIS, cf. **Figure 5**) and in French, the working language of the authors. The tier *category* assigns a label to each unit (lexical unit, types of transfer, pointing, or fingerspelling). The other tiers correspond to the non-manual parameters (gaze, facial expression, mouth patterns, and body posture).[18]

The most common transfer structures are described below. Our stake is twofold. We highlight the presence of these transfer structures across all the SLs examined, whether historically related or not[19]. At the same time, we illustrate what characterizes them as structures, namely, their compositionality. We first present the three main transfer structures and then illustrate some of the most common complex regular patterns that result from a combination of these structures either with each other or with other types of units. Taking into account for each simultaneous construction all its manual and non-manual parameters, analyzed according to the signer's intent[20], highlights the extreme potential linguistic body partitioning.

Figure 1 demonstrates a transfer of size and shape in LSF and in Libras. This transfer unit appears in the first part of the story, in the description of the background that includes a fence. Both signers use the same "four fingers spread" handshape for the dominant hand to represent the shape of the posts of the fence. In both signers, the fingers are pointing upward, but the palms of the hands are facing in different directions (outward for the LSF, inward for the Libras). This is a personal variant, not due to the norm of either language, and it does not affect the meaning conveyed. In both cases, the eye-gaze instantiates the shape described by the dominant hand (right hand) while the facial expression shows the length and delicacy of the shape described, also suggested by the squinting of the eyes.

The two images in **Figure 2** depict the same crucial moment of the story, the horse's jump and fall. Both signers, as most other signers of this corpus, chose to express this event using the same structure, a situational transfer. The choice is probably motivated by the inherent external point of view of this transfer, which allows the signer to emphasize the harsh trajectory of the horse relative to the fence. In these two images, the meaning conveyed is very similar: in LSF (on the left), the dominant (right) hand shows the horse jumping over the fence, which is represented by the

[17] We use the acronyms validated by the deaf communities involved.

[18] While tiers for manual and non-manual parameters are almost always present in SL literature, the tier for *category* (tag) seems to be original: it can be filled as the result of parametric field values (manual and non-manual). For example, if a unit of meaning incorporates some aspectual facial expression, a gaze directed at the dominant hand, which represents the moving agent while the non-dominant hand represents a stable entity, then we have a situational transfer, which will therefore be annotated as such in the category tier.

[19] We have chosen SLs that are relatively close, historically and geographically, like LSF and LIS, as well as unrelated SLs, such as LSF and NS.

[20] Non-manual parameters fulfill various functions depending on the intent used. Given space limitations, we refer the reader to Cuxac and Sallandre (2007) for further detail.

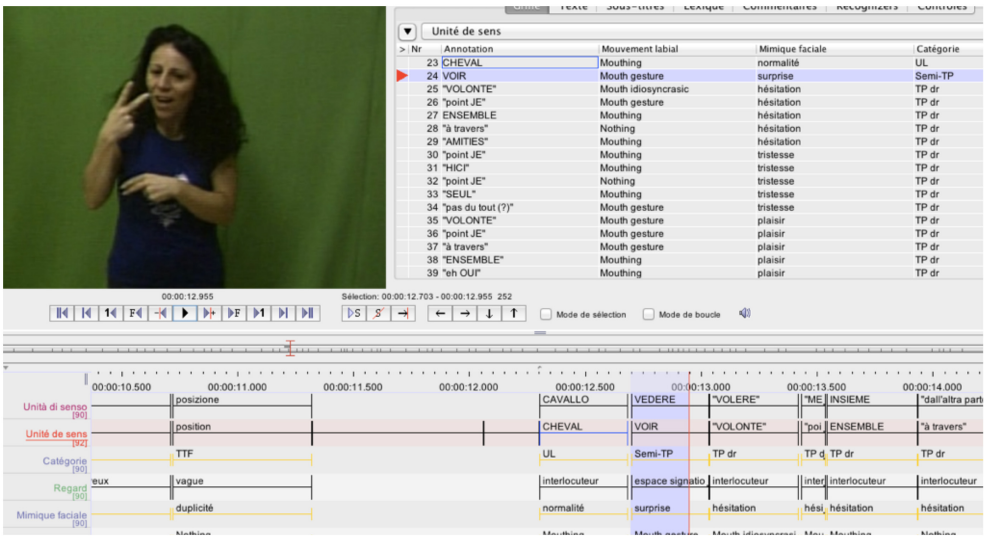


FIGURE 5 | Screen capture of ELAN annotation of a story in LIS. Focus on the meaning unit VEDERE (SEE) in a semi-personal transfer structure.

non-dominant hand; in Libras (on the right), the dominant hand also shows the horse's jump, while the non-dominant hand figures the ground onto which the horse stumbles awkwardly. The structural similarity between the two units is obvious: meaning conveyed by the dominant hand, which represents the action of the animate referent, movement of this hand over the non-dominant hand (locative), and gaze following the action carried out by the dominant hand. The minor differences are in the proform used for the dominant hand, in LSF, a V shaped form with two saliences (two fingers stretched open), and in Libras an X shape (two folded fingers), producing two slightly different representations of the form depending on whether or not the signer intends to show, at that particular moment, the bending of the horse's legs. The facial expressions providing aspectual value are also slightly different, expressing effort and speed in LSF and the shock of the fall in Libras.

Finally, the images in **Figure 3** present a very similar personal transfer of the horse galloping, in three languages, LSF, LIS, and NS. Again, we find striking structural similarities: directing the gaze away from the addressee, the postural involvement (chest, shoulders, head), and facial expression (that of the entity), all these elements indicating an internal and embodied point of view on the scene, in contrast to the external point of view inherent to situational transfer. Depending on SLs and signers, the handshape may differ (proform "fist" as here, or proform "two outstretched fingers" or even "flat hand"), according to the aspect they intent to show, but the attitude of the signers, moving away from the "Enunciation Domain" to embody the protagonist of the utterance, is extremely similar (for a definition of the term "Enunciation Domain," see the next section).

After having illustrated the three main transfer structures, let us move on to those that combine, simultaneously, either another transfer or a lexical or other unit. These structures, which will be outlined below, exhibit higher semantic density and more referents simultaneously present in the utterance. However, the role of the gaze is constant: in all these personal transfers, it is to represent the state of mind of the character(s) embodied as protagonists of the utterance (see section "Enunciation and Deixis-Anaphora: Key Role of Eye-Gaze").

We begin by what we call "double transfer" which is the simultaneous combination of a personal transfer and a situational transfer. Double transfer allows one to simultaneously express multiple perspectives (e.g., that of an agent and of a patient or that of a locative and that of an agent). **Figure 6** demonstrates an example of a double transfer in three SLs, PJM, LSM, and LCSh. This transfer is produced at the end of the Horse story, and its activation status is a simultaneous reintroduction of the two main protagonists of the narrative. It can be translated by the utterance "the cow bandages the horse's leg." This construction structurally combines a situational transfer locative (the horse's leg, represented by the non-dominant hand) and a personal transfer (the cow bandaging the leg, represented by the whole body, except for the non-dominant hand). Thus, two animated referents are simultaneously present in the utterance. In all images, the parameters are the same, with the gaze oriented toward the horse's leg, which was previously introduced into the narrative, i.e., bottom right for the signers in PJM, and bottom left for the LSM and LCSh.

The frequent integration of elements classically associated with the non-illustrative intent (e.g., a lexical unit) into a structure depending on the illustrative intent (e.g., a personal transfer) is analyzed in the Semiological Approach as a specific type of transfer structure (e.g., a semi-personal transfer, **Figures 5**, **7**). This type of highly frequent pattern (such as these structures) employs simultaneous constructions (i.e., units) carrying multiple references. It is consequently difficult to compare with SpL, including when multimodality is considered.

The examples in **Figure 7** illustrate a semi-personal transfer in LSR, VGT, and SASL, produced at different moments in the narrative. As it is the case in our previous example in

FIGURE 6 | Double transfer "the cow bandages the horse's leg" in PJM, LSM, and LCSh.

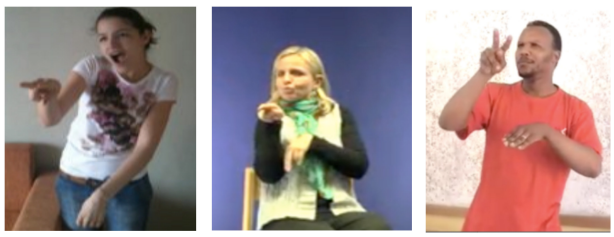


FIGURE 7 | Semi-personal transfer SEE, in LSR, VGT, and SASL.

LIS (**Figure 5**), this structure is defined by the simultaneous association of a personal transfer, here the character of the horse, and a lexical unit that specifies the action of the transferred entity. Here, while the body of the signers (including their facial expression and gaze) depicts the horse being surprised, the action of the dominant hand indicates that the horse "looks" at the cow. This is the lexical unit SEE that is used. A similar example from Auslan is presented in Hodge et al. (2019, Figure 4, unit 5). Thus, the personal semi-transfer is found in all the SLs examined (as does the semi-double transfer described below).

These corpora were assembled to examine one of the typological hypotheses underlying the Semiological Approach, namely, the existence of various transfer structures across SLs. Beyond certain variations between signers and between SL, our analyses confirm the existence of multiple structural similarities across the SLs examined. They also confirm the richness of the transfer constructions in each SL, as evidenced by the existence of double transfers or personal transfers with reported discourse[21]. Such outcomes, we must insist, require taking into account at any moment the functions performed by all the parameters, manual and non-manual, and particularly the gaze.

In the last part of this article, we will focus on the importance of gaze behavior (the behavior of the signer's gaze but also, crucially, that of the addressee's) for the functioning of reference in SL. We believe, however, that only an "enunciative" approach can provide an appropriate account thereof. The Semiological Approach, which has taken a functionalist perspective from the outset, also found significant resonance in the principles of Langacker's (1987, 1991) Cognitive Grammar, and in Lakoff's thinking (e.g., Lakoff and Johnson, 1980/1995). However, as stated in the introduction, an important aspect that distinguishes our conception from others which also align with a cognitive–functionalist approach is the adoption, from the outset, of the perspective of the *théories de l'énonciation*, a European approach that developed notably in France over the 20th century, its proponents dominating the field in the 1960s–1980s. This approach and its close links with deixis–anaphora are the focus of the following section. We highlight its specific relevance to SL, provided that the role of the gaze in these languages is fully taken into account.

ENUNCIATION AND DEIXIS–ANAPHORA: KEY ROLE OF EYE-GAZE

The term "enunciation theories" refers to a set of very diverse approaches which have in common that they have in-depth questioned the abstract notion of "language" ("la langue") posed by Saussure (as opposed to "la parole"). However, as Liddle (see Culioli, 1995) and Fuchs (2008), among some others, point out, while there are important points of intersection with Cognitive Grammar, these contributions have remained largely unknown in the United States. It is not possible to reconstruct here the historical roots of the notion of "enunciation," nor the specific contributions of its main representatives. With regard to the Semiological Approach, the key references are to be found in Jakobson (1957) and his concept of "shifter", in Benveniste (1970) and, concerning more particularly the concept of enunciator/co-enunciator, in Culioli (see note 24).

It is indeed from the concept of "shifter" masterfully developed by Jakobson (1957) for the study of the verbal forms of Russian that Benveniste elaborates his "enunciation theory." He shows that every utterance ("énoncé") necessarily contains a set of terms ("indices") whose specificity lies in the fact that they can only be defined by reference to what made it possible to produce the utterance itself, which he calls its "enunciation" ("énonciation"). These are the 1st- and 2nd-person pronouns that refer to the two interlocutors in the act of enunciation, "deictics" such as "here," "now," which refer to its place and time, the verbal tenses (the present tense, which designates a period of time as that of the enunciation), the way speakers embed their own personal assessment of their messages within them (the modalities), etc. Therefore, "Enunciation is this coming into service of language that is created by an individual instance of use."[22] (Benveniste, 1974, 80), the conditions of this activation being inscribed in the very system of language, through what is described as "the formal apparatus of enunciation" (Benveniste, 1970). Consequently, far from being a neutral and objective system, language ("la langue") contains "indices" that are to be considered as the very basis for constructing referential values.

By this very fact, every utterance carrying within it traces of its enunciation, it should be analyzed by taking into account two "layers," referred to as the "Enunciation Domain" ("plan de l'énonciation") and the "Utterance Domain" ("plan de l'énoncé").

[21] This last type of transfer is not presented here: we refer the reader to the detailed example below, **Figure 9**.

[22] French: "L'énonciation est cette mise en fonctionnement de la langue par un acte individuel d'utilisation." [our translation]

The first links and *linguistically* co-determines the speaker and the addressee: the very act of enunciation establishes them simultaneously (and reversibly) as 1st and 2nd person. The Utterance Domain is internal to the discourse being produced: it links the protagonists of the uttered process[23]. A key point is that the Enunciation Domain cannot be reduced to what is commonly referred to as "the utterance context," i.e., the context understood as the physical environment and all the actual circumstances in which an utterance is produced. The absolute, actual physical coordinates of the interlocutors are not relevant from an enunciative perspective. Personal shifters express the necessarily mutual co-determination of the two "co-enunciators"[24], and by contrast, that of the non-person (i.e., 3rd person). Pizzuto (2007), summarizing Benveniste's thought, underlines this "ineradicable subjectivity" introduced into language through the relationship between interlocutors (co-enunciators) established by the act of uttering (enunciation) and its necessarily universal nature.

Let us now recall that the issue of grammatical person marking was debated in SL linguistics very early on. As mentioned above, the long-dominant analysis identified in ASL (and later in other SLs) three personal pronouns, in 1st, 2nd, and 3rd person, formally characterized, for the 1st person by an index finger pointing toward the signer's chest, for the 2nd and 3rd person by an index finger pointing respectively toward the addressee (2nd) or to the addressee's right (3rd). The first to challenge this analysis was Meier (1990). Arguing that the interlocutor can alternately be one or the other of the participants (other than the signer) in an exchange, Meier points out that for both the 2nd and the 3rd person, the direction of pointing can be infinitely variable. This variation, he argues, poses a problem for a formal specification of these pronouns. He therefore proposes that ASL presents only a binary grammatical opposition, between 1st person and non-1st person. The debate also focused on the possibility of a formal analysis for the marking of person/arguments of the verb in directional verbs/agreement verbs. Recently revived, the discussion therefore focuses on the non-listable (non-morphemic) character of spatial points (loci) that can be created for doing reference. As mentioned above, Liddell (2003) provides the same analysis for both pointing signs and what he renames "indicating verbs": the theoretically unlimited variation of their direction indicates that they are gestural (i.e., according to him, non-symbolic) in this respect. Liddell thus joins Meier's position via another way: since manual pointing is assumed to be what formally marks the grammatical person, the unlimited variation of the actual location of the interlocutors would block the possibility of a formal distinction between the 2nd and 3rd person. However, Liddell goes further: he assimilates SL pointing signs to ostensive pointing gestures that can be found in SpL coverbal gesturing. In his words, "the directionality [of pointing signs] is an explicit instruction telling the addressee how to map the pronoun's semantic pole. *The addressee needs only to follow the directionality of the pronoun*, which will lead to the appropriate entity." (Liddell, 2003, 91, emphasis added).

Our analysis is very different. Adopting an enunciative perspective from the outset, Cuxac was particularly attentive to the gaze behavior of both interlocutors. He thus noticed that what specifies the addressee's gaze in SL is on the contrary its fixity (Cuxac, 2000, 217)[25]:

> "Anyone who has had the chance to observe signed communication cannot be but struck by the immobility that characterizes the receiver of the message: his/her body and face remain still (except for micro oscillations of the head that play a phatic function). What is most striking, however, is *the stillness of [the addressee's] gaze*. In order to capture the linguistic information provided by the signer's gaze and facial mimicry, the addressee maintains his/her gaze constantly focused (with respect to central vision) on the area around [the signer's] eyes. Most notably, *the addressee's gaze is never directed (in foveal vision) on the gestures that are produced, and it does never follow the movements of the signer's hands*"

This observation and its consequences are of crucial importance, as the fixity of the addressee's gaze attests (contra Liddell) to the deep difference in nature between the pointing sign in SL and the ostensive pointing in coverbal gesture. In parallel, Cuxac points out this other seemingly trivial fact that, in these visual face-to-face languages, no communication can take place without shared gaze. Now, what defines 1st and 2nd person as such in SL is this interlocked gaze which is also the very condition for the establishment of an act of enunciation in these languages. Indeed, according to the Semiological Approach, the *primary* means through which signers encode person reference distinctions is not pointing signs but, precisely, eye-gaze. These gaze patterns can be combined, for the 1st person, with a self-pointing and for the 2nd with a pointing toward the one being looked at (co-enunciator). However, as noted in the literature, these pointing signs, which have rather an emphasis value, are often optional. The 3rd person is, as opposed to the 1st and 2nd person, what is pointed at by the signer without being looked at (very literally the "non-person"). In an enunciative perspective, what is thus relevant is not pointing signs *per se* (nor a fortiori their actual direction) but their coupling/decoupling with gaze. This coupling/decoupling constitutes the basis of the distinction between the 1st and 2nd person and between them and the non-person, this operating in the two "Enunciation" and "Utterance" domains. The other salient feature is indeed, once the co-enunciators' gazes are "interlocked," thereby determining the Enunciation Domain, the extreme mobility of the signer's gaze, as opposed to the fixity of his/her addressee's: "(. . .) the signer's

[23]The distinction between these two domains/layers is of particular importance. Of course, we cannot pretend to account here for the specificities of the Enunciation Theories with regard, in particular, to Langacker's Cognitive Grammar, and we refer the reader to the rare studies on this point (see, in particular, Liddle's analysis of the proximities and differences between Culioli and Langacker, in Culioli, 1995). However, we will say that if, from an enunciative perspective language is similarly considered as a tool for conceptualizing the world, it is first and foremost a tool for interaction.

[24]Here lies what has been analyzed by several authors (e.g., De Voguë, 1992) and by Culioli himself as the major source of divergence between Benveniste and Culioli; namely the clearer and crucial difference established by Culioli between the "speaker"/the "addressee" (who refer to human beings in actual contexts of utterance but not to abstract coordinates in the "Enunciation Domain") and what he theorized by contrast as "enunciator" and "co-enunciator." The Semiological Approach is on this aspect closer to Culioli.

[25]Translation from Pizzuto (2007, 15, note 4); emphasis added.

gaze is extremely mobile, and meaningfully redirected toward the points in space that mark deictic-anaphoric reference in the 'third-person domain' in discourse" (Pizzuto, 2007, 19).

In fact, both prerequisite for the advent of any signed interaction and an anchor point for the personal deixis established by the very act of enunciation, the signer's gaze is also the key operator for creating and tracking (personal, temporal, and spatial) references in SL discourse. Thus, where the signer's intent is non-illustrative, it is his/her gaze (sometimes coupled with a manual pointing, either preceding, accompanying, or following) that activates a specific point in space (locus), prior to a lexical unit being spatialized there. The signer's gaze alone is subsequently sufficient to reactivate the locus, thereby reactivating the associated referent. In other words, it is primarily the gaze that *diagrammatizes* space, enabling a weave of semantic (grammatical) relations between entities associated with these loci. According to us, the use of space in SL is therefore of two types: (i) a topographical or descriptive space, which is an imagistic space and characterizes reference under the illustrative intent, and (ii) a diagrammatic space, typically involved in the construction and tracking of reference outside the illustrative intent. This, however, must be complemented by taking into account, from an enunciative perspective, the opposition mentioned above between the Enunciation Domain and the Utterance Domain and the different discursive frames of reference they generate. The following example[26] (**Figures 8**, **9**), which combines the two modes of meaning production, will illustrate these points, beginning with the degree of complexity the imbrication of the two types of spaces (imagistic and diagrammatic spaces) can achieve as well as the corollary finesse of the signer's management of his gaze. It should be pointed out that the sequences of images do not claim to represent the entire discourse, but rather selected moments.

The signer in this sequence is the director of an association teaching LSF to hearing people. In the example, he is explaining to the addressee the origin of his sign name. To do that, the signer refers back to the time he met his wife. He looks to some point on his left and thus activates a locus on which he immediately places the lexical unit MEET (**Figure 8A**) and then the lexical unit HUSBAND-WIFE (**Figure 8B**), which means "I meet [my] wife." Henceforth, this locus stands for the "wife" entity, following the diagrammatic logic mentioned above. Then follows a short sequence where the signer explains that at that point, his wife worked in a school for deaf children and that, in order to save money, he picked her up at work (**Figure 8C**). Looking at the pre-activated locus to his left, the signer produces the directional unit PICK-UP and orients it toward the "wife" locus with a motion of his chest ("I (therefore) was the one who picked her up"). In the immediately following sequence (not shown here), the locus "wife" is repeatedly reactivated and used as such. The signer explains that he was on this occasion regularly observed by the schoolchildren, and in particular by one of them. Therein lies the interesting point: switching to the other mode of meaning production (saying while showing) by breaking up the shared gaze, the signer continues his story by embodying himself in this child. To do this, he uses a personal transfer, enabling him to incarnate an entity distinct from himself (a 3rd p.) whose actions, thoughts, etc., he can show. The shift of his gaze away from his addressee's, which is typical of transfers, signals that the signer is no longer the enunciator; from that point on, his gaze is the gaze of the child he is transferred in. Following the imagistic logic which specifies the mode of meaning production within transfers, the signer (who became the "child" entity) articulates the directional lexical unit CALL through an orientation (reflected by the gaze) toward a point *higher* on his left (**Figure 8D**), meaning "the child calls my wife." What is noteworthy is that, respecting the logic of the previously elaborated and still active diagrammatic space, the signer positions his locus to his left, but he does so while simultaneously conforming to the logic of the imagistic space opened by the personal transfer: having become the "child" entity, he locates the "woman" entity higher up according to this latter logic. The two types of iconicity and the two corresponding types of space are thus combined in a way that is as economical as it is rigorous: the diagrammatic space of the relations between the actors of the utterance and the imagistic space opened by the personal transfer (space of the transferred entity).

Let us now illustrate, with the following part of the same sequence (**Figures 9A,B**), what we mean by "enunciative space" or "enunciative frame of reference" (enunciation space and utterance space) and the complexity of the constraints the signer must respect in managing his gaze within these intertwined spaces.

Having called the signer's wife (see above **Figure 8D**), the entity "child" engages in a dialogue with her (i.e., the entity represented by the locus "wife"). Becoming therefore a level 2 enunciator, the child–signer entity determines by this very fact the wife entity as co-enunciator (2^{nd} person) by looking at it (signer's gaze on the "wife" locus). In that level 2 Enunciation Domain thus opened within the first level utterance (reported speech), the "child"–enunciator produces a pointing sign (also in height, **Figure 9A**) outside the axis of interaction *while maintaining his gaze (raised) on his co-enunciator*, that is on the locus–wife (marking of the 3^{rd} person: "he") and then he produces the lexical unit BITE ONE'S NAILS (**Figure 9B**). The reported utterance can be translated as follows: "He (= level 1 enunciator-signer, now a 3^{rd} person) bites his nails."[27] What is noteworthy is the reiteration within space of the level 2 enunciation frame of reference of the principle described above for the formal marking of the grammatical person by the dynamics of the shared gaze ($1^{st}/2^{nd}$ person) and its decoupling from the direction of manual pointing (3^{rd} person).

The extreme logic and precision with which the signer manage the deictic functions of his gaze thus makes the whole discussion about the "real" coordinates of the interlocutors and the alleged infinite variability of the loci somewhat pointless. On the contrary, it seems to us that an enunciative analysis such as the one we are proposing, based on the key role of the gaze in these

[26]The examples in **Figures 8–10** are taken from another corpus, the Creagest corpus, consisting of dialogues between deaf adults (Garcia et al., 2015).

[27]The outcome of the sequence is to explain that it is this sign of "The one who bites his nails" that has become the signer's sign name.

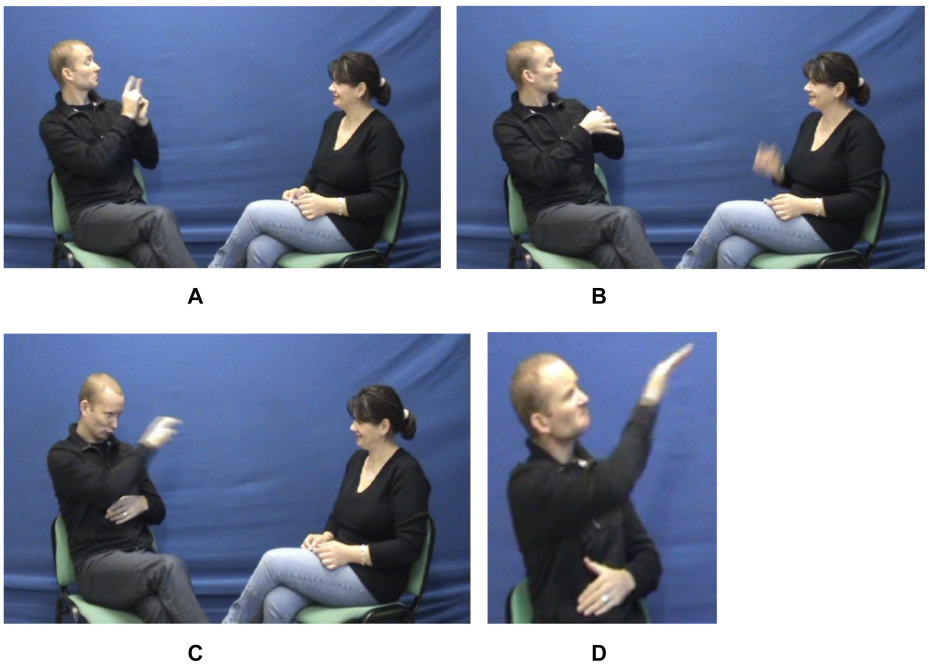


FIGURE 8 | Interweaving between both types of iconicity and space. **(A)** Lexical unit MEET, **(B)** lexical unit WIFE, **(C)** lexical unit PICK UP, **(D)** lexical unit CALL.

FIGURE 9 | Embedding a level 2 enunciation frame of reference within the Utterance Domain. **(A)** Pointing sign outside the axis of interaction (3rd person), **(B)** lexical unit BITE ONE'S NAILS.

languages of the visual and face-to-face that are, in essence, SLs, is able to give the simplest account of their own linguistic economy.

Let us come now to a key opposition in the gaze behavior, which is to signal the signer's semiological intent. As we mentioned earlier, while, outside the illustrative intent, the signer's gaze creates deixis (activating an/or reactivating loci in the signing space), it is yet primarily used to maintain eye contact with the addressee, particularly during the production of lexical units. At the opposite, the intent to say while showing (illustrative intent) requires the signer's gaze to be detached from the addressee, thereby signaling the temporary removal of the signer as enunciator. What distinguishes indeed the illustrative intent, and is therefore shared by the three main transfer types, is the prolonged break of eye contact between the signer and the addressee. By breaking the shared gaze, the signer literally erases him/herself from the Enunciation Domain. In personal transfer, the signer actually disappears as enunciator and embodies an entity referred to in the Utterance Domain, his/her gaze becoming that of the transferred entity (**Figure 3**). In situational transfer, the signer's gaze follows the movement of the entity being referred to by the dominant hand (**Figure 2**). In size and shape transfer, the signer's gaze accompanies the display of the shape (**Figure 1**). Therefore, the signer's gaze is a crucial clue to his/her semiological intent.

However, the analysis of gaze direction and the associated function requires a rather broad discursive context. An example will illustrate this point, while allowing us to refine our presentation of the roles of the gaze. Thus, while in the midst of producing a transfer structure, the signer can briefly direct the

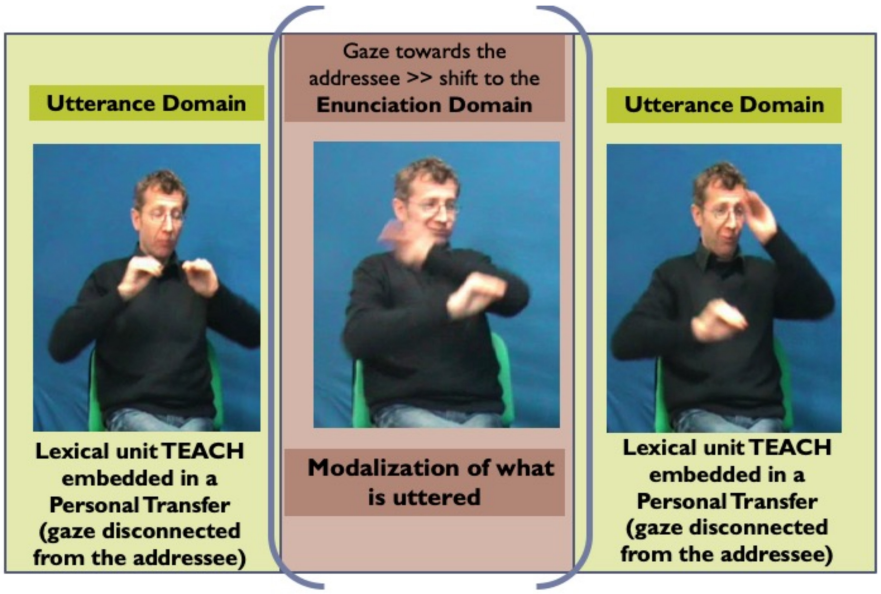


FIGURE 10 | Screenshot of the semi-personal transfer "TEACH awkwardly," from Garcia and Sallandre (2014, 330).

gaze toward the addressee intentionally, as if pausing the manual production, thereby momentarily reestablishing the Enunciation Domain; in this way, now reappearing as the enunciator, the signer can comment on the utterance, through facial expressions, thus "modalizing" it. Let us see the following example, pictured in **Figure 10**.

In this sequence, the signer describes his career as an LSF teacher to hearing adults. He resorts to a personal transfer of himself at the beginning of his career, showing himself as the clumsy professor he was. While embodying the young teacher he used to be, he produces the lexical unit TEACH. This embedding of a lexical unit in a broader illustrative context in which the signer is using a personal transfer[28] stems from a semi-personal transfer (see above **Figures 5**, **7**). As we have seen, this is a very economical structure precisely because the conventional and generic information carried by the lexical unit and the information conveyed by the iconic mode of meaning production (*saying while showing*) overlap, as witnessed by the manual and non-manual multi-linearity characteristic of the illustrative intent. However, complexity is further increased by the play of gaze (and facial expression), which allows the signer to shift from the Utterance Domain (where he stands as an embodied entity) to the Enunciation Domain (where he interacts with the addressee/co-enunciator). In fact, during the personal transfer of himself as a young professor (lexical unit TEACH), teaching awkwardly (hand movement and orientation), the signer's gaze and facial expression are alternatively: (i) those of the transferred character (himself at the time)—his gaze set on the moving hands, i.e., disconnected from the addressee, thereby signaling the transfer, and his facial expression depicting the muddled and awkward nature of the process (of teaching) (**Figure 10**, left and right images) and (ii) those of the signer/enunciator commenting to the addressee/co-enunciator on his teaching experience, displaying self-deprecation—his gaze set on the addressee, with a self-deprecating facial expression. Such sequences, whose complexity arises from the intertwining of lexical and transfer units, alternating between the two modes of saying, and from the interplay between the Enunciation and Utterance Domains are very characteristic of SL discourse.

Finally, acting as a rector for changing the frame of reference, the gaze is also what determines the shift from one intent to the other. In fact, it is often enough for the signer to direct his/her gaze on a lexical unit (by definition not looked at) so that, by switching to the illustrative intent, the latter shed its conventionalized nature and deploys its iconic potential, either by reactivating an original iconicity (in a lexical unit that stems from a transfer unit) or by re-motivation (reanalysis). To illustrate, in **Figure 4**, the gaze is initially directed toward the addressee during the production of the lexical unit TREE (unit 1); then, after a slight nod, the gaze turns toward the sign itself (unit 2); the lexical unit becomes a proform that iconically depicts the tree, and its branches, in particular (unit 2). This opens up the possibility of creating a construction around this locative proform, which is first activated by a manual pointing (unit 3) and becomes the situational transfer's locative on which sits the bird, i.e., animated agent of the utterance (unit 4). Thus, as indicated above, the iconicity present in many lexical signs is what allows the back and forth between the two modes of saying and thus between the main types of structures, in a particularly economical way.

In the end, as a condition for the advent of a signed interaction, anchoring of the personal deixis, rector of the

[28] As we have seen, the micro-sequence described here is part of a larger sequence with a clear illustrative intent both before and after the chosen example.

referential framework, marker of the signer's intent, key operator of the diagrammatization of space (activator and re-activator of reference), vector of the modalization of the utterance, the gaze is plurifunctional in SL. This is why, we insist once again, in order to properly analyze the function that the gaze fulfils at a given point, a sufficiently broad part of the discourse must be taken into account.

CONCLUSION

If we want to be able to compare crosslinguistic data on an equal basis to determine how reference operates in SL discourse, it seems to us urgent that SL linguists come to an agreement on how to segment sequences. From our perspective, namely, that of the Semiological Approach, segmentation requires an equal consideration of *all* meaning-carrying parameters, manual and non-manual alike, from a "vertical" view of the minimal unit of realization (see above, note 7, on Cuxac's, 2013 "multi-track body matrix"). While every parameter, non-manual ones in particular, plays a role in this matrix, this role is affected by the signer's intent, saying without showing, on the one hand, and saying while showing, on the other hand. Intent is defined by the gaze, whose role in this capacity is crucial.

We have stressed the multiple functions of the gaze and the importance of taking into account two key observations, from an enunciative perspective. The first observation is that, in these visual languages, where communication is by nature face-to-face, it is the shared eye-gaze that anchors deixis. The second one is the fixedness of the addressee's gaze, which maintains focused on the signer's. This is sufficient to highlight the profound difference in nature between (linguistic) pointing signs in SL and ostensive pointing gestures that can be found in SpL coverbal gesturing (*contra* Liddell, 2003). On this basis, we have endeavored to show how the distinction between Enunciation Domain and Utterance Domain, on which "enunciation theories" are based, is able to account for the most complex discourses in SL in a particularly economical way, provided that the signer's gaze is accurately taken into account.

Conducted in this perspective, our analyses of discourse sequences from corpora in several SLs attest to the multiple structural similarities notably with regard to transfer structures. This result strongly supports the typological hypothesis underlying the Semiological Approach, that all SLs share a substantial structural base, consisting notably of this type of structure.

Clarifying more precisely the difference in nature between the information conveyed by each of the two main types of "semiotics," saying without showing and saying by showing, should enable us to refine our understanding of "what reference is" and how it is established in SL. The fact that SL linguistically uses these two major semiological modes of saying offers linguists who are open to the multimodality of human language the opportunity to take an innovative look at the age-old theme of deixis–anaphora and thus renews the debate.

DATA AVAILABILITY STATEMENT

Most of the data sets presented in this study are available online. The name(s) of the deposit(s) and accession number(s) can be found below:

- Corpus LS-Colin: https://cocoon.huma-num.fr/exist/crdo/ark:/87895/1.17-483699;
- Corpus Creagest: https://www.ortolang.fr/market/corpora/ortolang-000926 ;
- VGT Corpus: https://www.corpusvgt.ugent.be/ and more specifically https://www.corpusvgt.ugent.be/nl/videoresultaat/1010 (**Figure 7**).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

BG designed the manuscript, wrote sections "Introduction; Reference, Deixis and Anaphora: Background; The Semiological Approach to Sign Language; and Enunciation and Deixis-Anaphora: Key Role of Eye-Gaze." M-AS wrote sections "Typological Scope of the Semiological Approach: Reference to Entities in Different Sign Languages; Conclusion," selected the data in twelve SLs and transcribed the LSF videos, designed and edited the figures. Both authors contributed to the conception of the study and thoroughly analyzed and discussed the data.

FUNDING

The publication of the manuscript was made possible thanks to the support of the Centre National de la Recherche Scientifique (CNRS), through its funding of International Research Network EURASIGN (2019–2022) for the year 2020 (accounting document no. 233807).

ACKNOWLEDGMENTS

We thank the deaf signers involved in this study from the various countries. We acknowledge the authors and annotators of the corpora for permission to use the source material in this article: Alessio Di Renzo, Ivani Fusellier, Robert Gavrilescu, Alexandre Daniel, Alain Gebert, Pamela Lattapiat, Julie Szczepaniak, Myriam Vermeerbergen, Mieke Van Herreveghe, and colleagues of the VGT corpus, Christian Cuxac and colleagues of the LS-COLIN corpus. We also thank Dana Cohen for her contribution to the translation process of this paper from French to English.

REFERENCES

Ahlgren, I. (1990). "Deictic pronouns in Swedish and Swedish Sign Language," in *Theoretical Issues in Syntactic Theory*, eds S. D. Fischer, and P. Siple, (Chicago: University of Chicago Press), 167–174.

Apothéloz, D. (1995). *Rôle et Fonctionnement de L'anaphore Dans La Dynamique Textuelle*. Genève: Droz.

Apothéloz, D., and Pekarek Doehler, S. (2003). Nouvelles perspectives sur la référence: des approches informationnelles aux approches interactionnelles. *Verbum* 25, 109–135.

Ariel, M. (1988). Referring and accessibility. *J. Linguist.* 24, 65–87. doi: 10.1017/s0022226700011567

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Armstrong, D., Stokoe, W. C., and Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge, MA: Cambridge University Press.

Benveniste, E. (1966). *Problèmes de Linguistique Générale*. Paris: Gallimard.

Benveniste, E. (1970). L'appareil formel de l'énonciation. *Langages* 5, 12–18. doi: 10.3406/lgge.1970.2572

Benveniste, E. (1974). *Problèmes de Linguistique Générale*. Paris: Gallimard.

Brennan, M. (1992). "The visual world of British sign language. An introduction," in *Dictionary of British Sign Language/English*, ed. D. Brien, (London: Faber and Faber), 1–118. doi: 10.1017/cbo9781139167048.002

Clark, H. (1996). *Using Language*. Cambridge, MA: Cambridge University Press.

Collins-Ahlgren, M. (1990). "Spatial-locative predicates in Thai Sign Language," in *Sign language research: Theoretical Issues*, ed. C. Lucas, (Washington, DC: Gallaudet University Press), 103–117.

Cormier, K., Smith, S., and Zwets, M. (2013). Framing constructed action in British Sign Language narratives. *J. Pragm.* 55, 119–139. doi: 10.1016/j.pragma.2013.06.002

Cornish, F. (1990). "Anaphore pragmatique, référence, et modèles du discours," in *L'anaphore et Ses Domaines*, eds G. Kleiber, and J.-E. Tyvaert, (Metz: Centre d'Analyse Syntaxique), 81–96.

Cornish, F. (2014). "Indexicals and context: context-bound pre-requisite(s), ongoing processing and aftermaths of the discourse referring act," in *Nouvelles Perspectives sur L'anaphore: Points de Vue Linguistique, Psycholinguistique et Acquisitionnel*, eds M. Fossard, and M.-J. Béguelin, (Berlin: Peter Lang), 5–33.

Croft, W. (2003). "Agreement as anaphora, anaphora as coreference," in *Languages Across Boundaries: Studies in Memory of Anna Siewerska*, eds D. Bakker, and M. Haspelmath Berlin, (Berlin: De Gruyter Mouton), 99–117.

Culioli, A. (1990). *Pour une Linguistique de L'énonciation*. Paris: Ophrys.

Culioli, A. (1995). "Cognition and representation in linguistic theory," in *Texts Selected*, ed. M. Liddle, (Amsterdam: John Benjamins).

Cuxac, C. (1985). "Esquisse d'une typologie des langues des signes," in *Autour de La Langue Des Signes*, ed. C. Cuxac, (Paris: Université René Descartes), 35–60.

Cuxac, C. (1993). Iconicité des langues des signes. *Faits de Langue.* 15, 47–56. doi: 10.3406/flang.1993.1034

Cuxac, C. (1999). "The expression of spatial relations and the spatialization of semantic relations in french sign language," in *Language Diversity and Cognitive Representations*, eds C. Fuchs, and S. Robert, (Amsterdam: John Benjamins), 123–142. doi: 10.1075/hcp.3.11cux

Cuxac, C. (2000). *La Langue des Signes Française (LSF). Les voies de l'iconicité*. Paris: Ophrys.

Cuxac, C. (2008). Langue des Signes et gestuelle co-verbale : pour un programme commun de recherches. *Les Cahiers de Linguistique Analogique* 5, 181–228.

Cuxac, C. (2013). Langues des signes : une modélisation sémiologique. *La Nouvelle Rev. l'adaptation et de la Scolarisation* 64, 65–80. doi: 10.3917/nras.064.0065

Cuxac, C., and Antinoro-Pizzuto, E. (2010). Emergence, norme et variation dans les langues des signes : vers une redéfinition notionnelle. *Lang. Soc.* 131, 37–53. doi: 10.3917/ls.131.0037

Cuxac, C., Braffort, A., Choisier, A., Collet, C., Dalle, P., Fusellier, I., et al. (2002). Available online at: https://cocoon.huma-num.fr/exist/crdo/ark:/87895/1.17-483699 (accessed September 1, 2020).

Cuxac, C., Fusellier-Souza, I., and Sallandre, M.-A. (1999). Iconicité des langues des signes et catégorisations. *Sémiotiques* 16, 143–166.

Cuxac, C., and Sallandre, M.-A. (2007). "Iconicity and arbitrariness, in French Sign Language: Highly Iconic Structures, degenerated iconicity and diagrammatic iconicity," in *Verbal and Signed Languages. Comparing Structures, Constructs and Methodologies*, eds E. Pizzuto, P. Pietrandrea, and R. Simone, (Berlin: Mouton de Gruyter), 13–33. doi: 10.1075/ill.12.04tur

De Voguë, S. (1992). Culioli après Benveniste : énonciation, langage, intégration. *Linx* 26, 77–108. doi: 10.3406/linx.1992.1238

Deuchar, M. (1984). *British Sign Language*. London: Routledge and Kegan Paul.

Ducrot, O. (1972). *Dire et ne pas dire*. Paris: Hermann.

Ducrot, O. (1980). *Les Mots du discours*. Paris: Minuit.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge, MA: Cambridge University Press.

Enfield, N. J. (2013). "Reference in conversation," in *The Handbook of Conversation Analysis*, eds J. Sidnell, and T. Stivers, (Malden, MA: Wiley-Blackwell), 433–454. doi: 10.1002/9781118325001.ch21

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum Press.

Engberg-Pedersen, E. (2003). "How composite is a fall? Adults' and children's descriptions of different types of falls in danish sign language," in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey, (Mahwah, NJ: Lawrence Erlbaum Associates), 311–332.

Fauconnier, G., and Turner, M. (1996). "Blending as a central process of grammar," in *Conceptual Structure, Discourse and Language*, ed. A. E. Goldberg, (Stanford: Center for the Study of Language and Information Publications), 113–130.

Fenlon, J., Cormier, K., and Brentari, D. (2018). "The phonology of sign languages," in *The Routledge Handbook of Phonological Theory*, eds S. J. Hannahs, and A. Bosch, (New York, NY: Routledge), 453–475.

Ferrara, L., and Hodge, G. (2018). Language as description, indication, and depiction. *Front. Psychol.* 9:716. doi: 10.3389/fpsyg.2018.00716

Fillmore, C. (1975). *Lectures on Deixis*. Santa Cruz: Indiana University Linguistics Club.

Friedman, L. (1975). On the semantics of space, time and person in American Sign Language. *Language* 51, 940–961. doi: 10.2307/412702

Friedman, L. (1977). *On the other hand: New Perspectives on American Sign Language*. New York, NY: Academic Press.

Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language* 51, 676–710.

Fuchs, C. (2008). "Linguistique française et cognition," in *Congrès Mondial de Linguistique Française 021*. doi: 10.1051/cmlf08340

Fusellier-Souza, I. (2004). *Sémiogenèse des Langues des Signes : Étude de Langues des Signes Primaires (LSP) Pratiquées Par des Sourds Brésiliens*. Ph. D. thesis, University of Paris, Saint-Denis.

Fusellier-Souza, I. (2006). Emergence and development of Signed Languages: from diachronic ontogenesis to diachronic phylogenesis. *Sign Lang. Stud.* 7-1, 30–56.

Fusellier-Souza, I. (2012). "Multiple perspectives on the emergence and development of human language: B. Comrie, C. Perdue and D. Slobin," in *Comparative perspectives on language acquisition: a tribute to Clive Perdue*, eds M. Watorek, S. Benazzo, and M. Hickmann, (Bristol: Multilingual Matters), 223–244. doi: 10.21832/9781847696045-014

Garcia, B. (2010). *Sourds, Surdité, Langue(s) Des Signes et Épistémologie des Sciences du Langage. Problématiques de la Scripturisation et Modélisation des Bas Niveaux en Langue des Signes Française (LSF)*. Habilitation thesis, University of Paris, Saint-Denis.

Garcia, B., L'Huillier, M.-T., and Vincent, C. (2015). *Corpus Creagest. Dialogue en LSF Entre Adultes Sourds*. Available online at: https://www.ortolang.fr/market/corpora/ortolang-000926 (accessed September 1, 2020).

Garcia, B., and Sallandre, M.-A. (2014). "Reference resolution in French Sign Language (LSF)," in *Crosslinguistic Studies on Noun Phrase Structure and Reference*, eds P. Cabredo Hofherr, and A. Zribi-Hertz, (Boston: Brill), 316–364. doi: 10.1163/9789004261440_012

Garcia, B., Sallandre, M.-A., and L'Huillier, M.-T. (2018). Impersonal human reference in French Sign Language (LSF). *Sign Lang. Linguist.* 21, 308–334.

Goldin-Meadow, S. (1991). "When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents," in *Tools, Language and Cognition in Human Evolution*, eds K. R. Gibson, and T. Ingold, (Cambridge, MA: Cambridge University Press), 63–85.

Goldin-Meadow, S. (2003). *The Resilience of Language*. New York, NY: Psychology Press.

Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign, and language: the coming of age of sign language and gesture studies. *Cambridge Core Behav. Brain Sci.* 40:e46. doi: 10.1017/S0140525X1600039X

Hockett, C. F. (1978). In search of Jove's brow. *Am. Speech* 53, 243–313. doi: 10.2307/455140

Hodge, G., Ferrara, L. N., and Anible, B. D. (2019). The semiotic diversity of doing reference in a deaf signed language. *J. Pragm.* 143, 33–53. doi: 10.1016/j.pragma.2019.01.025

Jacob, S. (2007). *Le Mouvement Référentiel Dans Des Narrations Enfantines en LSF : Conduite Descriptive Selon Une Trajectoire Développementale*. Ph.D. thesis, University of Paris, Saint-Denis.

Jakobson, R. (1957). *Shifters, Verbal Categories and the Russian Verb*. Cambridge, MA: Harvard University Press.

Jirou-Sylla, G. (2008). Description d'une langue des signes informelle micro-communautaire. Analyse lexicale du parler gestuel de Mbour (Senegal). *Cahiers Linguist. Analog.* 5, 135–180.

Johnston, T. (1991). Transcription and glossing of sign language texts: examples from Auslan (Australian sign language). *Int. J. Sign Linguist.* 2:1.

Johnston, T. (2008). "Corpus linguistics and signed languages: no lemmata, no corpus," in *Proceedings of The 3rd Workshop on the Representation and Processing of Sign Languages, (LREC 2008)*, Marrakech, 82–87.

Johnston, T. (2012). Lexical frequency in signed languages. *J. Deaf Stud. Deaf Educ.* 17, 163–193.

Johnston, T. (2019). *Auslan Corpus Annotation Guidelines [August 2019 Version]*. Sydney: Macquarie University.

Johnston, T., and Schembri, A. (1999). On defining lexeme in a sign language. *Sign Lang. Linguist.* 2, 115–185. doi: 10.1075/sll.2.2.03joh

Johnston, T., and Schembri, A. (2007). *Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics*. Cambridge, MA: Cambridge University Press.

Jouison, P. (1986/1995). *Écrits sur la Langue des Signes Française. Edition prepared by B. Garcia*. Paris: L'Harmattan.

Kegl, J. (1976). *Pronominalization in American Sign Language*. Master's thesis, MIT, Cambridge, MA.

Kegl, J., and Wilbur, R. (1976). "When does structure stop and style begin? Syntax, morphology, and phonology vs stylistic variation in American Sign Language," in *Papers from the Twelfth Regional Meeting*, eds C. Mufwene, and S. Steever, (Chicago: University of Chicago Press), 376–396.

Kendon, A. (1988). "How gestures can become like words," in *Cross-Cultural Perspectives in Nonverbal Communication*, ed. F. Poyatos, (Toronto: Hogrefe), 131–141.

Lakoff, G., and Johnson, M. L. (1980/1995). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Volume I, Theoretical Foundations*. Stanford: Stanford University Press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Volume II, Descriptive Application*. Stanford: Stanford University Press.

Levinson, S. C. (1983). *Pragmatics*. Cambridge, MA: Cambridge University Press.

Liddell, S. K. (1995). "Real, surrogate, and token space: grammatical consequences in ASL," in *Language, Gesture and Space*, eds K. Emmorey, and J. Reilly, (Hillsdale, NJ: Lawrence Erlbaum Associates), 19–41.

Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge, MA: Cambridge University Press.

Lillo-Martin, D., and Klima, E. S. (1990). "Pointing out differences: ASL pronouns in syntactic theory," in *Theoretical Issues in Syntactic Theory*, eds S. D. Fischer, and P. Siple, (Chicago, IL: University of Chicago Press), 191–210.

Lombardi Vallauri, E. (2007). "The deep relation between deixis and anaphora," in *Verbal and Signed Languages. Comparing Structures, Constructs and Methodologies*, eds E. Pizzuto, P. Pietrandrea, and R. Simone, (Berlin: Mouton de Gruyter), 309–338.

Lyons, J. (1977). *Semantics. Volume II*. Cambridge, MA: Cambridge University Press.

Mandel, M. (1977). "Iconic devices in american sign language," in *On the Other Hand*, ed. L. Friedman, (New York, NY: Academic Press), 57–107.

Martinod, E. (2019). *Approche Typologique Des Composants Minimaux Porteurs De Sens Dans Plusieurs Langues Des Signes (LS) Se Situant à Divers Degrés De Communautarisation*. Ph.D. thesis, University of Paris, Saint-Denis.

Martinod, E., Garcia, B., and Fusellier, I. (2020). "Meaningful handshapes in the emerging sign languages on marajó island (Brazil) in a typological perspective," in *Emerging Sign Languages of the Americas*, eds O. Le Guen, J. Safar, and M. Coppola, (Berlin: Mouton de Gruyter).

Mayer, M. (1969). *Frog, Where Are You?*. New York, NY: Dial Press.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: Chicago University Press.

Meier, R. P. (1990). "Person deixis in American Sign Language," in *Theoretical Issues in Syntactic Theory*, eds S. D. Fischer, and P. Siple, (Chicago: University of Chicago Press), 175–190.

Metzger, M. (1995). "Constructed dialogue and constructed action in American Sign Language," in *Sociolinguistics in Deaf Communities*, ed. C. Lucas, (Washington, DC: Gallaudet University Press), 255–271.

Newport, E., and Supalla, T. (2000). "Sign language research at the millennium," in *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, eds K. Emmorey, and H. Lane, (Mahwah, NJ: Lawrence Erlbaum Associates), 103–114.

Occhino, C. (2016). *A Cognitive Approach to Phonology: Evidence from Signed Languages*. Ph.D. thesis, University of New Mexico, Albuquerque.

Occhino, C. (2017). An introduction to embodied cognitive phonology: claw-5 handshape distribution in ASL and libras. *Complutense J. English Stud.* 25, 69–103. doi: 10.5209/CJES.57198

Occhino, C., and Wilcox, S. (2017). "Gesture or Sign? A categorization problem," in *Gesture, Sign and Language: The Coming of Age of Sign Language and Gesture Studies. Behavioral and Brain Sciences*, S. Goldin-Meadow, and D. Brentari, (Cambridge, MA: Cambridge University Press), 36–37.

Padden, C. (1986). "Verbs and role-shifting in ASL," in *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching*, ed. C. Padden, (Silver Spring, MD: National Association of the Deaf), 44–57.

Padden, C. (1990). "The relation between space and grammar in ASL verb morphology," in *Sign Language Research: Theoretical Issues*, ed. C. Lucas, (Washington, DC: Gallaudet University Press), 118–132.

Pizzuto, E. (1986). "The verb system of italian sign language (LIS)," in *Proceedings of the Second European Congress on Sign Language Research*, ed. B. T. Teervort, (Amsterdam: the Institute of General Linguistics), 17–31.

Pizzuto, E. (2007). "Deixis, anaphora and person reference in signed languages," in *Verbal and Signed Languages. Comparing Structures, Constructs and Methodologies*, eds E. Pizzuto, P. Pietrandrea, and R. Simone, (Berlin: Mouton de Gruyter), 275–308.

Pizzuto, E., Pietrandrea, P., and Simone, R. (eds) (2007). *Verbal and Signed Languages. Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter.

Pizzuto, E., Rossini, P., Sallandre, M.-A., and Wilkinson, E. (2008). "Deixis, anaphora and highly iconic structures: cross-linguistic evidence on American (ASL), French (LSF) and Italian (LIS) signed languages," in *Sign Languages: Spinning and Unraveling the Past, Present and Future*, ed. R. Muller de Quadros, (Petrópolis: Editora Arara Azul), 475–495.

Pizzuto, E., and Volterra, V. (2000). "Iconicity and transparency in sign languages : a cross-linguistic cross-cultural view," in *The Signs of Language Revisited: An Anthology in Honor of Ursula Bellugi and Edward Klima*, eds K. Emmorey, and H. Lane, (Mahwah, NJ: Lawrence Erlbaum Associates), 261–286.

Sallandre, M.-A. (2003). *Les Unités du Discours en Langue des Signes Française. Tentative de Catégorisation Dans le Cadre D'une Grammaire de L'iconicité*. Ph.D. thesis, University of Paris, Saint-Denis.

Sallandre, M.-A. (2007). "Simultaneity in French sign language discourse," in *Simultaneity in Signed Languages: Form and Function*, eds M. Vermeerbergen, L. Leeson, and O. Crasborn, (Amsterdam: John Benjamins), 103–125. doi: 10.1075/cilt.281.05sal

Sallandre, M.-A. (2014). *Compositionnalité des Unités Sémantiques en Langues Des Signes. Perspective Typologique et Développementale*. Habilitation thesis, University of Paris, Saint-Denis.

Sallandre, M.-A. (2020). Comparaisons typologiques entre les langues des signes, une approche sémiologique. *Lalies* 40, 9–30.

Sallandre, M.-A., Balvet, A., Besnard, G., and Garcia, B. (2019). *Etude Exploratoire de la Fréquence des Catégories linguistiques dans Quatre Genres Discursifs en LSF. LIDIL. 60*. Available online at: https://journals.openedition.org/lidil/7136 (accessed September 1, 2020).

Sallandre, M.-A., Di Renzo, A., and Gavrilescu, R. (2016). "Various types of personal transfers (constructed actions) in seven sign languages," in *Proceedings*

of the Theoretical Issues in Sign Language Research Conference (TISLR 12), (Melbourne: La Trobe University).

Schembri, A. (2003). "Rethinking 'Classifiers' in signed languages," in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey, (Mahwah, NJ: Lawrence Erlbaum Associates), 3–34.

Sidnell, J., and Enfield, N. (2016). "Deixis and the interactional foundations of reference," in *The Oxford Handbook of Pragmatics*, ed. Y. Huang, (Oxford: Oxford University Press), 217–239.

Sloetjes, H., and Wittenburg, P. (2018). *ELAN (version 5.9)*. Nijmegen: Max Planck Institute for Psycholinguistics.

Smith, C., Lentz, E. M., and Mikos, K. (1988). *Signing Naturally: Teacher's Curriculum Guide. Level I*. Berkeley, CA: DawnSign Press.

Stokoe, W. C. (1991). Semantic phonology. *Sign Lang. Stud.* 71, 107–114. doi: 10.1353/sls.1991.0032

Supalla, T. (1978). "Morphology of verbs of motion and location in american sign language," in *American Sign Language in a Bilingual, Bicultural Context*, eds F. Caccamise, and D. Hicks, (Coronado, CA: National Association of the Deaf), 27–46.

Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language. An Introduction*. Cambridge, MA: Cambridge University Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Vermeerbergen, M. (2006). Past and current trends in sign language research. *Lang. Com.* 26, 168–192. doi: 10.1016/j.langcom.2005.10.004

Wilbur, R. (1979). *American Sign Language and Sign Systems: Research and Application*. Baltimore: University Park Press.

Wilcox, S. (1996). Not from Jove's brow. *Lang. Commun.* 16, 179–192. doi: 10.1016/0271-5309(96)00006-7

Wilcox, S., and Wilcox, P. (1995). "The gestural expression of modality in American Sign Language," in *Modality in Grammar and Discourse*, eds J. Bybee, and S. Fleischman, (Amsterdam: John Benjamins), 135–162. doi: 10.1075/tsl.32.07wil

Winston, E. (1991). Spatial referencing and cohesion in an American Sign Language text. *Sign Lang. Stud.* 73, 397–410. doi: 10.1353/sls.1991.0003

Winston, E. (1995). "Spatial mapping in comparative discourse frames," in *Language, Gesture and Space*, eds K. Emmorey, and J. Reilly, (Hillsdale, NJ: Lawrence Erlbaum Associates), 97–114.

Woll, B., and Sutton-Spence, R. (2005). "Sign languages," in *Sociolinguistics: an International Handbook of the Science of Language and Society*, eds U. Ammon, N. Dittmar, K. Mattheier, and P. Trudgill, (Berlin: Walter de Gruyter), 677–683.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 November 2020
doi: 10.3389/fpsyg.2020.553226

Interpreting Estonian Demonstratives: The Effects of Referent's Distance and Visual Salience

Maria Reile[1], Kristiina Averin[2,3] and Nele Põldver[2]*

[1] Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia, [2] Institute of Psychology, University of Tartu, Tartu, Estonia, [3] Doctoral School of Behavioural, Social and Health Sciences, University of Tartu, Tartu, Estonia

Most of the research done with spatial demonstratives (words such as *this*, *here* and *that*, *there*) have focused on the production, not the interpretation, of these words. In addition, emphasis has been largely on demonstrative pronouns, leaving demonstrative adverbs with relatively little research attention. The present study explores the interpretation of both demonstrative pronouns and demonstrative adverbs in Estonian—a Finno-Ugric language with two dialectal-specific demonstrative pronoun systems. In the South-Estonian (SE) dialectal region, two demonstrative pronouns, *see*—"this" and *too*—"that", are used. In the North-Estonian (NE) region, only one, *see*—"this/that", is used. The aim of this study is twofold. First, we test if the distance and the visual salience of a referent have an effect on the interpretation of demonstratives. Second, we explore if there is a difference in the interpretation of demonstratives between native speakers from SE and NE. We used an interpretation experiment with 30 participants per group (total n = 60) and compared the SE and NE group responses. The results clearly show that the distance of the referent has an effect on how demonstratives are interpreted across the two groups, while the effect of visual salience is inconclusive. There is also a difference in the interpretation of demonstratives between the two dialectal groups. When using the Estonian with an influence of the SE dialect, the NE speakers rely on demonstrative adverbs in interpreting the referential utterance that includes demonstrative pronoun and adverb combinations, whereas the SE speakers also take into account the semantics of demonstrative pronouns. We show that, in addition to an already known difference in the production, there is also a difference in the interpretation of demonstratives between the two groups. In addition, our findings support the recognition that languages that have distance neutral demonstrative pronouns enforce the spatial meaning of a referring utterance by adding demonstrative adverbs. Not only is the interpretation of demonstrative pronouns affected, but the interpretation of demonstrative adverbs as well. The latter shows the importance of studying adverbs also, not just pronouns, and contributes to further knowledge of how demonstratives function.

Keywords: spatial demonstratives, demonstrative pronouns, demonstrative adverbs, referent distance, visual salience, experimental linguistics, interpretation experiment

OPEN ACCESS

Edited by:
Olga Capirci,
Italian National Research Council, Italy

Reviewed by:
David Peeters,
Tilburg University, Netherlands
Mikkel Wallentin,
Aarhus University, Denmark

****Correspondence:***
Maria Reile
maria.reile@ut.ee

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *17 April 2020*
Accepted: *01 October 2020*
Published: *20 November 2020*

Citation:
Reile M, Averin K and Põldver N (2020) Interpreting Estonian Demonstratives: The Effects of Referent's Distance and Visual Salience. Front. Psychol. 11:553226. doi: 10.3389/fpsyg.2020.553226

INTRODUCTION

Demonstratives, such as *this* and *there* in English, are one of the core elements of language as they belong to one of the first words that children acquire (Clark and Sengul, 1978), and they are used to indicate objects in the surroundings of interlocutors (Diessel, 1999, 2013). Diessel (1999) has even proposed that, in every language, there are at least two spatially contrastive demonstratives, demonstrative pronouns, demonstrative adverbs, or demonstrative particles. In this sense, demonstratives can be seen as language universals. However, their functions can differ between languages (e.g., Diessel, 1999; Dixon, 2003), and this makes them an interesting linguistic phenomenon. For instance, in some languages, demonstratives can indicate whether a referent is invisible (as in Khasi language) (Nagaraja, 1985 cited in Diessel, 1999), located down-river or up-river, down-hill or up-hill (as in Dyirbal language) (Dixon, 1972 cited in Diessel, 1999), while in other languages there are no specific demonstratives that would fulfill these functions. Also, in addition to distance indication, demonstratives can be used to express whether the intended referent is in the visual attention of the hearer, such as *şu* in Turkish (Özyürek, 1998; Küntay and Özyürek, 2006), or if the referent is in a joint focus of attention of the interlocutors (Diessel, 2006).

In addition to the different functions that demonstratives can fulfill, there are also different demonstrative pronoun systems. Diessel (1999) classifies demonstrative pronoun systems on the bases of the number of distance contrasts that adnominal demonstratives (demonstrative pronouns with an accompanying noun) make. This means that there can be demonstrative pronoun systems with one-way distance contrast (in these, demonstrative pronouns are distance-neutral, such as in German and in French) and demonstrative pronoun systems with even five distance-contrasts (such as in Koasati) (Diessel, 2013). There is a tendency that the more demonstrative pronouns a system has, the more different aspects of the referent the demonstrative pronouns express.

Empirical research has shown that, in spatial use, demonstratives indicate the distance of the referent from the speaker and the hearer (e.g., Dixon, 2003; Coventry et al., 2008; Diessel, 2013; Levinson, 2018). Moreover, there seems to be a connection between spatial perception and demonstratives as well as memory for object location (Coventry et al., 2008, 2014; Caldano and Coventry, 2019; Gudde et al., 2016). For example, it has been shown that in English and in Spanish the use of so-called distal demonstratives (*that* in English and *aquel* in Spanish) increased when the referent's distance increased (Coventry et al., 2008). In other words, when the referent was situated in the participant's extrapersonal space—the space outside one's grasping distance (di Pellegrino and Làdavas, 2015)—then distal demonstratives were used in referring to that object. Similar results have been found for other languages as well, such as Estonian and Võro language (Reile et al., 2020). In addition, in English, when an object is referred to with a distal demonstrative, then its location is remembered to be more distant than it actually was (Gudde et al., 2016). This highlights the importance of the referent's distance from the speaker in the choice and use of demonstrative pronouns.

Distance, however, is not the only factor contributing to the choice of demonstratives. Several authors have shown that also the visual salience of the referent or visual access to the referent (Diessel, 1999; Jarbou, 2010; Coventry et al., 2014) affects the use of demonstratives. Joint attention between interlocutors can have an effect on the use (Diessel, 2006) and the interpretation (Stevens and Zhang, 2013) of demonstratives. For example, in English, visually inaccessible objects are referred to with *that* (a distal demonstrative) (Coventry et al., 2014), whereas *this* can be interpreted that the interlocutors share a joint focus of attention (Stevens and Zhang, 2013). Nevertheless, the degree of which these factors influence the use of demonstratives differs between languages. For instance, in Khasi, a Mon Kher language, there is a specific demonstrative to express the invisibility of the referent (Nagaraja, 1985 cited in Diessel, 1999). In other languages, such as English, there are no specific demonstratives for this function, but the use of demonstratives is still influenced by these factors. In Estonian, the visual salience of the referent does not influence the choice of demonstrative pronouns, as in English, but seems to have an effect on how demonstrative adverbs are used (Reile, 2016, 2019). Therefore, the complexity of how different factors actually influence and how they contribute to demonstrative use is not yet fully understood.

While the empirical research on spatial demonstratives has increased, most of these studies have used production experiments to tackle the factors that have an effect on demonstrative use (e.g., Coventry et al., 2008, 2014; Piwek et al., 2008; Peeters et al., 2014; Tóth et al., 2014; Gudde et al., 2016). Nevertheless, there are some studies that focus on the interpretation of demonstrative pronouns (Bonfiglioli et al., 2009; Stevens and Zhang, 2013, 2014; Peeters et al., 2015; Rocca et al., 2020). These studies have shown that distance is not the only factor that can play a role in the interpretation of demonstratives. While distance has been shown to have an effect on the interpretation of demonstratives in several languages, i.e., the incongruent use of demonstratives (using a proximal instead of a distal one in referring to an object outside grasping distance) causes longer reaction times in participants' responses in Italian, English, and Japanese (Bonfiglioli et al., 2009; Stevens and Zhang, 2013, 2014); other factors, such as shared space of the interlocutors, can override the effects of egocentric distance, such as in Dutch (Peeters et al., 2015). In addition, the effects of distance can be relative in the sense that, when two referents are located in the peripersonal space of the participant, it is not appropriate to refer to both referents with the proximal pronoun (Bonfiglioli et al., 2009), at least in Italian. Even more so, a recent naturalistic fast fMRI experiment in Danish has shown that, while demonstratives are processed in the areas of the brain connected to visuospatial cognition, no statistically significant segregation was found between processing distal and proximal demonstratives (Rocca et al., 2020). Thus, similarly to the production of demonstratives, in the interpretation of demonstratives the effects of distance are also not as straightforward as previously thought.

The current study uses an interpretation experiment to pinpoint the factors that can have an effect on understanding demonstrative meaning in spatial reference, i.e., the use of spatial demonstratives, both demonstrative determiners and demonstrative adverbs, which has been seldom done with this methodological approach. We focus on Estonian which is a Finno-Ugric language that employs at least two demonstrative pronoun systems (Pajusalu, 2009). The use of these systems is related to the historical division of Estonian dialects (Pajusalu et al., 2009). In the North-Estonian (NE) dialectal region, a one-term system is used. This means that the sole demonstrative pronoun that is used is *see*—"this/that"—a distance-neutral demonstrative that refers to any referent regardless of its distance from the speaker. In this one-term demonstrative pronoun system, spatial contrasts are expressed through the use of demonstrative adverbs (see **Table 1**) (Pajusalu, 2009; Reile, 2015).

In the South-Estonian (SE) dialectal region, two demonstrative pronouns *see* and *too* are used. In this demonstrative pronoun system, *see* is the proximal and *too* is the distal demonstrative pronoun. However, *too* has a stronger anchorage to far distance than *see* to near distance (Reile, 2019; Reile et al., 2020). As in the one-term system, demonstrative pronouns can be accompanied by demonstrative adverbs also in the two-term system (see **Table 1**). While it is not impossible for a distal demonstrative pronoun to be combined with a proximal demonstrative adverb, it is still more common to be combined with a distal demonstrative adverb (Reile, 2016).

Both demonstrative pronouns are used as determiners in both demonstrative pronoun systems. Both demonstrative pronouns are also present in the written language of standard Estonian. However, *see* is far more frequent than *too* (Reile, 2019, p. 29). This suggests that while the Estonian speakers originating from the NE region have an exposure to the demonstrative pronoun *too*, at least in written form, it is highly likely that their interpretation of this demonstrative in spatial reference is different as compared to the Estonian speakers from the SE region.

Previous studies on the production of Estonian demonstratives have shown that, while distance has a straightforward effect on the choice of Estonian demonstratives (Reile et al., 2019, 2020), the effect of visual salience might manifest itself in a more indirect way, that is, rather than influencing the choice between distal or proximal demonstratives, the position of demonstrative adverbs in the word order of a referential utterance is affected (Reile, 2016). In other words, in referring to visually non-salient referents, the distal demonstrative adverb *seal*—"there" precedes the referential noun phrase (NP), and in referring to visually salient referents, the distal demonstrative *seal*—"there" comes after the referential NP (that might also include a demonstrative pronoun). For example, in referring to a visually non-salient book, one might say "*Vaata, seal see/too raamat*!" with a direct translation "Look, there this/that book!", and for a visually salient book, "*Vaata, see/too raamat seal*!" with a direct translation "Look, this/that book there!" In the current study, we manipulate the word order of the input sentence to put this finding under test.

TABLE 1 | Estonian demonstrative pronouns and demonstrative adverbs.

	One-term system (NE)	Two-term system (SE)
Demonstrative pronouns	*see*—"this"/"that"	*see*—"this"
		too—"that"
Demonstrative adverbs	*siia*—"to here"–*sinna*—"to there"	
	siin—"here"–*seal*—"there"	
	siit—"from here"–*sealt* from there"	

NE, North Estonia; SE, South Estonia.

To find out how participants interpret Estonian demonstrative determiners and adverbs, we conducted an interpretation experiment. The aim was to detect a possible association between the distance of the referent, the visual salience of the referent, and the interpretation of demonstratives. In other words, we were interested in whether some demonstratives in the demonstrative paradigm are preferred more for distant/visually non-salient referents than others.

Considering all the above-mentioned points, we have proposed the following hypotheses:

(1) The distance of the referent has an effect on demonstrative interpretation: when the demonstrative pronoun *too*—"that" or the adverb *sealt*—"(from) there" are heard, a distant referent is chosen, and when the demonstrative pronoun *see*—"this" and the adverb *siit*—"(from) here" are heard, a proximal referent is chosen.
(2) With demonstrative pronoun and adverb combinations, the choice of referent is based on the demonstrative adverbs when the visual scene is incongruent with the possible meaning of a demonstrative pronoun in a heard sentence, i.e., when a distal demonstrative pronoun is heard but the referents are in near space and when a proximal demonstrative pronoun is heard but the referents are in far space.
(3) The visual salience of the referent has an effect on demonstrative interpretation: with demonstrative pronoun and adverb combinations, the visually non-salient referent is chosen when the adverb precedes the pronoun in a heard sentence, and the visually salient referent is chosen when the pronoun precedes the adverb in a heard sentence.
(4) The choices for the referents between the NE and the SE speakers differ when the demonstrative *too* is heard. The SE speakers will show a pattern of choosing the farthest referent of the possible referents, while the NE speakers may choose any referent regardless of their distance.
(5) The reaction times (RTs) for choosing a referent when the demonstrative *too*—"that" is heard (with or without a demonstrative adverb) are slower for the NE speakers than for the SE speakers.

MATERIALS AND METHODS

The interpretation experiment consisted of two conditions. First, we tested the effects of the referent's distance from the speaker on the interpretation of demonstratives. Second, we looked for the effects of the visual salience of the referent on the interpretation of demonstratives.

Participants

Sixty volunteer participants (mean age, 29.7 years; *SD* = 6.5 years) with normal or corrected-to-normal vision took part in the experiment. Thirty participants originated from the one-way demonstrative system (NE) region and 30 from the two-way system (SE) region. In both groups, there were seven males and 23 females. It was explicitly explained to all the participants that their participation was voluntary and that they could leave the experiment at any time point, and an oral consent for participation was acquired.

Stimuli and Design

We used the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and its integrated OpenGL commands for Matlab (MathWorks, Natick, MA, United States) to create 3D images (hereafter pictures), run the experiments, and record the data. Every stimulus contained a picture of a table with three green construction bricks (see **Figures 1**, **2**) and a blue rectangle in the upper-right corner. In pictures 1–6, we manipulated the distance of the bricks from the near end of the table. In pictures 7–9, we also manipulated the visual salience of the bricks.

In the distance condition of the experiment, we manipulated the location of referents in six pictures that were designed in a way that proportionally mimicked peri- and extrapersonal space division. The latter has been shown to have an effect on the choice of demonstratives in production experiments (e.g., Coventry et al., 2008, 2014; Reile et al., 2020). This means that, in each of the pictures, the bricks were located either in the supposed peripersonal space of the participant (i.e., near to the participant, henceforth near space) or outside of it (i.e., far from the participant, henceforth far space) (see **Figure 1**).

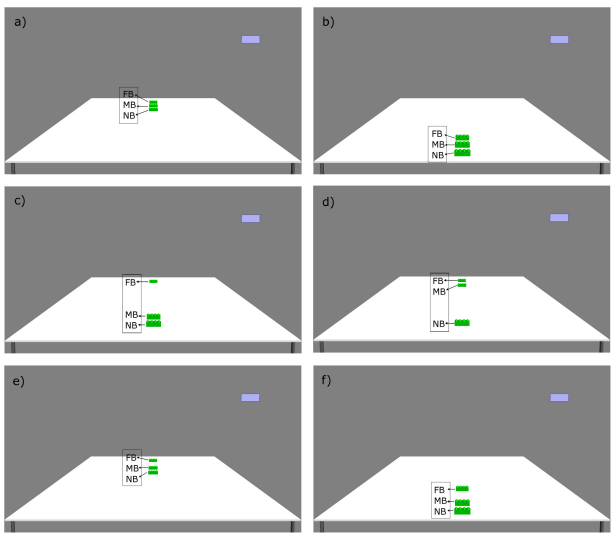


FIGURE 1 | The six picture stimuli used in the distance condition of the experiment, depicting the position of the bricks (green). The blue rectangle in the upper-right corner of the screen marks the response option in case of "do not know." Note that the text boxes (FB, MB, and NB, referring to the farthest, middle, and nearest brick, respectively) were not presented to the participants on the screen during the experiment. **(a)** Picture 1, **(b)** Picture 2, **(c)** Picture 3, **(d)** Picture 4, **(e)** Picture 5, and **(f)** Picture 6.

In the distance condition of the experiment, the three bricks were positioned on the table in the near space or the far space of the participants. In picture 1, all the three bricks were situated in the near space, that is, at the near the edge of the table. In picture 2, all the bricks were in the far space, which is at the far end of the table. In picture 3, two bricks were in the near space and one in the far space, and in picture 4, two bricks were in the far space and one in the near space. In picture 5, all the bricks were in far space, but one of them was a little farther away than the other two. In picture 6, all the bricks were in near space, but similarly to picture 5, one of the bricks was a little bit farther away than the other two (see **Figure 1**, pictures 1–6).

In the visual salience condition, we grouped the three referents together to create a figure-ground setting, that is, one in front (visually salient) and two at the back (visually non-salient). In addition to the visual salience, we manipulated the distance of the referent groups. In each picture, the whole brick group was either in the supposed peripersonal space of the participants or outside of it (see **Figure 2**). In picture 7, the grouped bricks were in far space. In picture 8, the bricks were nearer than in picture 7, but still in far space, and in picture 9, the bricks were placed in near space (see **Figure 2**, pictures 7–9). Changing the location of the brick group enabled us to test whether the visual salience effect would override the distance effect.

The participants' task was to look at the picture (one at a time) and choose a brick from the picture by clicking on the brick with a computer mouse. The participants had to make their choice based on the input sentence that they heard from the headphones. When it seemed that the sentence they heard did not match any of the bricks that they saw, they were allowed to click on the blue rectangle on the upper-right corner of the screen, which meant a response "do not know."

The input sentences that the participants heard were recorded by a female voice and went through an acoustic correction in Praat (Boersma and Weenink, 2007) to exclude the possible effect of intonation on the participants' choice of the referents. To do that, we overlaid each sentence with a neutral statement intonation contour with downstepped fundamental frequency (F0) peaks on the non-pronominal/content words (declining F0 peaks). The longer sentences (e.g., *võta sealt see klots*—"take from there this brick") were resynthesized with three F0 peaks, the shorter sentences (*võta see klots*—"take this brick") included only two F0 peaks. More specifically, we marked the onsets and offsets of every phrase and stressed vowel. F0 at the beginning of the sentence was set at 270 Hz, and at the end of the sentence, it was 190 Hz, regardless of sentence length. The F0 peaks were aligned with one-third of the vowel duration into the vowel. The peak heights from first to the final content word in

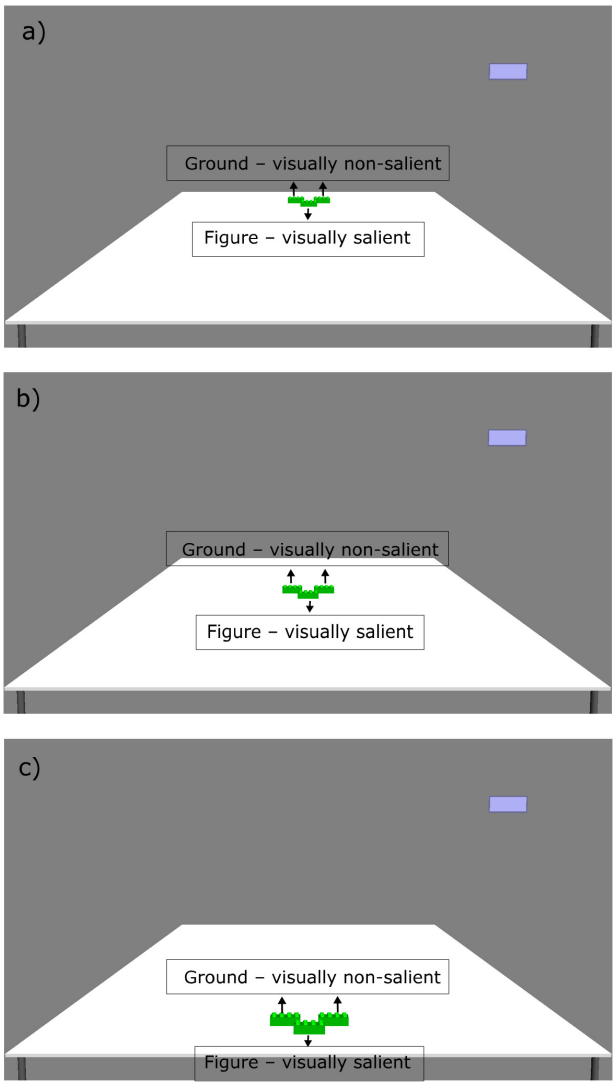


FIGURE 2 | The three picture stimuli used in the visual salience condition of the experiment, depicting the position of the bricks (green). The blue rectangle in the upper-right corner of the screen marks the response option in case of "do not know." Note that the text boxes (Figure, Ground) were not presented to the participants on the screen during the experiment. **(a)** Picture 7, **(b)** Picture 8, and **(c)** Picture 9.

long sentences were set to 277, 240, and 230 Hz and in short sentences to 270 and 230 Hz. F0 contour between these values was obtained by quadratic interpolation as provided in Praat PSOLA resynthesis method.

The input sentences consisted of either only an adnominal demonstrative pronoun (where a demonstrative pronoun precedes the noun) or combinations of adnominal demonstrative pronouns and demonstrative adverbs. The input sentences were as follows:

(1) *Võta **see** klots*—"take **this** brick"
(2) *Võta **see** klots **siit***—"take **this** brick from **here**"
(3) *Võta **siit see** klots*—"take from **here this** brick"
(4) *Võta **see** klots **sealt***—"take **this** brick from **there**"
(5) *Võta **sealt see** klots*—"take from **there this** brick"
(6) *Võta **too** klots*—"take **that** brick"
(7) *Võta **too** klots **siit***—"take **that** brick from **here**"
(8) *Võta **siit too** klots*—"take from **here that** brick"
(9) *Võta **too** klots **sealt***—"take **that** brick from **there**"
(10) *Võta **sealt too** klots*—"take from **there that** brick"
(11) *Võta väike jänku*—"take the little bunny"

In addition to the input sentences, we also used a filler sentence (no. 11) for control to keep the participants alert throughout the experiment. The filler sentence occurred six times per experiment series.

We used a different order of demonstrative pronouns and demonstrative adverbs in the input sentences because it has been shown that there is a tendency to use demonstrative adverbs in the first position of a referential utterance for visually non-salient referents (Reile, 2016, 2019).

Procedure

The experiment took place in a semi-darkened room on a Dell Precision M6500 laptop with a screen diameter of 17″ and 1,440 × 900-pixel resolution. The participants were instructed to sit in front of the laptop, put on the headphones, rest their heads on a chinrest in front of them, and hold a computer mouse with their dominant hand.

Before the experiment, the eye-tracking system was calibrated, and the participants were presented with four test-trials. In the test-trials, we did not use any demonstratives but had object descriptive phrases, such as *võta kollane klots*—"take (a) yellow brick." The eye-tracking measurement data are not analyzed in the scope of the current paper.

The procedure of the experiment was as follows: first, a picture of a table with bricks appeared on the computer screen for 5 s, during which the participants heard the auditory input sentence. After hearing the input sentence, the mouse cursor appeared on the right side of the screen. The participants had to choose one of the three bricks that they thought was the best match to the sentence heard and click on it using the computer mouse. If the participants felt that the input sentence did not apply to any of the bricks, they could click on a blue rectangle on the upper-right corner of the screen, indicating an answer "do not know." The time starting from the appearance of the cursor until the response (mouse click) was measured in milliseconds (reaction time) and recorded in a text file together with the relevant information per trial: the chosen brick, trial number, condition, and participant ID. After the response was given, a white screen was presented for 1 s, which was thereafter followed by a new trial.

All the pictures and input sentence sequences were blocked and randomized. All the input sentences had three repetitions with each of the pictures (except the filler sentence which was presented six times in each experiment series). Thus, there were 186 (3 × 6 × 10 + 6 controls) trials for the distance series and 96 (3 × 3 × 10 + 6 controls) trials for the visual salience series of the experiment. To minimize the order effect, half of the participants

started the experiment with the distance condition and half with the visual salience condition.

Data Analysis

The data analysis was carried out in R software version 4.0.2 (R Core Team, 2020) using generalized linear mixed effects regression models (GLMM). Similarly to linear mixed models, GLMM allows incorporating fixed and random factors. When fixed factors account for the systematic variability, random factors allow considering the variability from sources other than those in the scope of the present research interest. However, unlike LMM, GLMM does not require the dependent variable to follow a normal distribution (Lo and Andrews, 2015). Thus, for the reaction time analyses, we built general mixed models using the lme4 package (Bates et al., 2015), and for analyzing the choices of the participants, we used the MCMCglmm package that generates GLMM by utilizing Markov chain Monte Carlo and Bayesian methods. This allows specifying a variance structure with prior distributions for fixed and random factors (Hadfield, 2010). The advantages of the package in analyzing choice data include an option to create multinomial models and an ability to deal with issues arising from complete separation. The latter may occur if, in some of the conditions, some levels of the dependent variable have zero choosing frequency (Hadeld, 2012). For response data analysis, whether the participants chose a referent or clicked on the "do not know" rectangle, we built a generalized linear mixed effects regression model using the lme4 package glmer (Bates et al., 2015).

All the models were built separately for distance and visual salience condition. In the models that were based on the RT data and the choice data, the dependent variable was the participants' reaction time when choosing a brick (for the RT data) or the brick chosen by the participants (for the choice data), respectively. In the distance condition, the dependent variable had three levels: the nearest brick (reference category), the middle brick, and the farthest brick. In the visual salience condition, the dependent variable had two levels: non-salient brick (reference category) and salient brick. The independent variables for both conditions, as well as for both RTs and choice-based models, were as follows: origin of the participants, stimulus picture, and the interaction between the two. Note that, in the visual salience condition, word order was also added as an independent variable. The origin of the participants was a binary variable with the levels NE (reference category) and SE. The stimulus picture had six levels in distance condition, pictures 1 (reference category)–6, and three levels in visual salience condition, pictures 7 (reference category)–9. The word order in visual salience condition had two levels: adverb preceding a pronoun (reference category) and pronoun preceding an adverb. In all the models, the participant's ID was added as a random effect, and in the RT data-based model, we also included the trial number. From these analyses, the "do not know" rectangle click responses were excluded.

In the models that were based on the response data in distance and visual salience conditions, the dependent variable was the response of the participants, either the "do not know" rectangle click (reference category) or the choice of a brick. The independent variables were the origin of the participant, stimulus picture (the levels are the same as in previous models for both variables), and the input sentence. The input sentence had four levels: *võta see klots siit*—"take this brik from here" (reference category), *võta see klots sealt*—"take this brick from there", *võta too klots siit*—"take that brick from here", and *võta too klots sealt*—"take that brick from there". Note that for a better comparison of the models, we left out the input sentences that included only demonstrative pronouns and merged the input sentences that included the same demonstratives but had a different word order (e.g., *võta see klots siit*—"take this brick from here" was merged with *võta siit see klots*—"take from here this brick").

RESULTS

Distance Condition: Choices of Bricks

We tested hypotheses 1, 2, and 4 using the data from the distance condition. As the aim of hypotheses 2 and 4 was to pinpoint the differences between the two participant groups in the interpretation of specific Estonian demonstratives, we built a separate model for each input sentence (see **Table 2**).

The input sentence *võta see klots siit*—"take this brick from here"

While the variable origin of the participants had no effect on the choice of the participants, the stimulus picture proved to be statistically significant. Pictures 2 and 4 decreased the likelihood for the participants to choose the middle or the farthest brick over the nearest brick as compared to picture 1 ($p < 0.001$ and $p < 0.01$, respectively, for with picture 2 and $p < 0.001$ and $p < 0.001$, respectively, for with picture 4). Picture 5 decreased the likelihood for the participants to choose the middle brick over the nearest brick ($p < 0.05$) but had no effect on the choice of the farthest brick. Picture 6 increased the likelihood for the participants to choose the farthest brick over the nearest brick ($p < 0.05$) but had no effect on the choice of the middle brick over the nearest brick. There were no statistically significant interactions.

The input sentence *võta too klots siit*—"take that brick from here"

Both independent variables origin of the participants and stimulus picture proved to be statistically significant in predicting the choice of the participants. The SE participants were more likely to choose the farthest or the middle brick over the nearest brick ($p < 0.001$ and $p < 0.01$, respectively) than the NE participants. Pictures 2–6 showed a statistically significant effect on the choice of the participants as compared to picture 1. Pictures 2 and 4 decreased the likelihood for the participants to choose the middle or the farthest brick over the nearest brick (all $p < 0.001$). Picture 3 increased the likelihood for the participants to choose the middle brick over the nearest brick ($p < 0.001$) but decreased the likelihood for the participants to choose the farthest brick over the nearest brick ($p < 0.05$). Picture 5 decreased the likelihood

TABLE 2 | Results of generalized linear mixed effects regression models predicting the choices of the participants in distance condition.

	võta see klots siit "take this brick from here"		*võta too klots siit* "take that brick from here"		*võta see klots sealt* "take this brick from there"		*võta too klots sealt* "take that brick from there"	
	Middle	**Farthest**	**Middle**	**Farthest**	**Middle**	**Farthest**	**Middle**	**Farthest**
Intercept	−3.30***	−3.16***	−1.59***	−1.34***	−0.35	3.25***	0.63	4.22***
Origin SE vs. NE	−0.31	0.11	0.99*	1.71***	1.22*	−0.47	0.42	1.41*
Picture 2 vs. 1	−2.10***	−1.44**	−1.35***	−1.97***	−1.45**	−3.56***	−1.14*	−3.05***
Picture 3 vs. 1	0.15	−0.35	1.13***	−0.63*	0.74	3.57***	1.25	3.39***
Picture 4 vs. 1	−3.11***	−4.26***	−2.51***	−4.32***	2.36***	−2.19***	1.48**	−1.56***
Picture 5 vs. 1	−1.98***	0.08	−0.05	−0.77**	−0.31	−0.73	1.17	1.08*
Picture 6 vs. 1	0.57	1.29***	0.98***	1.00***	1.55	2.84***	1.66	3.13***
Origin SE: picture 2 vs. origin NE: picture 1	NA	NA	NA	NA	−0.34	0.63	NA	NA
Origin SE: picture 3 vs. origin NE: picture 1	NA	NA	NA	NA	2.12*	0.40	NA	NA
Origin SE: picture 4 vs. origin NE: picture 1	NA	NA	NA	NA	−1.33*	0.42	NA	NA
Origin SE: picture 5 vs. origin NE: picture 1	NA	NA	NA	NA	−0.26	0.23	NA	NA
Origin SE: picture 6 vs. origin NE: picture 1	NA	NA	NA	NA	1.42	1.06	NA	NA
Observations	1,662		1,648		1,803		1,806	
Model accuracy (%)	85.61		62.74		76.32		86.82	

*The reference category of the dependent variable is the nearest brick. The accuracy of the models was calculated based on the agreement between the bricks actually chosen and the choices suggested by the probabilities predicted by the model. *p < 0.05, **p < 0.01, ***p < 0.001.*

for the participants to choose the farthest brick over the nearest brick ($p < 0.01$) but had no effect on the choice of the middle brick. Picture 6 increased the likelihood for the participants to choose the middle or the farthest brick over the nearest brick (both $p < 0.001$). The interaction terms were statistically not significant and thus excluded from the final model.

The input sentence *võta see klots sealt*—"take this brick from there"

Both independent variables origin of the participants and stimulus picture proved to be statistically significant in predicting the choices of the participants. The SE participants were more likely to choose the middle brick over the nearest brick ($p < 0.05$) than the NE participants, but there was no effect on the choice of the farthest brick. Pictures 2, 3, 4, and 6 showed a statistically significant effect on the choice of the participants as compared to picture 1. Picture 2 decreased the likelihood of the participants to choose the middle or the farthest brick ($p < 0.01$ and $p < 0.001$, respectively) over the nearest brick. Pictures 3 and 6 increased the likelihood for the participants to choose the farthest brick over the nearest brick (both $p < 0.001$) but had no effect on the choice of the middle brick over the nearest brick. Picture 4 increased the likelihood for the participants to choose the middle brick over the nearest brick (both $p < 0.001$) and decreased the likelihood for the participants to choose the farthest brick over the nearest brick ($p < 0.001$). As for interactions, the interaction between origin of the participants and picture 4 and the origin of the participants and picture 3 proved to be statistically significant. The SE participants were more likely to choose the middle brick ($p < 0.05$) than the NE participants when seeing picture 3 as compared to seeing picture 1. However, when seeing picture 4 as compared to seeing picture 1, the SE participants were less likely to choose the middle brick ($p < 0.05$) than the NE participants. None of the interactions had an effect in predicting the choice for the farthest brick.

The input sentence *võta too klots sealt*—"take that brick from there"

Both independent variables origin of the participants and stimulus picture proved to be statistically significant in predicting the choices of the participants. The SE participants were more likely to choose a farthest brick over the nearest brick ($p < 0.05$) than the NE participants, but there was no effect on the choice of the middle brick over the nearest brick. Pictures 2–6 showed a statistically significant effect on the choice of the participants as compared to picture 1. Picture 2 decreased the likelihood of the participants to choose the middle or the farthest brick over the nearest brick ($p < 0.05$ and $p < 0.001$, respectively). Pictures 3, 5, and 6 increased the likelihood for the participants to choose the farthest brick over the nearest brick ($p < 0.001$, $p < 0.05$, and $p < 0.001$, respectively) but had no effect on the choice of the middle brick. Picture 4 decreased the likelihood for the participants to choose the farthest brick over the nearest brick ($p < 0.01$) and increased the likelihood for the participants to choose the middle brick over the nearest brick ($p < 0.001$). The interaction terms were statistically not significant and thus excluded from the final model.

Visual Salience Condition: Choices of Bricks

To test hypothesis 3, we used the data from the visual salience condition. As the aim of this hypothesis was to pinpoint the differences in the interpretation of specific Estonian demonstratives in regard to the visual salience of the referents, we built a separate model for each input sentence.

TABLE 3 | Results of generalized linear mixed effects regression models predicting the choices of the participants in the visual salience condition.

	***võta see klots siit* "take this brick from here"**	***võta too klots siit* "take that brick from here"**	***võta see klots sealt* "take this brick from there"**	***võta too klots sealt* "take that brick from there"**
Intercept	3.00***	2.38***	0.41	−1.17
Origin SE vs. NE	−1.19	−1.51	−0.32	−1.16
Word order Pron-Adv vs. Adv-Pron	0.21	−0.44	−0.41	−0.29
Picture 8 vs. 7	0.80	0.24	0.18	−0.09
Picture 9 vs. 7	−0.06	−0.63	−2.48***	−2.45**
Origin SE: picture 8 vs. origin NE: picture 7	NA	NA	NA	−0.01
Origin SE: picture 9 vs. origin NE: picture 7	NA	NA	NA	1.94*
Observations	1,662	1,648	1,803	1,806
Model accuracy (%)	92.59	63.30	59.02	72.73

*The reference category of the dependent variable is the salient brick. The accuracy of the models was calculated based on the agreement between the bricks actually chosen and the choices suggested by the probabilities predicted by the model. *p < 0.05, **p < 0.01, ***p < 0.001.*

TABLE 4 | Results of generalized linear mixed effects models predicting the participants' reaction times.

	***võta see klots siit* "take this brick from here"**	***võta too klots siit* "take that brick from here"**	***võta see klots sealt* "take this brick from there"**	***võta too klots sealt* "take that brick from there"**
Intercept	1.39***	1.71***	1.52***	1.62***
Origin SE vs. NE	−0.00	−0.02	0.06	−0.05
Picture 2 vs. 1	0.11	−0.03	0.07	0.02
Picture 3 vs. 1	0.07	−0.13	−0.18***	−0.26***
Picture 4 vs. 1	0.00	−0.28***	0.03	−0.10
Picture 5 vs. 1	0.07	−0.00	−0.10	−0.14**
Picture 6 vs. 1	0.05	−0.05	−0.10	−0.26***
Origin SE: picture 2 vs. origin NE: picture 1	0.00	0.11	NA	NA
Origin SE: picture 3 vs. origin NE: picture 1	−0.07	0.13	NA	NA
Origin SE: picture 4 vs. origin NE: picture 1	−0.15*	0.34***	NA	NA
Origin SE: picture 5 vs. origin NE: picture 1	0.09	0.11	NA	NA
Origin SE: picture 6 vs. origin NE: picture 1	−0.09	0.07	NA	NA
Observations	1,662	1,648	1,803	1,806

**p < 0.05, **p < 0.01, ***p < 0.001.*

None of the variables included in the models built for the input sentences *võta see klots siit*—"take this brick from here" and *võta too klots siit*—"take that brick from here" had a statistically significant effect on the choices of the participants (see **Table 3**). The salient brick, which was also the closest one to the participants from the three bricks, was chosen considerably more frequently than the non-salient bricks regardless of which of the pictures were seen or what was the origin of the participants.

The input sentence *võta see klots sealt*—"take this brick from there"

The only independent variable that was statistically significant in predicting the choice of the salient brick was the stimulus picture. Picture 9 decreased the likelihood for the participants to choose the salient brick as compared to picture 7 ($p < 0.001$).

The input sentence *võta too klots sealt*—"take that brick from there"

There were two statistically significant variables in predicting the choice of the salient brick. First, picture 9 decreased the likelihood for the participants to choose the salient brick as compared to picture 7 ($p < 0.01$). Second, the interaction between origin of the participants and picture 9 increased the likelihood for choosing the salient brick ($p < 0.05$), that is, the SE participants were more likely to choose the salient brick than the NE participants when seeing picture 9 as compared to seeing picture 7.

Results of the Reaction Time Analyses

Since the aim of hypothesis 5 was to test the differences between the two speaker groups while hearing specific Estonian demonstratives, we built a separate model for each input sentence for both distance and visual salience conditions. None of the independent variables proved to be statistically significant in the models of visual salience condition; therefore, we only present the results of the distance condition (see **Table 4**).

The input sentence *võta see klots siit*—"take this brick from here"

The interaction between origin of the participants and picture 4 decreased ($p < 0.05$) the participants' RTs. The SE participants were quicker in making a choice than the NE participants while seeing picture 4 as compared to seeing picture 1. No other

variables or interactions proved to be statistically significant (all *p*-values exceeded the 0.05 threshold).

The input sentence *võta too klots siit*—"take that brick from here"

Picture 4 decreased the RTs significantly as compared to picture 1 ($p < 0.001$) in both participant groups. However, with interaction between origin of the participants and picture 4, the SE participants' RTs got slower than the NE participants' RT's when seeing picture 1 as compared to seeing picture 4. No other variables or interactions proved to be statistically significant (all *p*-values exceeded the 0.05 threshold).

The input sentence *võta see klots sealt*—"take this brick from there"

Only stimulus picture 3 had a statistically significant effect on the participants' RTs. Both participant groups were quicker when they saw picture 3 as compared to picture 1 ($p < 0.01$). This means that making a choice while hearing the input sentences *võta see klots sealt*—"take this brick (from) there" was easier when seeing picture 3 as compared to picture 1. There were no statistically significant interactions between the variables (all *p*-values exceeded the 0.05 threshold).

The input sentence *võta too klots sealt*—"take that brick from there"

Stimulus pictures 3, 5, and 6 had a statistically significant effect on the participants' RTs. Both participant groups were quicker when they saw pictures 3, 5, and 6 as compared to picture 1 ($p < 0.001$, $p < 0.01$, and $p < 0.001$, respectively). This means that making a choice while hearing the input sentences *võta see klots sealt*—"take this brick (from) there" was easier when seeing pictures 3, 5, and 6 as compared to picture 1. There were no statistically significant interactions between the variables (all *p*-values exceeded the 0.05 threshold).

Responses of the Participants: A Choice for a Brick vs. "Do Not Know"

The results show that there was also a slight difference in the overall responses, whether to choose a brick or opt for the "do not know" rectangle, between the two participant groups. **Table 5** presents the results of the response data in the distance and visual salience conditions. Most of the "do not know" responses were for the pictures in which the bricks were all in the same distance, picture 1 and picture 2 (35.2% for NE and 27.9% for SE speakers and 42.9% for NE and 31.9% for SE speakers of the choices, respectively). Thus, for the participants, it was easier to make a choice between the bricks if they were divided between far and near space or if one of the bricks stood out from the rest.

Similarly, the "do not know" answers were frequent in the visual salience condition (see **Table 5**) (note that the total count of choices is 720 for each picture). Differences in the responses between the SE and the NE groups indicated that the participants could also have a different behavior in their decision to choose or not choose a brick. Therefore, we tested this separately for distance and visual salience conditions.

TABLE 5 | The proportion of "do not know" responses by origin of the participants, distance stimuli, and visual salience stimuli.

	Stimulus picture	NE	SE
		n (%)	*n* (%)
Distance	Picture 1	254 (35.2)	201 (27.9)
	Picture 2	309 (42.9)	230 (31.9)
	Picture 3	70 (9.7)	56 (7.7)
	Picture 4	57 (7.9)	57 (7.9)
	Picture 5	180 (25.0)	136 (18.9)
	Picture 6	108 (15.0)	63 (8.7)
Visual salience	Picture 7	285 (39.6)	152 (21.1)
	Picture 8	252 (35.0)	87 (12.1)
	Picture 9	265 (36.8)	115 (15.9)

Distance Condition: A Choice for a Brick vs. "Do Not Know"

As seen in **Table 6**, the origin of the participants (NE vs. SE) did not have a statistically significant effect in predicting the participants' response. However, other independent variables proved to be statistically significant. While pictures 3–6 increased (all $p < 0.001$) the likelihood of the participants to choose a brick rather than clicking on the "do not know" rectangle as compared to picture 1, picture 2 decreased ($p < 0.001$) the same. All input sentences, apart from the input sentence *võta too klots siit*—"take that brick from here," had an effect on the participants' responses. The input sentences *võta see klots sealt*—"take this brick from here" and *võta too klots sealt*—"take that brick from there" increased (all $p < 0.001$) the likelihood for the participants to choose a brick when compared to the input sentence *võta see klots siit*—"take this brick from here".

We also tested for interactions between origin of the participants and stimulus picture (**Table 6**). All interactions proved to be statistically not significant (all $p > 0.05$) apart from the interaction between the variables origin of the participants and picture 6 ($p < 0.05$). The SE participants were more likely to choose a brick instead of the "do not know" rectangle than the NE participants when seeing picture 6 as compared to seeing picture 1.

Visual Salience Condition: A Choice for a Brick vs. "Do Not Know"

As seen in **Table 6**, all independent variables had a statistically significant effect on predicting the participants' response in visual salience condition. The SE participants were more likely to choose a brick rather than the "do not know" rectangle ($p < 0.05$) when compared to the NE participants. Also, hearing the input sentences *võta see klots sealt*—"take this brick from there" and *võta too klots sealt*—"take that brick from there" increased the likelihood ($p < 0.001$ and $p < 0.001$, respectively) for the participants to choose a brick as compared to hearing the input sentence *võta see klots siit*—"take this brick from here." However, hearing the input sentence *võta too klots siit*—"take that brick from here" decreased the likelihood ($p < 0.05$) for the participants to choose a brick as compared to hearing the input sentence *võta see klots siit*—"take this brick from here." As for

TABLE 6 | Results of generalized linear mixed effects models predicting a choice for a brick over "do not know" rectangle clicks in distance and visual salience conditions.

Distance		Visual salience	
Independent variables	**Brick chosen**	**Independent variables**	**Brick chosen**
Intercept	1.14	Intercept	0.67
Origin SE vs. NE	1.40	Origin SE vs. NE	2.01*
Picture 2 vs. 1	−0.59***	Picture 8 vs. 7	0.34*
Picture 3 vs. 1	2.49***	Picture 9 vs. 7	0.21
Picture 4 vs. 1	2.79***	ProxPDistA vs. ProxPProxA	0.60***
Picture 5 vs. 1	0.83***	DistPProxA vs. ProxPProxA	−0.29*
Picture 6 vs. 1	1.80***	DistPDistA vs. ProxPProxA	0.48***
ProxPDistA vs. ProxPProxA	0.77***	Origin SE: picture 8 vs. origin NE: picture 7	0.76**
DistPProxA vs. ProxPProxA	−0.07	Origin SE: picture 9 vs. origin NE: picture 7	0.37
DistPDistA vs. ProxPProxA	0.79***		
Origin SE: picture 2 vs. origin NE: picture 1	0.16		
Origin SE: picture 3 vs. origin NE: picture 1	0.12		
Origin SE: picture 4 vs. origin NE: picture 1	−0.21		
Origin SE: picture 5 vs. origin NE: picture 1	0.17		
Origin SE: picture 6 vs. origin NE: picture 1	0.63*		
Observations	8,640		4,320

*ProxPDistA, võta see klots sealt—"take this brick from there"; DistPDistA, võta too klots sealt—"take that brick from there"; ProxPProxA, võta see klots siit—"take this brick from here"; DistPProxA, võta too klots siit—"take that brick from here." *p < 0.05, **p < 0.01, ***p < 0.001.*

the effect of stimulus pictures, picture 8 increased the likelihood ($p < 0.05$) for the participants to choose a brick rather than a "do not know" rectangle as compared to picture 7. Picture 9, however, had no effect.

We also tested for the effects of interactions. There was one statistically significant interaction between the variables origin of the participants and picture 8 ($p < 0.01$). The SE participants were more likely to choose a brick rather than the "do not know" rectangle than the NE participants when seeing picture 8 as compared to seeing picture 7.

DISCUSSION

While the empirical research on demonstratives in spatial use has increased in recent years (e.g., Stevens and Zhang, 2013; Coventry et al., 2014; Peeters et al., 2014; Gudde et al., 2016; Levinson, 2018; Caldano and Coventry, 2019), there are not many studies that concentrate on demonstrative interpretation. Furthermore, most of the studies focus on demonstrative pronouns rather than demonstrative adverbs in well-studied languages, such as English, Dutch, and Japanese. Considering the fact that there are languages that lack the distance contrast on the level of demonstrative pronouns (Diessel, 1999, 2013), it is essential to investigate both demonstrative pronouns and adverbs to gain full understanding on how demonstrative systems work.

This study focuses on the interpretation of Estonian demonstratives—demonstrative determiners and demonstrative adverbs. We conducted an interpretation experiment where we (1) tested for the effects of referent's distance and visual salience on the interpretation of demonstratives and (2) explored the possible differences in the interpretation of demonstratives between two Estonian native speakers' groups originating from South Estonia (SE, two-term system users) and North Estonia (NE, one-term system users). Studying two speaker groups, who use different demonstrative pronoun systems, provided us with a better insight on how the demonstrative systems work. We were able to do this since the SE speakers tend to use two demonstrative pronouns, *see*—"this" and *too*—"that," while the NE speakers use only *see*—"this." Both speaker groups use all demonstrative adverbs. We will now discuss our findings in the light of each hypothesis.

The results support our first hypothesis, that is, the distance of the referent has an effect on the interpretation of demonstratives: when the demonstrative pronoun *too*—"that" and the adverb *sealt*—"(from) there" are heard, a distant referent is chosen and when the demonstrative pronoun *see*—"this" and the adverb *siit*—"(from) here" are heard, a proximal referent is chosen. This is in line with the previous findings of interpretation (Bonfiglioli et al., 2009) as well as production studies (e.g., Coventry et al., 2008, 2014; Piwek et al., 2008; Tóth et al., 2014; Gudde et al., 2016; Meira and Guirardello-Damian, 2018) in different languages. In our study, the participants tended to choose the farthest referent when hearing input sentences that included the demonstrative pronoun *too*—"that" and the adverb *sealt*—"from there," and the proximal referent tended to be chosen when hearing the demonstrative pronoun *see*—"this" and the adverb *siit*—"from here." At the same time, the effect of distance was relative, that is, the participants still chose a referent according to the input sentence heard even if the demonstrative determiner and/or adverb heard was not congruent with the displayed stimulus picture. For example, with picture 6 where the referents were in near space but the last one was a bit farther than the first two, the likelihood for the participants to choose the farthest referent over the nearest one was even higher than with picture 1, where all the referents were in far space. Similar findings on distance

relativity have been reported for Italian demonstrative pronouns in a comprehension experiment conducted by Bonfiglioli et al. (2009), where the RTs of the participants were slower if the object that was referred to with a proximal demonstrative pronoun was positioned in the far distance and *vice versa*. This indicates that while a single referent in near space (in peripersonal space) can be referred to with proximal demonstrative, as has been found for several languages (see Coventry et al., 2008, 2014 and Reile et al., 2020), with multiple referents in near space, the distance between the referents themselves will start to influence the interpretation of demonstratives and probably production as well.

The results also support our second hypothesis that, with demonstrative pronoun and adverb combinations, the decision of choosing the referent is based on the demonstrative adverbs when the visual scene is incongruent with a demonstrative pronoun in a sentence heard. However, this hypothesis did not hold true with all the pictures. For example, when the participants heard the input sentence *võta too klots siit*—"take that brick from here" and saw picture 2, where all the referents were in near space, the likelihood for the participants to choose the farthest referent decreased as compared to picture 1, where all the referents were in far space. On the other hand, seeing picture 6, where all the referents were in near space but one was a bit farther than the others, the likelihood for choosing the farthest referent increased as compared to picture 1. This different pattern of choices between picture 2 and picture 6 while hearing the same input sentence suggests that the interpretation of demonstrative pronoun and adverb combinations is more complex than we first predicted. While previous findings with Estonian demonstratives show that demonstrative adverbs have a stronger association with the distance of the referent than demonstrative pronouns (Reile, 2019; Reile et al., 2019), our results suggest that this applies only when the intended referents are all in the same distance with each other. Similar findings to our interpretation study have been found by Meira and Terrill (2005) who argued that, in Lavukaleve language, when the referents are positioned in the same length from each other, these are seen as being in one region, and speakers tend to refer to them with the same demonstrative.

While distance had an effect on demonstrative interpretation across the two participant groups, there was a difference in how they interpreted the meaning of demonstrative determiner and adverb combinations. This brings us to our fourth and fifth hypotheses. The fourth hypothesis was as follows: the choices for the referents between the NE and the SE speakers differ when the demonstrative *too* is heard. The SE speakers will show a pattern of choosing the farthest referent, while the NE speakers may choose any referent regardless of their distance. The results partly support this hypothesis. The origin of the participants proved to make a difference with the input sentences where the distal demonstrative pronoun was combined with a proximal demonstrative adverb and *vice versa* and also when the distal demonstrative pronoun was combined with a distal demonstrative adverb. The SE participants were more likely to choose the farthest or the middle referent when hearing a distal pronoun and proximal adverb combination than the NE participants. With the proximal pronoun and the distal adverb combination, they were more likely to choose the middle referent, and with the distal pronoun and distal adverb combination, they were more likely to choose the farthest referent than the NE participants. This suggests that the SE speakers needed to match the region from where to look for the referent with the demonstrative adverb as well as to decide which of the possible referents in this region would best match the demonstrative pronoun heard. This proposition is further supported by findings that imply distance-neutral use (Larjavaara, 2007; Pajusalu, 2009) or spatial unmarkedness (Reile, 2019; Reile et al., 2019) of demonstrative pronoun *see*—"this/that" in the Estonian one-term system (NE), whereas in the Estonian two-term system (SE), demonstrative pronouns are both argued to be spatially anchored (Reile, 2019; Reile et al., 2020). Moreover, a similar tendency to first mark the region with a demonstrative adverb and then the referent with a demonstrative pronoun has been detected in the production of Estonian demonstratives (Reile, 2016) by the SE speakers as well as in the use of Finnish demonstratives by native Finnish speakers. For Finnish, a Finno-Ugric language with three demonstrative stems, Laury (1996) has argued that the locative demonstratives are used for referents that are conceptualized rather as a ground (i.e., a region) than a figure (i.e., the referred object), whereas demonstrative pronouns are used for figure-like referents.

Our additional finding that the SE speakers were more likely to choose a referent rather than opting for a "do not know" answer indicates that having more demonstratives in a demonstrative paradigm can help the speakers to handle ambiguous referential situations. This is in line with Diessel's (1999) proposition that the more demonstratives a language has, the more aspects of the referent they can express. Although the Estonian demonstratives do not express visual salience of the referent *per se*, having an additional demonstrative pronoun seems to aid the speaker to reach a decision, i.e., to choose a referent.

Our fifth hypothesis stating that the RTs for choosing a referent, when the demonstrative *too*—"that" is heard, are slower for the NE speakers than for the SE speakers did not hold true. The origin of the participants proved to be significant only in the interaction with stimulus picture 4, where one of the referents was in near space and two were in far space. The tendency in the RTs seemed to be that the SE speakers were slower in choosing the referent than the NE speakers while hearing the input sentence *võta too klots siit*—"take that brick from here" and seeing picture 4 as compared to seeing picture 1, where the referents were all in far space. The SE speakers were quicker than the NE speakers while hearing the input sentence *võta see klots siit*—"take this brick from here" and seeing picture 4 as compared to seeing picture 1. All differences in the RTs were induced by stimulus pictures. This was especially pronounced when the participants heard the input sentences that included distal demonstratives (both determiner and adverb). The participants were quicker with almost all other pictures than with picture 1 (where all the referents were located in far space), and it was probably harder to make a choice based on this particular input sentence. It is also possible that the difference in the RTs between the two speaker groups is so subtle that the design of the experiment did not capture it. Using a button press instead of a computer mouse to measure the RTs of the participants would have given us more accurate results.

Therefore, the fifth hypothesis could still hold true when using a different measuring technique and is worth further research.

Comparing the two systems in one language already suggests differences in conceptualization of space through language. Adding the evidence that has been reported for other languages suggests that different demonstrative systems can define the way speakers conceptualize space. When more tools are available, the speakers are provided with a more clearly carved up space. For example, demonstrative adverbs are not only used to reinforce the meaning of demonstrative pronouns, as has been suggested by Diessel (2013), but they can also indicate whether the referent is near the addressee or the speaker as in spoken Brazilian Portuguese (Meira and Guirardello-Damian, 2018) (note that demonstrative pronouns are not used for that purpose in spoken Brazilian Portuguese). This further shows the importance of studying demonstrative adverbs in addition to pronouns to better understand the mechanisms of the demonstrative systems.

In addition to the distance of the referent, we tested for the effects of visual salience on the interpretation of demonstratives. Previous research on Estonian demonstratives suggested that the effect of visual salience manifests itself in the word order of the referential utterance rather than in the choice of specific demonstratives (Reile, 2016, 2019), that is, if a referent is not visually salient, then in the word order of the referential utterance the demonstrative adverb precedes the referential noun phrase (that might include a demonstrative pronoun). We tested this in hypothesis 3, and the results show that the visual salience had no impact on which of the referents were chosen. Although this might be true for Estonian demonstratives, research on English demonstrative pronouns has shown that visual access has an effect on demonstrative choice (Coventry et al., 2014). Moreover, there are languages that even have a demonstrative that is specifically used for invisible referents (Diessel, 1999). Therefore, visual salience or access might be less strong of a factor and more language specific than the distance of the referent.

CONCLUSION

To conclude, our study shows that distance has an effect on the interpretation of Estonian demonstratives, and at the same time, the effect of visual salience is overridden by the distance. In addition, the results suggest that there are differences on how the different demonstrative pronoun system users conceptualize space. When using the Estonian with an influence of the SE dialect, the NE speakers rely on demonstrative adverbs in interpreting the referential utterance that includes demonstrative pronoun and adverb combinations, whereas the SE speakers also take into account the semantics of demonstrative pronouns. This shows the importance of studying demonstrative pronouns and adverbs together when tackling the working mechanisms of demonstrative systems.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MR, KA, and NP designed the experiments. MR and KA conducted the experiments. KA analyzed the data. MR interpreted the data. The manuscript text was prepared by MR, NP, and KA. Figures and the tables were prepared by KA. All the authors reviewed the manuscript and approved the final version of the manuscript to be published.

FUNDING

This study was supported by the Estonian Research Council (grants IUT02-13, PRG770, and PUT701) and by the European Union through the European Regional Development Fund (Center of Excellence in Estonian Studies).

ACKNOWLEDGMENTS

We would like to thank Nele Ots (Ph.D.) for her help in processing the auditory input sentences.

REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2007). *PRAAT: Doing Phonetics by Computer (Version 5.2.34)*. Avaliabe at: http://www.fon.hum.uva.nl/praat/ (accessed January, 2016).

Bonfiglioli, C., Finocchiaro, C., Gesierich, B., Rositani, F., and Vescovi, M. (2009). A kinematic approach to the conceptual representations of this and that. *Cognition* 111, 270–274. doi: 10.1016/j.cognition.2009.01.006

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Caldano, M., and Coventry, K. R. (2019). Spatial demonstratives and perceptual space: to reach or not to reach? *Cognition* 191:103989. doi: 10.1016/j.cognition.2019.06.001

Clark, E. V., and Sengul, C. J. (1978). Strategies in the acquisition of deixis. *J. Child Lang.* 5, 457–475. doi: 10.1017/S0305000900002099

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near–far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

di Pellegrino, G., and Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia* 66, 126–133. doi: 10.1016/j.neuropsychologia.2014.11.011

Diessel, H. (1999). *Demonstratives. Form, Function and Grammaticalization*. Amsterdam: John Benjamins Publishing Company, doi: 10.1075/tsl.42

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489. doi: 10.1515/COG.2006.015

Diessel, H. (2013). "Distance contrasts in demonstratives," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Dixon, R. M. W. (2003). Demonstratives: a cross-linguistic typology. *Stud. Lang.* 27, 61–112. doi: 10.1075/sl.27.1.04dix

Gudde, H. B., Coventry, K. R., and Engelhardt, P. E. (2016). Language and memory for object location. *Cognition* 153, 99–107. doi: 10.1016/j.cognition.2016.04.016

Hadeld, J. (2012). *MCMCglmm Course Notes*. Avaliable at: https://www.semanticscholar.org/paper/MCMCglmm-Course-Notes-Hadeld/81c0f45d1b638501bdccc1e1eb7f9a6bcd263180 (accessed August, 2020).

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the {MCMCglmm} {R} Package. *J. Stat. Softw.* 33, 1–22.

Jarbou, S. O. (2010). Accessibility vs. physical proximity: an analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *J. Pragmat.* 42, 3078–3097. doi: 10.1016/j.pragma.2010.04.014

Kleiner, M., Brainard, D., and Pelli, D. (2007). "What's new in Psychtoolbox-3?". *Perception* 36, 1–16.

Küntay, A. C., and Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *J. Child Lang.* 33, 303–320. doi: 10.1017/S0305000906007380

Larjavaara, M. (2007). *Pragmasemantiikka [Pragmasemantics]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Laury, R. (1996). "Pronouns and adverbs, figure and ground: the local case forms and locative forms of Finnish demonstratives in spoken discourse," in *SKY 1996. Yearbook of the Linguistic Association of Finland*, eds T. Haukioja, M.-L. Helasvuo, and E. Kärkkäinen Helsinski, 65–92.

Levinson, S. C. (2018). "Introduction: demonstratives: patterns in diversity," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781108333818.002

Lo, S., and Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6:1171. doi: 10.3389/fpsyg.2015.01171

Meira, S., and Guirardello-Damian, R. (2018). "Brazilian Portugese: non-contrastive exophoric use of demonstratives in the spoken language," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 116–133. doi: 10.1017/9781108333818.006

Meira, S., and Terrill, A. (2005). Contrasting contrastive demonstratives in Tiriyó and Lavukaleve. *Linguistics* 43, 1131–1152. doi: 10.1515/ling.2005.43.6.1131

Nagaraja, K. S. (1985). *Khasi: A Descriptive Analysis*. Dissertation, Deccan College, Pune.

Özyürek, A. (1998). "An analysis of the basic meaning of turkish demonstratives in face-to-face conversational interaction," in *Oralite et Gestualite: Communication Multimodale, interaction*, eds S. Santi, I. Guaitella, C. Cave, and G. Konopczynski (Paris: L'Hartmann), 604–614.

Pajusalu, K., Hennoste, T., Niit, E., Päll, P., and Viikberg, J. (2009). *Eesti Murded ja Kohanimed [Estonian dialects and toponyms]*, 2nd Edn. Tallinn: Eesti Keele Sihtasutus.

Pajusalu, R. (2009). Pronouns and reference in Estonian. *STUF – Sprachtypol. Univ.* 62, 122–139.

Peeters, D., Azar, Z., and Özyürek, A. (2014). "The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, eds P. Bello, M. Guarini, M. McShane, and B. Scassellati (Austin, TX: Cognitive Science Society), 1144–1149.

Peeters, D., Hagoort, P., and Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition* 136, 64–84. doi: 10.1016/j.cognition.2014.10.010

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vis.* 10, 437–442. doi: 10.1163/156856897x00366

Piwek, P., Beun, R.-J., and Cremers, A. (2008). "Proximal" and "distal" in language and cognition: evidence from deictic demonstratives in Dutch. *J. Pragmat.* 40, 694–718. doi: 10.1016/j.pragma.2007.05.001

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Avaliable at: https://www.r-project.org/ (accessed July, 2020).

Reile, M. (2015). Space and demonstratives: an experiment with Estonian exophoric demonstratives. *J. Estonian Finno Ugric Linguist.* 6, 137–165. doi: 10.12697/jeful.2015.6.2.06

Reile, M. (2016). Distance, visual salience and contrast expressed through different demonstrative systems: an experimental study in Estonian. *SKY J. Linguist.* 29, 63–94.

Reile, M. (2019). *Estonian Demonstratives in Exophoric use: An Experimental Approach*. Dissertation, University of Tartu, Tartu.

Reile, M., Plado, H., Gudde, H. B., and Coventry, K. R. (2020). Demonstratives as spatial deictics or something more? Evidence from Common Estonian and Võro. *Folia Linguist.* 54, 167–195. doi: 10.1515/flin-2020-2030

Reile, M., Taremaa, P., Nahkola, T., and Pajusalu, R. (2019). Reference in the borderline of space and discourse: a free production experiment in Estonian. Russian, and Finnish. *Linguist. Uralica* 55, 185–208. doi: 10.3176/lu.2019.3.02

Rocca, R., Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., and Walletin, M. (2020). Language beyond the language system: dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI. *NeuroImage* 216:116128. doi: 10.1016/j.neuroimage.2019.116128

Stevens, J., and Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: an event-related potential study. *J. Neurolinguistics* 26, 31–45. doi: 10.1016/j.jneuroling.2012.02.005

Stevens, J., and Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: a cross-language study of spatial demonstratives. *J. Neurolinguistics* 30, 27–47. doi: 10.1016/j.jneuroling.2014.03.003

Tóth, E., Csatár, P., and Banga, A. (2014). "Exploring hungarian and dutch gestural demonstratives," in *Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014: Language Use and Linguistic Structure*, eds L. Veselovská and M. Janebová (Olomouc: Palacku University), 607–626.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 25 November 2020
doi: 10.3389/fpsyg.2020.555265

Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review

Holger Diessel[1]* *and Kenny R. Coventry*[2]

[1] *Department of English, Friedrich-Schiller-Universität Jena, Jena, Germany,* [2] *School of Psychology, University of East Anglia, Norwich, United Kingdom*

This paper offers a review of research on demonstratives from an interdisciplinary perspective. In particular, we consider the role of demonstratives in current research on language universals, language evolution, language acquisition, multimodal communication, signed language, language and perception, language in interaction, spatial imagery, and discourse processing. Traditionally, demonstratives are analyzed as a particular class of spatial deictics. Yet, a number of recent studies have argued that space is largely irrelevant to deixis and that demonstratives are primarily used for social and interactive purposes. Synthesizing findings in the literature, we conclude that demonstratives are a very special class of linguistic items that are foundational to both spatial and social aspects of language and cognition.

Keywords: demonstrative, deixis, joint attention, peripersonal action space, language acquisition, language universals, spatial cognition, embodied cognition

OPEN ACCESS

Edited by:
Andriy Myachykov, Northumbria University, United Kingdom

Reviewed by:
Bencie Woll, University College London, United Kingdom
Yang Zhang, University of Minnesota Health Twin Cities, United States

***Correspondence:**
Holger Diessel
holger.diessel@uni-jena.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *24 April 2020*
Accepted: *22 October 2020*
Published: *25 November 2020*

Citation:
Diessel H and Coventry KR (2020) Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review. Front. Psychol. 11:555265. doi: 10.3389/fpsyg.2020.555265

INTRODUCTION

The term "demonstrative" refers to a small class of expressions that are commonly divided into two basic types: nominal demonstratives such as English *this* and *that* and adverbial demonstratives such as *here* and *there* (Dixon, 2003). The two types of demonstratives are closely related. They usually include the same deictic roots (Diessel, 1999) and are defined by two basic concepts of language and cognition, i.e., joint attention and deixis (Levinson, 2004; Diessel, 2014).

Joint attention is a key concept of social cognition (e.g., Carpenter et al., 1998; Tomasello et al., 2005). In order to communicate, actor and addressee must coordinate their attention so that they are jointly focused on the same referent. This is not a trivial task because it presupposes that the participants in a communicative act conceive of each other as mental or intentional agents who look at the world from different perspectives. The ability to understand another person's perspective is a basic capacity of the human mind that evolves only gradually in preschool children and is much less developed in other species (Tomasello, 1999). Most research on joint attention has been concerned with nonverbal means of communication, notably with pointing and eye gaze (e.g., Carpenter et al., 1998; Liszkowski et al., 2006); but, of course, joint attention can also be coordinated by linguistic means. In particular, demonstratives serve to create and to manipulate joint attention in face-to-face communication (Clark, 1996; Diessel, 2006).

The term "deixis" is used in different ways by different scholars (see Fricke, 2014 for discussion). Following Bühler (1934), many researchers apply the term to linguistic expressions that are semantically contingent on a particular point of reference, which Bühler called the "origo" (Bühler, 1934: 117). The origo is the center of a coordinate system, i.e., a deictic frame of reference, which, in the unmarked case, is grounded by a speaker's body, but which can be shifted to another person and construed in flexible ways (cf. Diessel, 2014; Stukenbrock, 2015).

Bühler's work has been very influential (cf. Diessel, 2012a; Fricke, 2014), but in some of the recent literature, the term deixis is used in a more general way than proposed by Bühler. According to Levinson (2004), deictic expressions are linguistic elements with "built-in-contextual parameters" that must be specified by the context to be fully understood. While the speaker's body may provide contextual cues for the interpretation of demonstratives in a particular speech situation, Levinson and others have explicitly argued against an egocentric, body-oriented concept of deixis (e.g., Levinson, 2003: 71; Peeters and Özyürek, 2016).

The different views of deixis are key to understanding why there is so much disagreement about the nature of demonstratives in the current literature. As we will see, while some researchers conceive of demonstratives as a particular class of spatial terms that are ultimately based on our bodily experience with concrete objects in space, other researchers argue that demonstratives are primarily used for social and interactive purposes and that space and embodiment are much less important to the study of deixis than commonly assumed.

This paper provides a critical review of current linguistic and psycholinguistic research on demonstratives. We begin with research on demonstratives in linguistic typology, historical linguistics, language acquisition, and signed language, and then turn to the debate about the use of demonstratives in spatial language and social interaction. To preview our conclusion, we will argue that an egocentric, body-centered view of deixis is perfectly compatible with the view that demonstratives are used for both spatial and interactive purposes.

Universality

Demonstratives have a number of important properties that characterize them as a very special class of linguistic expressions (Diessel, 2006). To begin with, demonstratives are likely to be universal. Recent research in typology has argued that languages are much more diverse than commonly assumed in theoretical linguistics and cognitive science. According to Evans and Levinson (2009: 2), "languages differ so fundamentally from one another at every level of description (sound, grammar, lexicon, meaning) that it is very hard to find any single structural property they share."

Yet, while language universals are rare and difficult to find, they DO exist. One aspect all languages seem to share is a particular class of demonstratives. Although Evans and Levinson do not mention demonstratives in their programmatic paper on the "myth of language universals," there have been several large-scale typological studies suggesting that demonstratives are very likely to exist in all languages (e.g., Himmelmann, 1997; Diessel, 1999; Dixon, 2003; Breunesse, 2019; see also Levinson, 2018). The universality of demonstratives stands in sharp contrast to the cross-linguistic distribution of other closed-class items. As Evans and Levinson (and others) have noted, many languages lack adpositions, determiners, auxiliaries, conjunctions, case markers, copulas and third person pronouns. Yet, demonstratives seem to be universal.

Note that this does not concern the word class functions of demonstratives. Above we have mentioned the distinction between nominal demonstratives (e.g., *this/that*) and adverbial demonstratives (e.g., *here/there*), which concerns the analysis of demonstrative word classes. Like English, many other languages distinguish between nominal demonstratives functioning as pronouns or determiners and adverbial demonstratives functioning as spatial adverbs (Dixon, 2003). Yet, while this distinction is common, it is NOT universal. Acehnese, for instance, has three deictic particles, *nyoe, nyan* an *jêh*, glossed by (1985: 130) as "this," "that, close" and "that, far," respectively, that can be used as pronouns (cf. 1a) or spatial adverbs (cf. 1b). However, there are no language internal criteria that would justify a categorical division between nominal and adverbial demonstratives in Acehnese (Durie, 1985: 130–4).

(1) Acehnese (Austronesian, Indonesia)

a. *bek* *neu = peugah* ***nyan*** *bak = lôn*
don't 2 = tell that to = I
"Don't tell me that." (Durie, 1985: 49)
b. ***nyoe*** *na* *peng*
here be money
"Here is my money." (Durie, 1985: 132).

Moreover, the word class categories of demonstratives do not only comprise pronouns, determiners and spatial adverbs. If we look at demonstratives from a cross-linguistic perspective, we also find manner demonstrative adverbs (König, 2012), demonstrative identifiers (Diessel, 1999), demonstrative presentatives (Treis, 2018) and demonstrative verbs (Guérin, 2015). Mauwake, for example, has demonstrative verbs that occur with tense and verbal agreement affixes, as in example (2). Demonstrative verbs are rare, but have also been found in various other languages including Dyirbal (Dixon, 2003), Mapuche (Smeets, 1989), Komnzo (Döhler, 2018), Yukaghir (Maslova, 2003) and Quechua (Shimelman, 2017) (see Guérin, 2015 for a cross-linguistic overview).

(2) Mauwake (Trans New Guinea, Papua New Guinea)

nomokowa *unowa* ***fan-e-mik***, . . .
2SG/PL.brother many here-PST-1/3PL . . .
"Many of your brothers are here, ..." (Berghäll, 2015: 266).

In general, some languages use a single series of demonstratives across a wide range of contexts, but other languages have elaborate systems of demonstrative word classes (cf. **Table 1**). Yet, while the word class categories of demonstratives are language- and construction-particular, typologists agree that all languages have a special class of demonstratives.

What is more, not only the existence of demonstratives is likely to be universal, but also the distinction between proximal and distal terms may be a universal property of language (Diessel, 1999; Dixon, 2003; Breunesse, 2019). Some languages have "neutral demonstratives" that are not deictically contrastive. The French demonstrative *ça*, for instance, is distance-neutral. Recent research suggests that neutral demonstratives are cross-linguistically more common than previously assumed in the

TABLE 1 | Examples of demonstrative word class systems.

		DET	PRO	ADV.SPACE	ADV.MANNER	IDENTIFIER	VERB
French	PROX	*celui/celle-(ci)*	*ce/cette-(ci)*	*ici*	*ainsi*	*ce*	*(voici)*
	DIST	*celui/celle-(là)*	*ce/cette-(là)*	*lá*			*(voilà)*
Mauwake	PROX	*fain*		*fan*	*feenap*		*fan*-PST-AGR
	DIST	*nain*		*nan*	*naap*		*nan*-PST-AGR
Acehnese	PROX		*nyoe*				
	MED		*nyan*				
	DIST		*jêh*				

typological literature (Levinson, 2018). Nevertheless, while demonstratives are not generally marked for distance, the available data suggest that all languages have at least two distance-marked demonstratives that correspond to English *here* and *there*. **Figure 1** shows the number of distance contrasts in spatial demonstrative adverbs in a representative sample of 150 languages (the language sample is described in **Supplementary Datasheet 1** in the **Supplementary Materials**).

As can be seen, the majority of languages in this sample have two or three distance terms (two-term: $N = 72$; three-term: $N = 66$). Larger systems with four or more terms are rare ($N = 12$); but note that some languages have spatial demonstrative adverbs indicating elevation, direction or visibility (not shown in **Figure 1**), in addition to distance (cf. Forker, 2019, this volume).

Language Evolution

Another aspect that characterizes demonstratives as a special class is their role in language change and language evolution. Both linguists and cognitive scientists have often argued that language has evolved from gesture (e.g., Arbib, 2012; Liszkowski et al., 2012). The hypothesis is intriguing, but difficult to evaluate. Since there are no historical records of early human communication, it is impossible to study the evolution of gesture and speech directly. Nevertheless, there is good evidence from diachronic linguistics that demonstratives, which are commonly accompanied by deictic gestures (see below), have emerged early in language evolution (Diessel, 2013).

In the historical literature, it is often assumed that all grammatical function morphemes are ultimately based on content words, notably on nouns and verbs (Bybee, 2003; Hopper and Traugott, 2003). Yet, several studies have pointed out that although demonstratives are closed-class items, they are not etymologically related to nouns and verbs (Himmelmann, 1997: 20; Dixon, 2003). In particular, Diessel (1999, 2006, 2013, 2014) has argued that the diachronic origins of demonstratives are unknown. Considering data from several hundred languages, Diessel did not find a single language in which demonstratives are derived from content words, suggesting that demonstratives are fundamentally distinct from other closed-class items.

Heine and Kuteva (2007) have challenged this claim, arguing that demonstratives have evolved from motion verbs (see also Frajzyngier, 1996: 159; Heine and Kuteva, 2002: 146). The main piece of evidence for this hypothesis comes from a few African languages, in particular from !Xun, in which a verb meaning "go" is phonetically similar to a distal demonstrative. There are no historical records to investigate the proposed development in these languages. However, since motion entails distance, Heine and Kuteva (2007: 76–7) maintain that their analysis is not only suggested by the phonetic overlap between the verb "go" and the distal demonstrative "that," but also by semantic factors.

We are not convinced by this analysis. To the best of our knowledge, there is no widespread phonetic similarity between demonstratives and motion verbs and the conceptual link between motion and distance is not sufficient to postulate a general grammaticalization path from "go" to "that." Moreover, even if it turns out that demonstratives and motion verbs are diachronically related in some languages, the direction of the relationship could be the other way around. Heine and Kuteva assume that motion implies distance, but it is equally plausible that the indication of a distant referent is interpreted as a request for movement.

As it stands, we are not aware of any language for which we can be certain that demonstratives have evolved from motion verbs. What we do find in some languages are demonstratives that have coalesced with verbs (Brugmann, 1904; Evans, 1990; Vindenes, 2017). In French, for example, the deictic presentatives *voici* and *voilà* are historically derived from the singular imperative form of the verb *voir* "see/look" and the spatial demonstratives *ici* "here" and *là* "there." In other languages, demonstratives have merged with copulas (e.g., in Komnzo; see Döhler, 2018: 126-7) or posture verbs (e.g., in Gunwingguan; see Evans, 1990). However, while these developments explain why the demonstratives of some languages include a verb root, or why they are inflected for tense, they do not explain where demonstratives come from.

In general, although there is no apriori reason to exclude the possibility that a demonstrative may evolve from a motion verb, the available data suggest that, if this has ever happened, it is a rare phenomenon that does not explain the diachronic origins of demonstratives as a cross-linguistic class (Diessel, 2006, 2013).

What is more, demonstratives are not only old and non-derived, they also play a key role in the diachronic evolution of grammar. Research on grammaticalization has been mainly concerned with the development of function morphemes from nouns and verbs. There is plenty of evidence that adpositions, auxiliaries, case markers and many other types of grammatical morphemes have evolved from content words. Yet, what is often overlooked, or not sufficiently explained in the grammaticalization literature, is that demonstratives provide a second major source for grammatical function morphemes (Himmelmann, 1997: 115–155, Diessel, 1999, 2019: 167–171;

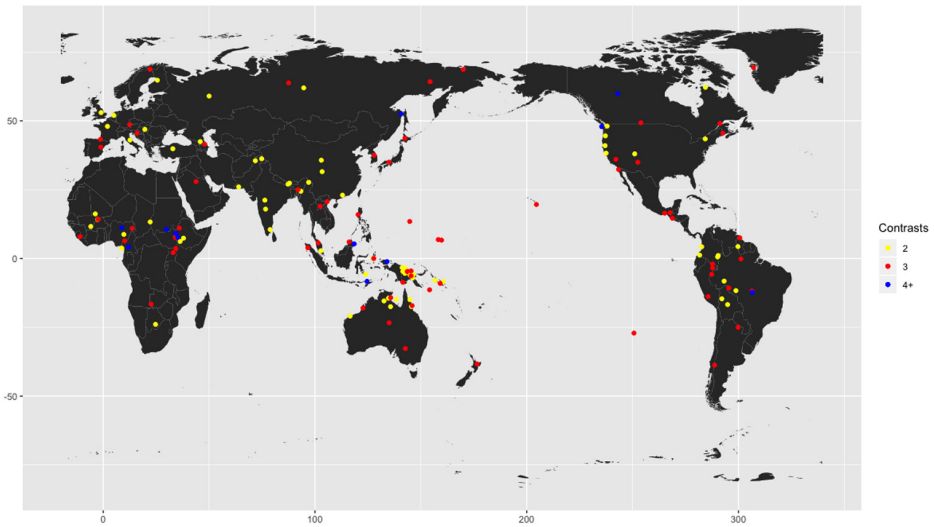


FIGURE 1 | Number of distance contrasts in spatial demonstrative adverbs in a 150 language sample.

Diessel and Breunesse, 2020). Across languages, demonstratives are commonly reanalyzed as definite articles, third person pronouns, relative pronouns, quotative markers and nonverbal copulas, which in turn often develop into noun class markers, agreement affixes, subordinate conjunctions, complementizers and focus markers (see **Figure 2**). While some of these morphemes may also arise from content words, there can be no doubt that demonstratives are of fundamental significance to the diachronic evolution of grammar, as already suggested in the classic works of Brugmann (1904) and Bühler (1934).

Language Acquisition

There is little research on the acquisition of demonstratives, but in a classic paper, Clark (1978) made three important claims. First, she argued that the acquisition of verbal deixis builds on children's prior use of deictic gesture; second, she claimed that demonstratives are usually among children's first words; and third, she argued that demonstratives are very frequent in early child language:

"Among the earliest words acquired is usually at least one deictic word—a word invariably used together with a deictic gesture. A deictic word based on *there* or *that* . . . often appears in the first ten words of English-speaking children, certainly within the first 50 words" (Clark, 1978: 95).

These claims are widely cited in the literature, but to date only few studies have examined the acquisition of demonstratives in development (empirically). Several studies show that the development of demonstratives seems to be quite protracted, with adult-like uses emerging long after children start producing demonstratives (de Villiers and Villiers, 1974; Webb and Abrahamson, 1976; Clark and Sengul, 1978; Tanz, 1980; Küntay and Özyürek, 2006). However, the only study we know that is specifically concerned with the relationship between pointing and demonstratives in young children is a recent paper by Todisco et al. (2020).

Using data from reading sessions in which participants referred to animals in a picture book, Todisco et al. (2020) analyzed the interaction between verbal and nonverbal means of deictic reference in Italian-speaking children aged 20 to 31 months. While these children are (already) too old to examine the transition from gesture to speech, Todisco et al. observed that young children frequently combine demonstratives and pointing gesture, and do so in a synchronous manner, with the peak of the pointing gesture produced at the same time as the deictic vocalization. Moreover, joint attention on an object (both infant and caregiver looking at the intended referent) was found to immediately precede deictic communication in the vast majority of deictic episodes.

The main piece of evidence for Clark's hypothesis that demonstratives are among children's first words comes from diary and observational research (Nelson, 1973; Bates, 1976). Specifically, these studies report that children's early pointing gestures are frequently accompanied by vocalizations such as [e], [aʔ], or [da] that may be seen as precursors of demonstratives (Clark, 1978: 95). If this is correct, demonstratives are usually among the earliest words children produce.

Caselli et al. (1995) presented data that raised doubts about this claim. Analyzing parent reports of young English and Italian children, compiled with the MacArthur-Bates Communicative Development Inventory (CDI), these researchers did not find any demonstratives among children's first 50 words. However, since parent reports may not provide a reliable measure for the appearance of closed-class function words (Salerni et al., 2007), we decided to look at children's early demonstratives in naturally occurring child speech.

Using data from the CHILDES database (MacWhinney, 2000), we investigated records of spontaneous speech from 20 children, learning four different languages: English (N = 10), Dutch (N = 3), Hebrew (N = 4), and Japanese (N = 3). We selected these children based on two criteria: age and the amount of data

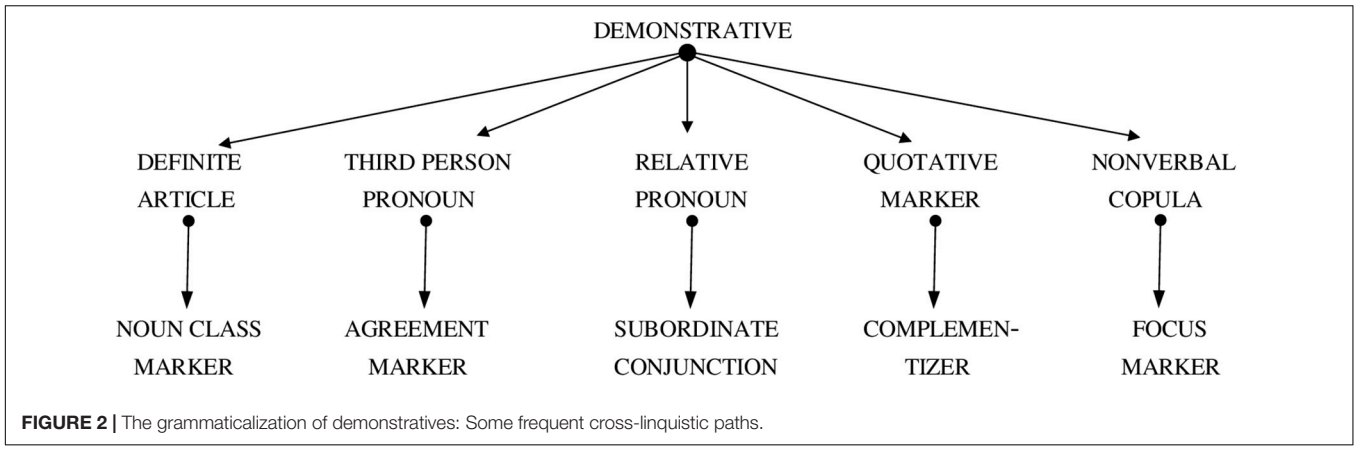


FIGURE 2 | The grammaticalization of demonstratives: Some frequent cross-linquistic paths.

available for each child. All 20 children were younger than 25 months and their data include a minimum of 3400 words per child. The results of this study are summarized in **Supplementary Datasheet 2** of the **Supplementary Materials**.

Overall, the data comprise 206,188 child words. Since most of these words were produced by children beyond the one-word stage, the data are not fully appropriate to examine the appearance of children's very first words. Nevertheless, while one would need other types of data to determine the precise age when children begin to use demonstratives, the data strongly suggest that demonstratives are generally among children's early words. As it turns out, there was at least one demonstrative in the first file of all 20 children regardless of their age. Even the youngest children, aged 10 to 17 months (Laura, Naomi, Judith, Peter, Meinder, Smadar, Lior, Nanami, Asoto, Kiichan), used demonstratives from early on.

The vast majority of children's early demonstratives occur in one-word utterances, or less frequently, together with a noun. They are embedded in parent-child interactions in which the participants seek to focus the other person's attention onto a referent in the surrounding situation. Here are some typical examples from the English data:

(3) Naomi (14 months)
*CHI: Kit(ty) kit(ty).
*MOT: Okay, are you done with looking at the pictures?
*MOT: Are you going to give them to me?
*CHI: **Dere** [= there].
*MOT: Okay you can give them to me.

(4) Laura (18 months)
*CHI: Matthoo [= Matthew].
*MOT: And Matthew.
*MOT: Where are they?
*CHI: **There**.
*MOT: There?

(5) Eve (18 months)
*MOT: I don't think so.
*MOT: Mr. Fraser has coffee.
*MOT: Mr. Fraser's drinking coffee.
*CHI: **That**.
*MOT: What is that?

(6) Eve (18 months)
*MOT: The ducks say what?
*CHI: **That**.
*MOT: What is that?
*CHI: **That** radio.
*MOT: What?

(7) Eve (18 months)
*ADL: That's very good.
*CHI: I did it.
*CHI: **There**.
*CHI: **There** Fraser.
*ADL: That's a nice box of books.

While the CHILDES transcripts do not provide (systematic) information about the context and use of gesture, it is clear from the data that children's early demonstratives refer to objects and locations in their vicinity and that many of these early uses are accompanied by gesture (as indicated on the "action tier"), consistent with the findings of Todisco et al. (2020). The data also show that demonstratives are very frequent in early child language. As can be seen in **Table 2**, demonstratives account for a very large proportion of children's early words, ranging from a mean of 5.9% in Dutch to a mean of 8.3% in English.

Moreover, if we look at the frequencies of individual words, we find a demonstrative at the top of the word frequency lists of 8 of the 20 children. Apart from demonstratives, children made extensive use of pronouns (e.g., *it*), determiners (e.g., *the*) and interjections (e.g., *oh, yeah, no*); but with the exception of "mummy" and some proper names, there were hardly any nouns (or verbs) among the 20 most frequent words at this age, suggesting that demonstratives are the preferred means of linguistic reference in early child language.

Comparing children across the four languages, we found a conspicuous asymmetry in the use of proximal and distal terms. The English- and Dutch-speaking children used more distal demonstratives than proximal demonstratives (English: 26.2% proximal vs. 73.8% distal; Dutch: 30.3% proximal vs. 69.7% distal), but the Hebrew- and Japanese-speaking children used primarily proximal terms (Hebrew: 97.1% proximal; Japanese:

TABLE 2 | Raw frequencies and mean proportions of demonstratives in early child speech.

	Number of children	Age range	Corpus size (child)	DEM total	Mean proportions
English	10	1.02–2.0	103329	8478	8.27%
Dutch	3	0.10–2.0	20991	869	5.88%
Hebrew	4	1.04–2.0	34852	3101	7.76%
Japanese	3	1.00–2.0	47016	3277	7.82%
Total	20		206188	15725	

90.1% proximal). There were also some medial demonstratives in the Japanese data (3.9% medial), but distal demonstratives were rare in both Hebrew (2.8% distal) and Japanese (5.2% distal). Since children's mothers used very similar proportions of proximal and distal terms, it seems reasonable to assume that children tend to use demonstratives that are frequent in the ambient language.

More research is needed to investigate the acquisition of demonstratives and the alternation between proximal and distal terms in early child language. However, while our data are not sufficient to verify Clark's claim that demonstratives are always among children's first 50 words, they strongly suggest that children begin to produce demonstratives early and that demonstratives are among the most frequent words in early child language.

Multimodality

One of the most conspicuous properties of demonstratives is that they are frequently accompanied by nonverbal means of deictic reference, notably by pointing and eye gaze (Enfield, 2003; Stukenbrock, 2015; Levinson, 2018). The multimodal use of demonstratives has been investigated from different perspectives with a variety of methods.

First, linguistic field workers have developed particular elicitation tools to examine the interaction between demonstratives, pointing and gaze in different contexts. Of particular importance is the questionnaire developed by Wilkins (2018), which has been used in a large number of studies on languages across the world (see the recent collection of articles in Levinson et al., 2018). While the Wilkins questionnaire is not specifically designed to probe into multimodal communication, this research strongly suggests that the combination of demonstratives with pointing and gaze is cross-linguistically very common.

However, while multimodality may be a universal trait of demonstrative reference, there are interesting differences in the way demonstratives are combined with nonverbal strategies of deixis. For instance, while it is by no means uncommon for speakers of English to use (exophoric) demonstratives without a co-occurring gesture, reports of linguistic field workers suggest that there are languages in which certain types of demonstratives are generally accompanied by pointing or gaze. In Yélî Dnye, for example, proximal demonstratives seem to require a pointing gesture, or at least gaze, unless the referent is being manipulated (Levinson, 2018: 32). Other languages in which pointing or gaze appear to be "obligatory" with certain demonstratives include Goemai (Hellwig, 2003: 263), Kilivila (Senft, 2004: 62), Yucatec (Bohnemeyer, 2018), Warao (Herrmann, 2018), and Tiriyó (Meira, 2018). Interestingly, in many of these languages it is the proximal demonstrative that is tied to gesture (Levinson, 2018: 32-3).

Second, the multimodal use of demonstratives has been examined with methods of conversational analysis (e.g., Laury, 1997; Hindmarsh and Heath, 2000; Strauss, 2002; Enfield, 2003; Eriksson, 2008; Etelämäki, 2009; Jarbou, 2010; Stukenbrock, 2015; Gipper, 2017). Using video recordings of naturally occurring speech, these studies provide in-depth analyses of multimodal demonstratives in different contexts. One aspect that is emphasized in this research is that demonstratives are primarily used for interactive purposes rather than for spatial reference (see below). Another finding is that demonstratives are not only combined with prototypical pointing gestures, involving the extended arm and index finger, but also with various forms of "bodily displays" including touching, reaching, holding and picking up (Eriksson, 2008; see also Hindmarsh and Heath, 2000; Enfield, 2003; Talmy, 2018).

Finally, there are a number of psycholinguistic studies that have examined the multimodal use of demonstratives with experimental methods (e.g., Bangerter, 2004; Piwek et al., 2008; Lücking et al., 2015; Cooperrider, 2016; Garciá et al., 2017). Most of this research is concerned with demonstratives and pointing, but Garciá et al. (2017) looked at the interaction between demonstratives and gaze (see also Todisco et al., 2020 and Stukenbrock, this volume). Using an experiment in which participants had to instruct another person to move an object on a tablet, they tested the effect of eye gaze on spatial language under two conditions: the gaze condition, in which participants could see each other's eyes, and the no-gaze condition in which their eyes were hidden behind goggles. As expected, in the gaze condition, participants made extensive use of demonstratives, but in the no-gaze condition, they resorted to other, non-deictic means of spatial language, suggesting that speakers shun away from verbal deixis when gaze is not available as a communicative device.

In accordance with this finding, Bangerter (2004) observed that the availability of gesture has a significant impact on speakers' use of demonstratives. When gesture is available, speakers prefer short deictic descriptions; but when gesture is not available, they tend to use longer nondeictic descriptions. In addition, Bangerter found that the combination of demonstrative and gesture varies with distance. Other things being equal, the gestural use of demonstratives is much more frequent for nearby referents than for referents far away (see also Piwek et al., 2008). Since far-away referents are often difficult

to identify by gesture (Lücking et al., 2015), one might hypothesize that the correlation between distance and pointing is ultimately motivated by the ambiguity of distant pointing. Good evidence for this hypothesis comes from a study by Cooperrider (2016), who found that speakers extend the gestural use of demonstratives to far-away referents if they are given a laser pointer, making it possible to identify a distant referent that cannot be unambiguously identified by manual gesture.

Signed Language

Signed language abounds with pointing gestures, but the status of pointing signs is controversial (e.g., Liddell, 2000; Cormier et al., 2013; Johnston, 2013; Fenlon et al., 2019). Like demonstratives, pointing signs can target either a perceptually accessible referent in the surrounding situation or a discourse referent in signed space. In the latter case, an absent referent is located on a horizontal plane in front of the signer, where it is available as a "locus" for subsequent reference through pointing.

Following Friedman (1975) and others, it has long been assumed that pointing gestures have the status of lexical signs in signed language. They are commonly analyzed as particular types of words functioning as pronouns, determiners, adverbs and other word classes (Friedman, 1975; Klima and Bellugi, 1979). However, some of the more recent literature has questioned this view, arguing that pointing signs are distinct from lexical signs (Liddell, 2000; Cormier et al., 2013; Johnston, 2013).

All researchers agree that pointing is of central significance to reference in signed language, but given that pointing is also commonly used as co-speech of spoken language, it is not immediately clear why pointing signs should be regarded as words rather than as genuine gestures (Engberg-Pedersen, 2003). There is an ongoing debate about this issue in the current literature.

Some researchers claim that pointing signs are fundamentally distinct from pronouns or determiners (Liddell, 2000) and emphasize that most of the features that characterize pointing signs in signed language are also characteristic of deictic points in co-speech (Liddell, 2000; see also Engberg-Pedersen, 2003; Johnston, 2013). Other researchers argue that, while pointing signs are (superficially) similar to pointing in spoken language, they are more constrained in meaning and form than ordinary pointing gestures. For instance, in one study, Fenlon et al. (2019) compared video data of 24 signers of British signed language to video data of 27 speakers of American English and found that, on balance, signers' pointing signs were more reduced, more consistent, and more integrated with other aspects of the linguistic system than the pointing gestures of speakers' co-speech.

As it stands, the issue is unresolved (see Cormier et al., 2013 for discussion); but irrespective of the view a particular researcher holds in this debate, there is widespread consensus that pointing signs have a particular status in signed language: they are "semi-conventionalized" (Johnston, 2013) and combine aspects of "word and gesture within a single form" (Meier and Lillo-Martin, 2010: 356; see also Cormier et al., 2013).

What is more, some recent studies explain the particular status of pointing signs in signed language by grammaticalization (Pfau and Steinbach, 2011; Fenlon et al., 2019). More precisely, Pfau and Steinbach (2011) hypothesized that (many) pointing signs can be seen as grammaticalized pointing gestures that have evolved along a grammaticalization path leading from genuine pointing gestures via locative pointing signs to determiners, personal pronouns and agreement markers. Since there are almost no diachronic data of signed languages, the hypothesis is difficult to verify (Pfau and Steinbach, 2011: 384). However, interestingly, Coppola and Senghas (2010) present data from Nicaraguan sign language that could be interpreted as evidence for the proposed grammaticalization path.

Nicaraguan sign language emerged as a new language in the late 1970s when deaf children were brought together for the first time at school. Earlier, deaf people had very little contact with each other and signed only at home to communicate with hearing people around them. Comparing pointing signs of four "homesigners" with pointing signs of different cohorts of signers who used Nicaraguan sign language at school (starting at different stages of the emerging language), Coppola and Senghas found that homesigners' pointing signs were almost exclusively used to indicate a place, whereas the pointing signs of the three cohorts who used Nicaraguan sign language at school also included abstract points functioning as determiners and personal pronouns which seem to have evolved from locative points by grammaticalization.

Finally, while pointing signs share many properties with demonstratives in spoken language, it is unclear if the deictic points of signed language are marked for distance, like most demonstratives in spoken language, or if they are distance-neutral, like the deictic points of co-speech. The only study we know that has been explicitly concerned with the expression of distance in signed language is Morford et al. (2019). Using an experimental paradigm in which bilingual signers of American sign language had to coordinate their actions in a cooperative task, these researchers observed that points to distal referents were often accompanied by "facial compressions" such as eye squinting, head tilt and cheek raising, which only rarely appeared with points to proximal referents. However, the same facial compressions also occurred when the addressee had misunderstood a previous referent. In addition, Morford et al. observed that pointing signs of distal referents were more often used with a straight index handshape and an arc trajectory than pointing signs of proximal referents, but this was not statistically significant. Only future research can show if the distinction between proximal and distal deictics also occurs in signed language or if the pointing signs of signed language are distance-neutral (possibly because distance is an emergent property of deictic pointing that has not yet been grammaticalized in signed languages, which tend to be much younger than spoken languages).

Perceptual Space

We now turn to the above-mentioned debate about the nature of deixis and demonstrative reference. Recall that there are two different views of deixis. Some researchers conceive of

(spatial) deixis as an egocentric, body-oriented strategy to provide orientation in space (e.g., Bühler, 1934; Coventry et al., 2008; Diessel, 2014); but other researchers dispute the pivotal role of speakers' body for the study of deixis and argue that demonstratives are primarily used for social and interactive purposes rather than for spatial reference (e.g., Jarbou, 2010; Peeters and Özyürek, 2016; Gipper, 2017).

In what follows, we discuss studies from both sides of the debate. We begin with research supporting the egocentric, body-oriented view of spatial deixis and then turn to research that has emphasized the social and interactive functions of demonstratives.

At the heart of the current debate about demonstrative reference is the alternation between proximal and distal terms. Traditionally, this alternation is explained by the relative distance between referent and origo. Crucially, "relative distance" must not be confused with "physical distance." As Bühler (and many others) have pointed out, speakers' choice between proximal and distal demonstratives is contingent on language users' conceptualization of the speech situation rather than on physical properties of space. Consider, for instance, the following examples of the proximal demonstrative *here* (cf. Diessel, 2012b: 2410).

(3) a. Here on my leg
b. Here in this room
c. Here in London
d. Here in Europe
c. Here on this planet.

What these examples show is that the region included in the origo varies with the construal of the speech situation. In example (3a), *here* refers to a small spot on speaker's leg, but in all other examples, it refers to a much larger region, which generally includes the speaker but may also include the addressee. The distal term *there* is used in contrast to *here*. It can refer to any location as long as the location is not included in the region conceptualized as the origo. Thus, while *here* and *there* refer to locations in different distance to the speaker, the distance features of these terms are determined by interlocutors' conceptualization of the speech situation rather than by objective properties of metrical space (e.g., Bühler, 1934; Enfield, 2003).

However, while the alternation between proximal and distal demonstratives cannot be defined in terms of physical space, a number of recent studies have argued that the encoding of distance in demonstratives is ultimately based on our bodily interaction with concrete objects in (real) space. These studies draw on research in neuropsychology indicating that objects in peripersonal space are processed in fundamentally different ways from objects in extrapersonal space (e.g., Goodale and Milner, 1992). Peripersonal space is the region of space in which a person can interact with objects and animate beings, by reaching, grasping or touching. Extrapersonal space, in contrast, is the region of space in which objects and animate beings are only perceptually accessible but not available for (physical) interaction (see Bufacchi and Iannetti, 2018 for a review).

Considering this distinction, Coventry and colleagues hypothesized that the universal contrast between proximal and distal demonstratives has its roots in the neuropsychological distinction between peripersonal and extrapersonal space (Coventry et al., 2008, 2014; Gudde et al., 2016; see also Rocca et al., 2019b). This hypothesis was first proposed by Kemmerer (1999), who then dismissed it, mainly because proximal demonstratives are not only used for reachable objects.

However, while the alternation between proximal and distal demonstratives is very flexible, there is good reason to assume that demonstrative choice is ultimately grounded in the vision and action systems, i.e., in the distinction between peripersonal and extrapersonal space. In early research on this topic, peripersonal space was primarily defined in terms of the actor's body, notably the actor's perimeter of arm's reach (Kemmerer, 1999); but more recent research suggests that the distinction between peripersonal and extrapersonal space is mainly determined by the way a person interacts with objects rather than by concrete body parts (see Bufacchi and Iannetti, 2018 for a recent discussion of relevant findings).

Building on these considerations, Coventry et al. (2008) conducted a series of experiments with speakers of English and Spanish in order to investigate the potential influence of peripersonal space on speakers' choice of a particular demonstrative. Using a new experimental paradigm in which participants could choose between proximal and distal terms in order to refer to objects at different distances from the speaker (**Figure 3** left panel), they found a strong preference for proximal demonstratives if speakers could reach the referent. Crucially, while *this* was usually confined to objects in arm's reach, Coventry et al. showed that speakers extend the use of proximal demonstratives to referents at a greater distance if they can use a tool, e.g., a stick, in order to reach it (**Figure 3** right panel), indicating that it is not speakers' body *per se* but the (in)ability to interact with an object that affects their choice of a particular deictic term.

The results were replicated in several follow-up studies with speakers of other languages under somewhat different experimental conditions (cf. Coventry et al., 2014; Gudde et al., 2016; Caldano and Coventry, 2019; Reile et al., 2020). Other evidence supporting a distance-based analysis of demonstrative reference comes from an EEG experiment by Stevens and Zhang (2013) and a behavioral study by Bonfiglioli et al. (2009).

Taken together, this research provides compelling evidence for an egocentric, body-oriented view of spatial deixis in which speakers' choice of a demonstrative is (often) motivated by the possibility of interaction between referent and origo, which ordinarily correlates with distance. However, in addition to relative distance, there are various other factors that can influence demonstrative choice (cf. Coventry et al., 2014).

Interactional Space

If one were to adopt a purely distance-based account, one important aspect "missing" is the role of the hearer (Jungbluth, 2003). Accounts of the demonstrative systems with three or more terms often consider whether such systems are "distance-based" (e.g., with a medial distance term) or whether they

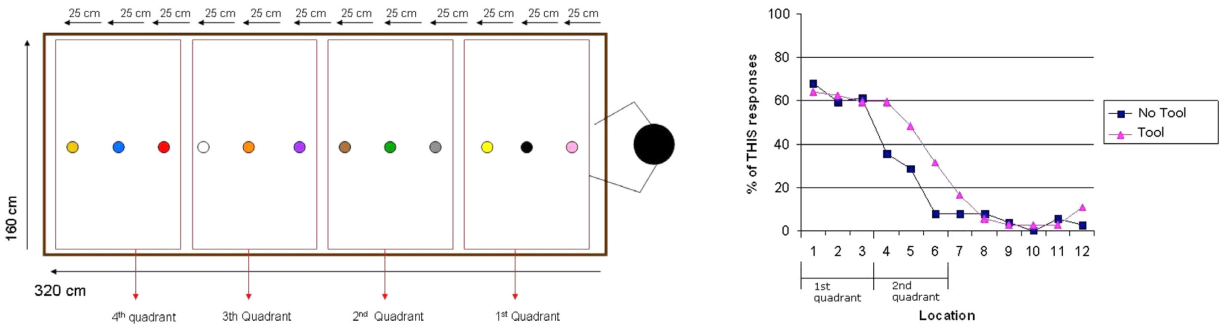


FIGURE 3 | Experimental set-up used by Coventry et al. (2008) (left panel) and associated results (right panel).

might be "person-centered." For example, Spanish has three demonstratives, *este, ese* and *aquel*, that are often described in grammars as being parallel to the distinction between first, second and third person. On this view, *este* refers to an object near the speaker, *ese* indicates a referent near the hearer, and *aquel* specifies a referent far away from both speech participants (Anderson and Keenan, 1985). In contrast, Anderson and Keenan (1985) consider the Japanese three-term demonstrative system a distance-based system, with the terms *kore*, *sore* and *are* representing increasing distance from the speaker (but see Hasegawa, 2012).

The distinction between person-oriented and distance-oriented systems has been prominent in cross-linguistic research on demonstratives (Anderson and Keenan, 1985; Diessel, 1999; Dixon, 2003). Yet, recent research suggests that this distinction is note quite appropriate to characterize demonstratives in three- and four-term systems (cf. Levinson, 2018). One reason for this is that the position of the hearer can influence the conceptualization of space in several different ways. As Jungbluth (2005) has demonstrated, based on data from Spanish, the influence of hearers' position on demonstrative choice varies with the constellation of speaker and addressee in a particular situation (see also Jungbluth, 2003).

Using an elicitation task, Jungbluth examined the use of *este, ese* and *aquel* in three basic constellations: face-to-face, face-to-back and side-by-side. In face-to-face conversation, every object included in interlocutors' shared field of vision was referred to by *este*, even if the referent was close to the addressee; yet, referents outside of interlocutors' shared space were referred to by *aquel* (**Figure 4** left panel). In face-to-back conversation, *ese* referred to objects in hearers' immediate field of vision, whereas *este* was preferred for referents near the speaker, which the hearer could not see (**Figure 4** middle panel). Finally, in side-by-side conversation, *este, ese* and *aquel* were used to differentiate between three different referents on a relative distance scale (**Figure 4** right panel).

Thus, while there are situations in which *ese* indicates a referent near the hearer, Jungbluth maintained that Spanish does not have a simple hearer-oriented system as commonly assumed in the typological literature. Rather, the use of all three Spanish demonstratives varies with the constellation of the speech participants and the location of the intended referent (see also Coventry et al., 2008). Generalizing across the constellations shown in **Figure 4**, Jungbluth argued that the main determinant for speakers' choice of a particular demonstrative in Spanish is the "conversational dyad" or "shared conversational space."

Similar analyses have been proposed by other scholars for other languages with both two- and three-term systems (Hanks, 1990; Burenhult, 2003; Enfield, 2003; Piwek et al., 2008; Peeters et al., 2015). For instance, Peeters et al. (2015) have argued, based on data from EEG experiments, that the "construal of shared space" determines the alternation between proximal and distal demonstratives in Dutch. The results of this study are complex, but Peeters et al. interpret N400 effects as evidence that proximal demonstratives are preferred in face-to-face constellations for referents in shared space (but only if there is no alternative referent outside of the conversational dyad). Since these effects occurred regardless of the relative distance between speaker and referent, Peeters et al. claim that their results are not consistent with an egocentric and distance-based account of spatial deixis (see also Peeters and Özyürek, 2016).

In a similar vein, Piwek et al. (2008) argue that cognitive accessibility, rather than distance in space, motivates speakers' choice of a particular demonstrative in Dutch. Using a dialogue game in which participants instructed an experimental collaborator to build a lego model, they found that distal demonstratives are preferred for highly accessible referents, whereas low-accessible referents are commonly referred to by proximal demonstratives (see also Kirsner, 1979).

In general, there is a large body of research indicating that demonstrative choice is influenced by shared space and accessibility (Ariel, 1990; Gundel et al., 1993; Laury, 1997; Burenhult, 2003; Enfield, 2003). However, while this sheds new light on demonstrative reference, it does not undermine an egocentric, body-oriented account of spatial deixis. On the contrary, what these studies show is that demonstratives are commonly used to coordinate interlocutors' joint focus of attention, which typically involves body-oriented means of communication such as pointing and gaze that are produced from an egocentric perspective. Assuming that demonstratives are commonly used to manipulate joint attention in multimodal

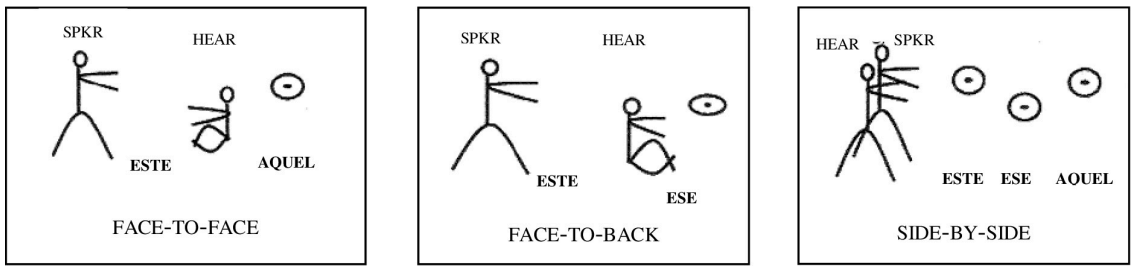


FIGURE 4 | Constellation of speech participants in Jungbluth's study of Spanish demonstratives (adopted from Jungbluth, 2003).

communication, we contend that demonstratives are best analyzed within an egocentric, body-oriented frame of reference.

However, crucially, while the deictic frame of reference is usually grounded in the speaker's body (in the unmarked case), linguistic reference is never immediately determined by physical properties of the outside world—it is always contingent on the conceptualization of space (Talmy, 2000, 2018). Like all other aspects of meaning, deixis is the product of conceptual processes, such as the figure-ground organization, that influence the choice of a particular term (see also Diessel, 2019: 27-30).

What is more, while the speaker's body is commonly interpreted as the origin of a deictic frame of reference, it must be emphasized that the origo, or deictic center, can be shifted from the speaker to another person.

Projected Space

One of Bühler's most important discoveries was that, while demonstratives are commonly used in perceptual space, they can also be used in spatial imagery. Bühler called this "Deixis am Phantasma" and analyzed several distinct cases (Bühler, 1934: 121-140). In the most basic case, the origo is shifted from the speaker onto another person or viewer. The phenomenon is well-known from narratives and spatial descriptions. In narrative discourse, the origo is projected from the speaker, or writer, to a protagonist or narrator who uses demonstratives with reference to objects in the story world (Ehlich, 1979). A similar phenomenon has been observed in linguistic and psycholinguistic research on spatial descriptions (e.g., Ullmer-Ehrich, 1982).

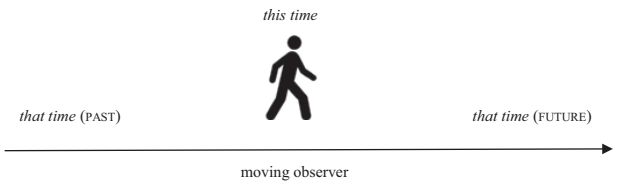
Deictic projections have been investigated in narratology and discourse analysis (Linde and Labov, 1975), but there is little (recent) research on this topic in linguistics and psychology. Stukenbrock (2014, 2015) analyzed deictic projections and other forms of Deixis am Phantasma in video recordings of conversational German. One important finding that has emerged from this research is that space deixis, time deixis and person deixis are not always aligned in spatial imagery. There are interesting blends of deictic projections in Stukenbrock's data in which the deictic dimensions of space, time and person are disassociated from one another.

Another recent study that illustrates the importance of deictic projections for the analysis of demonstrative reference is Rocca et al. (2019c). Using a new interactional paradigm, these researchers found that participants shifted their deictic coordinate system onto a collaborator during a spatial coordination task. Considering this finding, Rocca et al. argue that speakers remap their "action space," or "peripersonal space," onto their "partners' action space" in order to facilitate the collaboration between them.

Deictic projections are crucial to the current debate about the nature of demonstrative reference because they show that even in an egocentric, body-oriented theory of deixis, demonstratives are not always grounded by the speaker. If the origo is shifted, the alternation between proximal and distal terms is determined by the target of the projection rather than the speaker's body or location. Indeed, such an approach is consistent with evidence from other spatial terms—projective adpositions (e.g., *to the left/right*; *in front of*)—where it has been shown empirically that another person's perspective is frequently used to assign direction when speaker and hearer are misaligned (i.e., *the cup is on the (hearer's) left*; see Tversky and Hard, 2009; Tosi et al., 2020).

Beyond Space

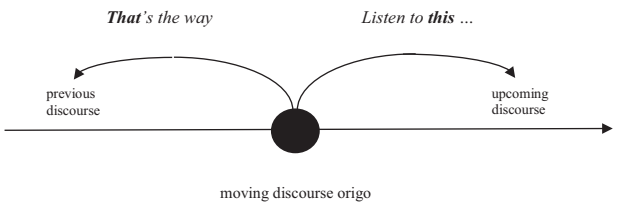
Finally, like many other types of spatial expressions, demonstratives can be extended from the domain of space into non-spatial domains (cf. Bühler, 1934; Fillmore, 1997; Griffiths et al., 2019; Rocca et al., 2019a). To begin with, across languages demonstratives are commonly used with reference to time (cf. *this/that time, month, year*). There is little research on temporal demonstratives, but the extension from space to time is well-known from research on other types of expressions (Haspelmath, 1997; Boroditsky, 2002). If we think of time as a "time line," demonstratives refer to an earlier or later point on that line.



Note that the temporal use of demonstratives involves a radical reconstruction of the conceptual frame that underlies the interpretation of deixis. Spatial demonstratives are interpreted within a coordinate system that is usually evoked by speakers' body, gaze and gesture, but can also be derived from other aspects of the context (in spatial imagery for instance) (Diessel,

2014). However, in contrast to the conceptualization of space deixis, the conceptualization of time deixis does not involve a body-oriented frame of reference, as evidenced by the fact that temporal demonstratives are not accompanied by pointing, gaze or body posture. Only one study to date has examined how demonstratives are used temporally and spatially within the same context. Griffiths et al. (2019) ran a series of experiments eliciting demonstratives to refer to objects, manipulating *where* objects were located in (virtual) space and also *when* objects appeared (e.g., objects appeared and disappeared at different times). They found that demonstratives were used according to whether the object was reachable or not, but there were no effects of time of object appearance/disappearance on demonstrative choice. One interpretation of these findings is that the spatial determinants of demonstrative use take precedence over non-spatial uses, consistent with conceptual metaphor theory (Lakoff and Johnson, 1980).

Like time deixis, discourse deixis often involves demonstratives. Since language unfolds in time, discourse is commonly conceptualized as a continuous stream of linguistic elements. There is a close connection between time deixis and discourse deixis. Both are construed as a band of successive elements that is divided into distinct areas by a moving origo. However, while the origo of time deixis is determined by the moment of speaking, the origo of discourse deixis is determined by the location of a demonstrative in the unfolding discourse.



As Bühler (1934: 390) put it:

If discourse deictic expressions could speak, they "would speak more or less as follows: look ahead or back along the band of the present utterance. There something will be found that actually belongs here, where I am, so that it can be connected with what now follows. Or the other way around: what comes after me belongs there, it was only displaced from that position for relief." [English translation from Goodwin 1990: 443]

The discourse use of demonstratives has been investigated in a large number of studies using both corpus and experimental methods (e.g., Gundel et al., 1993; Himmelmann, 1996; Kaiser and Trueswell, 2008; Kehler et al., 2008; Cornish, 2011; Fossard et al., 2012; Kehler and Rohde, 2013, 2019). The results of this research are too complex to be reviewed in this paper, but there is one finding we'd like to mention as it concerns the current debate about the nature of demonstrative reference. While the notion of relative distance is not immediately relevant to the discourse use of demonstratives, there is evidence that speakers' choice between proximal and distal terms in discourse is influenced by the same psychological factors as demonstrative choice in perceptual space, i.e., by accessibility (Ariel, 1990), common ground (Kaiser and Trueswell, 2008) and manual affordances (Rocca et al., 2019a). In fact, Talmy (2018) argues that the discourse use of demonstratives (which he calls "anaphoric") involves the same cognitive processes as the perceptual use of demonstratives (which he calls "deictic"). On Talmy's account, language includes a single "targeting system" that underlies demonstrative reference in both perceptual space and discourse processing.

CONCLUSION

In conclusion, in this article we have reviewed linguistic and psycholinguistic research on demonstratives from many different perspectives. There is widespread consensus in the literature that demonstratives constitute a unique class of expressions that are crucially distinct from other closed-class items: Demonstratives are likely to be universal and not derived from content words, they seem to be among the earliest and most frequent words in L1 acquisition, they are closely related to pointing, gaze and body posture, and they are of fundamental significance to the diachronic evolution of grammar.

But why are demonstratives so special? What distinguishes them from adpositions, auxiliaries and other closed-class items. We suggest that demonstratives have a particular status in language because of their communicative function to create and to manipulate joint attention (Diessel, 2006).

Joint attention is a prerequisite for social interaction, language acquisition and language evolution (Tomasello, 1999), and it is closely related to spatial deixis. While joint attention is defined as a social phenomenon, it is created by nonverbal means of deictic reference such as pointing and gaze that involve the human body, notably the actor's body, as a source of spatial orientation. Since demonstratives are commonly combined with pointing and gaze, it is reasonable to assume that, in the unmarked case, it is the speaker's body and gesture that provide a (deictic) frame of reference for the semantic interpretation of demonstratives.

There is compelling evidence that demonstrative reference has its roots in our bodily experience with objects in space (Coventry et al., 2008, 2014); but, crucially, deictic space must not be confused with physical space. Some recent studies have criticized the egocentric, body-oriented view of deixis because not all uses of demonstratives involve speakers' body in physical space (e.g., Peeters and Özyürek, 2016). However, this critique is unfounded as it does not recognize the role of conceptualization in the creation of deixis. As Bühler (and many others) have noted, while the deictic center is usually grounded by a speaker's body at the time of an utterance, it can be construed in flexible ways. As we have seen, the origo may be a small spot or a large region, it may or may not include the addressee, and it can be shifted to another person or viewer and mapped onto nonspatial domains, notably the domains of time and discourse.

It should be noted that it is often hard to compare results from studies that employ such a wide range of methodologies – from linguistic work in the field, often (by necessity) with small numbers of informants that makes generalization difficult, to controlled experimental studies with increased (statistical) power, but sampling linguistic behavior in more circumscribed situations. Nevertheless, it would seem that the factors that influence the conceptualization of a deictic

frame of reference are many. Of particular importance is interlocutors' common ground (Clark et al., 1983), but salience in sensory perception and language users' experience with particular types of expressions are also important (Talmy, 2000). This explains why demonstrative reference is so tremendously variable. However, like many other aspects of meaning, the meaning of demonstratives is ultimately based on our bodily experience with objects in space.

AUTHOR CONTRIBUTIONS

The manuscript was jointly written by HD and KC. Both authors contributed to the article and approved the submitted version.

FUNDING

This paper is an integral part of the research project on deictic communication funded by the European Commission: Deictic Communication—A Multidisciplinary Training (DCOMM), Project Number: 676063.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.555265/full#supplementary-material

REFERENCES

Anderson, S. R., and Keenan, E. L. (1985). "Deixis," in *Language Typology and Syntactic Description*, Vol. 3, ed. T. Shopen (Cambridge: Cambridge University Press), 259–308.

Arbib, M. (2012). *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi: 10.1111/j.0956-7976.2004.00694.x

Bates, E. (1976). *Language in Context*. New York, NY: Academic Press.

Berghäll, L. (2015). *A Grammar of Mauwake*. Leipzig: Language Science Press.

Bohnemeyer, J. (2018). "Yucatec demonstratives in interaction: spontaneous vs. elicited data," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 176–205. doi: 10.1017/9781108333818.009

Bonfiglioli, C., Finnocchiaro, C., Gesierich, B., Rositani, F., and Vescovi, M. (2009). A kinematic approach to the conceptual representation of this and that. *Cognition* 111, 270–274.

Boroditsky, L. (2002). Metaphorical structuring: understanding time through spatial metaphors. *Cognition* 75, 1–28. doi: 10.1016/s0010-0277(99)00073-6

Breunesse, M. (2019). *Demonstratives in Space and Discourse: A Synchronic and Diachronic Analysis*. PhD dissertation, University of Jena, Jena.

Brugmann, K. (1904). *Demonstrativpronomina der Indogermanischen Sprachen*. Leipzig: Teubner.

Bufacchi, R. J., and Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends Cogn. Sci.* 22, 1076–1090. doi: 10.1016/j.tics.2018.09.004

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.

Burenhult, N. (2003). Attention, accessibility, and the addressee: the case of the Jahai demonstrative ton. *Pragmatics* 13, 363–379. doi: 10.1075/prag.13.3.01bur

Bybee, J. (2003). "Cognitive processes in grammaticalization," in *The New Psychology of Language*, ed. M. Tomasello (Mahwah: Lawrence Erlbaum), 145–168.

Caldano, M., and Coventry, K. R. (2019). Spatial demonstratives and perceptual space: to reach or not to reach. *Cognition* 191, 1–6. doi: 10.1016/j.reach.2018.11.001

Carpenter, M., Tomasello, M., and Savage-Rumbaugh, S. (1998). Joint attention and imitative learning in children, chimpanzees and enculturated chimpanzees. *Soc. Dev.* 4, 217–237. doi: 10.1111/j.1467-9507.1995.tb00063.x

Caselli, M. C., Bates, B., Casadio, P., Fenson, L., Sanderl, L., and Weir, J. (1995). A cross-linguistic study of early lexical development. *Cogn. Dev.* 10, 159–199. doi: 10.1016/0885-2014(95)90008-x

Clark, E. V. (1978). "From gesture to word: on the natural history of deixis in language acquisition," in *Human Growth and Development*, eds J. S. Bruner and A. Garton (Oxford: Oxford University Press), 85–120.

Clark, E. V., and Sengul, C. J. (1978). Strategies in the acquisition of deixis. *J. Child Lang.* 5, 457–475. doi: 10.1017/s0305000900002099

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *J. Verb. Learn. Verb. Behav.* 22, 245–258. doi: 10.1016/s0022-5371(83)90189-5

Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discour. Process.* 53, 632–656. doi: 10.1080/0163853x.2015.1094280

Coppola, M., and Senghas, A. (2010). "Deixis in an emerging sign language," in *Sign Languages*, ed. D. Brentari (Cambridge: Cambridge University Press), 543–569. doi: 10.1017/cbo9780511712203.025

Cormier, K., Schembri, A., and Woll, B. (2013). Pronouns and pointing in sign language. *Lingua* 36, 230–247. doi: 10.1016/j.lingua.2013.09.010

Cornish, F. (2011). 'Strict' anadeixis, discourse deixis and text structuring. *Lang. Sci.* 33, 753–767. doi: 10.1016/j.langsci.2011.01.001

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near-far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

de Villiers, P., and Villiers, J. (1974). On this, that, and the other: nonegocentrism in very young children. *J. Exp. Child Psychol.* 18, 438–447. doi: 10.1016/0022-0965(74)90122-2

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Diessel, H. (2012a). "Buehlers two-field theory of pointing and naming and the deictic origins of grammatical morphemes," in *New Perspectives on Grammaticalization: Theoretical Understanding and Empirical Description*, eds T. Breban, L. Brems, K. Davidse, and T. Mortelmans (Amsterdam: John Benjamins), 35–48.

Diessel, H. (2012b). "Deixis and demonstratives," in *An International Handbook of Natural Language Meaning*, Vol. 3, eds C. Maienborn, K. von Heusinger, and P. Portner (Berlin: Mouton de Gruyter), 2407–2431.

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Lang. Cogn.* 5, 239–249. doi: 10.1515/langcog-2013-0017

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Lang. Linguist. Compass* 8, 116–132. doi: 10.1111/lnc3.12066

Diessel, H. (2019). *The Grammar Network. How Linguistic Structure is Shaped by Language Use*. Cambridge: Cambridge University Press. doi: 10.1017/9781108671040

Diessel, H., and Breunesse, M. (2020). "A typology of demonstrative clause linkers," in *Demonstratives in Discourse*, eds Å Næss, A. Margetts, and Y. Treis (Berlin: Language Science Press), 305–341.

Dixon, R. M. W. (2003). Demonstratives. A cross-linguistic typology. *Stud. Lang.* 27, 61–122. doi: 10.1075/sl.27.1.04dix

Döhler, C. (2018). A *Grammar of Komnzo*. Leipzig: Language Science Press. doi: 10.5281/zenodo.1477799

Durie, M. (1985). *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht: Foris.

Ehlich, K. (1979). *Verwendung der Deixis beim sprachlichen Handeln. Linguistisch-philologische Untersuchungen zum hebäischen deiktischen System*. Frankfurt: Peter Lang.

Enfield, N. (2003). Demonstratives in space and interaction: data from Lao speakers and implications for semantic analysis. *Language* 79, 82–117. doi: 10.1353/lan.2003.0075

Engberg-Pedersen, E. (2003). "From pointing to reference and predication: pointing signs, eye gaze, and head and body orientation in Danish sign language," in *Pointing. Where Language, Culture and Cognition Meet*, ed. S. Kita (Makwah, NJ: Lawrence Erlbaum), 269–292.

Eriksson, M. (2008). Referring as interaction: on the interplay between linguistic and bodily practices. *J. Pragm.* 41, 240–262. doi: 10.1016/j.pragma.2008.10.011

Etelämäki, M. (2009). The Finnish demonstrative pronouns in light of interaction. *J. Pragm.* 41, 25–46. doi: 10.1016/j.pragma.2008.09.005

Evans, N. (1990). "Without this child – some regularities of semantic change in the Australian linguistic area," in *Language and History: Essays in Honour of Luise A. Hercuse*, eds P. Austin, R. M. W. Dixon, T. Dutton, and I. White (Canberra: Pacific Linguistics), 137–155.

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–448. doi: 10.1017/s0140525x0999094x

Fenlon, J., Cooperrider, K., Keane, J., Brentari, D., and Goldin-Meadow, S. (2019). Comparing sign language and gesture: insights from pointing. *Glossa* 4:2. doi: 10.5334/gjgl.499

Fillmore, C. J. (1997). *Lectures on Deixis*. Stanford: CSLI Publications.

Forker, D. (2019). Elevation as a category of grammar: sanzhi dargwa and beyond. *Linguist. Typol.* 23, 59–106. doi: 10.1515/lingty-2019-0001

Fossard, M., Garnham, A., and Cowles, H. W. (2012). Between anaphora and deixis . . . The resolution of the demonstrative noun phrase "that N". *Lang. Cogn. Process.* 27, 1385–1404. doi: 10.1080/01690965.2011.606668

Frajzyngier, Z. (1996). "On sources of demonstratives and anaphors," in *Studies in Anaphora*, ed. B. Fox (Amsterdam: John Benjamins), 169–203. doi: 10.1075/tsl.33.07fra

Fricke, E. (2014). "Deixis, gesture, and embodiment from a linguistic point of view," in *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Vol. 1, eds M. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Teßendorf (Berlin: Mouton de Gruyter), 1803–1823.

Friedman, L. A. (1975). Space, time, and person reference in American sign language. *Language* 51, 940–961. doi: 10.2307/412702

Garciá, J. L. P., Rosendal Ehlers, K., and Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: the contribution of a de-pigmented sclera to proto-declaratives. *Lang. Commun.* 54, 73–81. doi: 10.1016/j.langcom.2016.10.007

Gipper, S. (2017). Pre-semantic pragmatics encoded: a non-spatial account of Yurakaré demonstratives. *J. Prag.* 120, 122–143. doi: 10.1016/j.pragma.2017.08.012

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Griffiths, D., Bester, A., and Coventry, K. R. (2019). Space trumps time when talking about objects. *Cogn. Sci.* 43:e12719. doi: 10.1111/cogs.12719

Gudde, H. B., Coventry, K. R., and Engelhardt, P. E. (2016). Language and memory for object location. *Cognition* 153, 99–107. doi: 10.1016/j.cognition.2016.04.016

Guérin, V. (2015). Demonstrative verbs: a typology of verbal manner deixis. *Linguist. Typol.* 19, 141–199.

Gundel, J., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535

Hanks, W. F. (1990). *Referential Practice*. Chicago, IL: Chicago University Press.

Hasegawa, Y. (2012). Deictic and anaphoric uses of the Japanese demonstratives ko-so-a. *J. Jpn. Linguist.* 28, 44–61.

Haspelmath, M. (1997). *From Space to Time*. Munich: Lincom.

Heine, B., and Kuteva, T. (2002). *World Lexicon of Grammaticalization*. Oxford: Oxford University Press.

Heine, B., and Kuteva, T. (2007). *The Genesis of Grammar. A Reconstruction*. Oxford: Oxford University Press.

Hellwig, B. (2003). *The Grammatical Coding of Postural Semantics in Goemai (a West Chadic Language of Nigeria)*. Nijmegen: Max Planck Series in Psycholinguistics.

Herrmann, S. (2018). "Warao demonstratives," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 282–302.

Himmelmann, N. (1996). "Demonstratives in narrative discourse: a taxonomy of universal uses," in *Studies in Anaphora*, ed. B. Fox (Amsterdam: John Benjamins), 205–254. doi: 10.1075/tsl.33.08him

Himmelmann, N. (1997). *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur*. Tübingen: Narr.

Hindmarsh, J., and Heath, C. (2000). Embodied reference: a study of deixis in workplace interaction. *J. Prag.* 32, 1855–1878. doi: 10.1016/s0378-2166(99)00122-8

Hopper, P. J., and Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Jarbou, S. O. (2010). Accessibility vs. physical proximity: an analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *J. Prag.* 42, 3078–2097. doi: 10.1016/j.pragma.2010.04.014

Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of 'pronouns' in Auslan? *Corpus Linguist. Linguist. Theory* 9, 109–159.

Jungbluth, K. (2003). "Deictics in the conversational dyad," in *Deictic Conceptualizations of Space, Time and Person*, ed. F. Lenz (Amsterdam: John Benjamins), 13–40. doi: 10.1075/pbns.112.04jun

Jungbluth, K. (2005). *Die Pragmatik der Demonstrativepronomina in den Iberoromanischen Sprachen*. Berlin: Mouton de Gruyter.

Kaiser, E., and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference. *Lang. Cogn. Process.* 23, 709–748. doi: 10.1080/01690960701771220

Kehler, A., Kertz, L., Rohde, H., and Elman, J. (2008). Coherence and coreference revisited. *J. Semant.* 25, 1–44. doi: 10.1093/jos/ffm018

Kehler, A., and Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theor. Linguist.* 39, 1–37. doi: 10.1515/tl-2013-0001

Kehler, A., and Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *J. Prag.* 154, 63–74. doi: 10.1016/j.pragma.2018.04.006

Kemmerer, D. (1999). "Near" and "far" in language and perception. *Cognition* 73, 35–63. doi: 10.1016/s0010-0277(99)00040-2

Kirsner, R. S. (1979). "Deixis in discourse: an explanatory quantitative study of the modern Dutch demonstrative adjectives," in *Syntax and Semantics, Vol. 12: Discourse and Syntax*, ed. T. Givón (New York: Academic Press), 355–375. doi: 10.1163/9789004368897_016

Klima, E. S., and Bellugi, U. (1979). *The Signs of Language*. Cambridge: Harvard University Press.

König, E. (2012). Le rôle des déictiques de manière dans le cadre d'une typologie de la déixis. *Bull. Soc. Linguist. Paris* 107, 11–42.

Küntay, A., and Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *J. Child Lang.* 33, 303–320. doi: 10.1017/s0305000906007380

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Laury, R. (1997). *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish*. Amsterdam: John Benjamins.

Levinson, S. C. (2003). *Space in Language and Cognition*. Cambridge: Cambridge University Press.

Levinson, S. C. (2004). "Deixis and pragmatic," in *The Handbook of Pragmatics*, eds L. Horn and G. Ward (Oxford: Blackwell), 97–121.

Levinson, S. C. (2018). "Demonstratives – Patterns of diversity," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781108333818.002

Levinson, S. C., Cutfield, S., Dunn, M., Enfield, E., and Meira, S. (eds) (2018). *Demonstratives in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Liddell, S. K. (2000). "Blended spaces and deixis in sign language discourse," in *Language and Gesture*, ed. D. MacNeill (Cambridge: Cambridge University Press), 331–357. doi: 10.1017/cbo9780511620850.021

Linde, C., and Labov, W. (1975). Spatial structure as a site for the study of language and thought. *Language* 51, 924–939. doi: 10.2307/412701

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., and de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cogn. Sci.* 36, 698–713. doi: 10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Striano, T., and Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *J. Cogn. Dev.* 7, 173–187. doi: 10.1207/s15327647jcd0702_2

Lücking, A., Pfeiffer, T., and Rieser, H. (2015). Pointing and reference reconsidered. *J. Prag.* 77, 56–79. doi: 10.1016/j.pragma.2014.12.013

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, Vol. 2, the Database*. Hillsdale NJ: Lawrence Erlbaum.

Maslova, E. (2003). *A Grammar of Kolyma Yukaghir*. Berlin: Mouton de Gruyter.

Meier, R. P., and Lillo-Martin, D. (2010). "Does spatial language make it special? On the grammar of pointing signs in American sign language," in *HypothesisA/Hypothesis B. Linguistic Explorations in Honor of David M. Perlmutter*, eds D. B. Gerdts, J. C. Moore, and M. Polinsky (Cambridge: MIT Press), 345–360.

Meira, S. (2018). "Tiriyó – Non-contrastive exophoric uses of demonstratives," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 222–241. doi: 10.1017/9781108333818.011

Morford, J. P., Shaffer, B., Shin, N., Twitchell, P., and Petersen, B. T. (2019). An exploratory study of ASL demonstratives. *Languages* 4:80. doi: 10.3390/languages4040080

Nelson, K. (1973). Structure and Strategy in Learning to Talk. *Monogr. Soc. Res. Child Dev.* 38, 1–135. doi: 10.2307/1165788

Peeters, D., Hagoort, P., and Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition* 136, 64–84. doi: 10.1016/j.cognition.2014.10.010

Peeters, D., and Özyürek, A. (2016). This and that revisited: a social and multimodal approach to spatial demonstratives. *Front. Psychol.* 7:222. doi: 10.3389/fpsyg.2016.00222

Pfau, R., and Steinbach, M. (2011). "Grammaticalization in sign languages," in *The Oxford Handbook of Grammaticalization*, eds H. Narrog and B. Heine (Oxford: Oxford University Press), 683–695.

Piwek, P., Beun, R. J., and Cremers, A. (2008). Proximal and distal language and cognition: evidence from deictic demonstratives in Dutch. *J. Prag.* 40, 694–718. doi: 10.1016/j.pragma.2007.05.001

Reile, M., Plado, H., Gudde, H. B., and Coventry, K. R. (2020). Demonstratives as spatial deictics or something more? Evidence from Common Estonian and Võro. *Folia Linguistica* 54, 167–195. doi: 10.1515/flin-2020-2030

Rocca, R., Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., and Wallentin, M. (2019a). Language beyond the language system: dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fMRI. *NeuroImage* 216:116128. doi: 10.1016/j.neuroimage.2019.116128

Rocca, R., Tylén, K., and Wallentin, M. (2019b). This shoe, that tiger: semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLoS One* 14:e0210333. doi: 10.1371/journal.pone.0210333

Rocca, R., Wallentin, M., Vesper, C., and Tylén, K. (2019c). This is for you: social modulations of proximal vs. distal space in collaborative interactions. *Sci. Rep.* 9, 1–14.

Salerni, N., Assanelli, A., D'Odorico, L., and Rossi, G. (2007). Qualitative aspects of productive vocabulary at the 200- and 500-word stages: a comparison between spontaneous speech and parental report data. *First Lang.* 27, 75–87. doi: 10.1177/0142723707067545

Senft, G. (2004). "Aspects of spatial deixis in Kilivila," in *Deixis and Demonstratives in Oceanic Languages*, ed. G. Senft (Canberra: Australian National University), 59–80.

Shimelman, A. (2017). *A Grammar of Yauyos Quechua*. Berlin: Language Science Press.

Smeets, I. (1989). *A Mapuche Grammar*. PhD thesis, University of Leiden, Leiden.

Stevens, J., and Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: an event-related potential study. *J. Neurolinguist.* 26, 31–45. doi: 10.1016/j.jneuroling.2012.02.005

Strauss, S. (2002). This, that, and it in spoken American English: a demonstrative system of gradient focus. *Lang. Sci.* 24, 131–152.

Stukenbrock, A. (2014). Pointing to an 'empty' space: deixis am phantasma in face-to-face interaction. *J. Prag.* 74, 70–93. doi: 10.1016/j.pragma.2014.08.001

Stukenbrock, A. (2015). *Deixis in der face-to-face-Interaktion*. Berlin: Mouton de Gruyter.

Talmy, L. (2000). *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

Talmy, L. (2018). *The Targeting System of Language*. Cambridge, MA: MIT Press.

Tanz, C. (1980). *Studies in the Acquisition of Deictic Terms*. Cambridge, MA: Cambridge University Press.

Todisco, E., Guijarro-Fuentes, P., Collier, J., and Coventry, K. R. (2020). The temporal dynamics of deictic communication. *First Lang.* [Epub ahead of print].

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 721–727. doi: 10.1017/S0140525X05000129

Tosi, A., Pickering, M. J., and Brannigan, H. P. (2020). Speakers' use of agency and visual context in spatial descriptions. *Cognition* 194:104070. doi: 10.1016/j.cognition.2019.104070

Treis, Y. (2018). Presentative demonstratives in Kambaata from a Cushitic perspective. Archive Ouverte HAL hal-01505569v2.

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008

Ullmer-Ehrich, V. (1982). "The structure of living room descriptions," in *Speech, Place and Action. Studies in Deixis and Related Topics*, eds R. J. Jarvella and W. Klein (New York: Wiley), 219–215.

Vindenes, U. (2017). *Complex demonstratives and cyclic change in Norwegian*. PhD dissertation, University of Oslo, Oslo.

Webb, P., and Abrahamson, A. (1976). Stages of egocentrism in children's use of this and that: a different point of view. *J. Child Lang.* 3, 349–367. doi: 10.1017/s0305000900007236

Wilkins, D. P. (2018). "The demonstrative questionnaire: "THIS" and "THAT"," in *comparative perspectives. In Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Cambridge: Cambridge University Press), 43–71. doi: 10.1017/9781108333818.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 02 December 2020
doi: 10.3389/fpsyg.2020.575497

Shrinking Your Deictic System: How Far Can You Go?

Mila Vulchanova[1], Pedro Guijarro-Fuentes[2], Jacqueline Collier[3] and Valentin Vulchanov[1]*

[1] Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science & Technology, Trondheim, Norway, [2] Departamento de Filología Española, Moderna y Clásica, Universidad de Islas Baleares, Palma, Spain, [3] School of Psychology, University of East Anglia, Norwich, United Kingdom

Languages around the world differ in terms of the number of adnominal and pronominal demonstratives they require, as well as the factors that impact on their felicitous use. Given this cross-linguistic variation in deictic demonstrative terms, and the features that determine their felicitous use, an open question is how this is accommodated within bilingual cognition and language. In particular, we were interested in the extent to which bilingual language exposure and practice might alter the way in which a bilingual is using deictic demonstratives in their first language. Recent research on language attrition suggests that L2 learning selectively affects aspects of the native language, with some domains of language competence being more vulnerable than others. If demonstratives are basic, and acquired relatively early, they should be less susceptible to change and attrition. This was the hypothesis we went on to test in the current study. We tested two groups of native Spanish speakers, a control group living in Spain and an experimental group living in Norway using the (Spatial) Memory game paradigm. Contra to our expectations, the results indicate a significant difference between the two groups in use of deictic terms, indicative of a change in the preferred number of terms used. This suggests that deictic referential systems may change over time under pressure from bilingual language exposure.

Keywords: demonstratives, bilingualism, language attrition, norwegian, spatial memory game, spanish sample

OPEN ACCESS

Edited by:
Holger Diessel, Friedrich Schiller University Jena, Germany

Reviewed by:
Merlijn Breunesse, University of Amsterdam, Netherlands
Antonella Sorace, University of Edinburgh, United Kingdom

****Correspondence:***
Mila Vulchanova
mila.vulchanova@ntnu.no

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *23 June 2020*
Accepted: *20 October 2020*
Published: *02 December 2020*

Citation:
Vulchanova M, Guijarro-Fuentes P, Collier J and Vulchanov V (2020) Shrinking Your Deictic System: How Far Can You Go? Front. Psychol. 11:575497. doi: 10.3389/fpsyg.2020.575497

INTRODUCTION

Demonstratives are function words typically used to refer to physical, concrete entities in a real-world speech situation. Utterance of the demonstrative, often accompanied by a pointing gesture (Bühler, 1934; Diessel, 1999, 2006; Levinson, 2004), has an important communicative upshot. It aims to focus the attention of the addressee on a particular entity in the shared perceptual or visual field of the interlocutors. Languages around the world differ in terms of the number of adnominal and pronominal demonstratives they require (Diessel, 1999), as well as the factors that impact on their felicitous use. Distance from the deictic center (i.e., the speaker) has been identified as the most common feature encoded in demonstratives cross-linguistically (Lyons, 1977; Anderson and Keenan, 1985; Diessel, 1999). Following recent in-depth empirical and experimental research, this tradition has been called into question, often referred to as the "spatial bias" in accounts of demonstratives (Levinson, 2018). Thus, in addition to distance from speaker, and/or hearer, visibility, ownership, possibility to interact with the reference object and other features of the speaker-hearer constellation have been shown to be relevant for deictic term use, even in a language

like English, which does not encode these features lexically (Coventry et al., 2008, 2014; Levinson et al., 2018).

Given the cross-linguistic variation in deictic demonstrative terms, and the features that determine their felicitous use, an open question is how this is accommodated within bilingual cognition and language. In particular, we are interested in the extent to which daily immersive exposure to a second language and practice might alter the way in which bilingual speakers are using deictic demonstratives in their first language. Recent research on language attrition suggests that L2 learning selectively affects aspects of the native language, with some domains of language competence being more vulnerable than others (Jakobson, 1941; de Bot and Weltens, 1991; Keijzer, 2007). Native language vulnerability is subject to individual variation, and specific factors in the bilingual speaker's background, such as level of education, literacy etc. (Köpke, 2007). At the same time, it has been shown that the more robust aspects of language are those that are typically acquired early and sub-serve basic language functions (Jakobson, 1941; Keijzer, 2007). If demonstratives are basic (Dixon, 2003), and acquired relatively early (Clark, 1978; Diessel, 2006), they should be less susceptible to change and attrition. This was the hypothesis we went on to test in the current study.

The languages in the current study are a three-term language, Spanish, and a two- term language, Norwegian. Diessel (1999; 2005; 2013) and Dixon (2003) provide a comprehensive survey of cross-linguistic variation in relation to the system of demonstratives and the parameters affecting the choice of demonstratives in specific contexts. In Diessel's schematization (2005, 2013), for instance, a two-termed proximal/distance contrast system has a higher frequency (54.4%) than the three-termed contrast (37.4%), and other combinations of demonstratives (8%). In addition, within the frame of proximal/distal opposition, the distance-oriented system is the most widespread (two thirds of the languages analyzed; Diessel, 2005, 2013) in comparison to the person-oriented system.

Spanish features a tripartite demonstrative system with three elements (*este*, *ese,* and *aquel*) (Jungbluth and Da Milano, 2015), which can inflect for gender and number and are used adnominally. In addition, Spanish has three demonstrative pronouns (*esto*, *eso,* and *aquello*), which do not inflect and have, nevertheless, been traditionally labeled as neuter demonstrative pronouns in the Spanish grammatical tradition (although there is not clearly a neuter grammatical gender in Spanish *per se*). The Spanish demonstrative terms are commonly characterized as conveying different degrees of distance with respect to the deictic center (the speaker): *este* ("this") is proximal, *ese* ("that") medial, and *aquel* ("that yonder") is the distal demonstrative of the tripartite system. The Spanish demonstrative system, can thus be seen as gravitating toward an egocentric, distance-oriented preference usage, which accounts for the proximal, medial and distal forms in relation to the speaker, with little or no consideration of the position of the hearer (Diessel, 1999; Jungbluth, 2003; Coventry et al., 2008; Jungbluth and Da Milano, 2015). This is also consistent with Hottenroth (1982) who suggests that the "proximal-medial-distal form designates increasingly remote concentric circles around the speaker" (p. 133). Jungbluth (2003); Coventry et al. (2008), and Jungbluth and Da Milano (2015) presented a more detailed description of the Spanish demonstrative system, taking into account the effect of the hearer's position in the choice of demonstratives. Jungbluth (2003) and Jungbluth and Da Milano (2015), for instance, suggested a dual-oriented system of interaction with three possible conditions ("constellations") with respect to the hearer: face-to-face, side-by-side, and face-to-back. During semi-naturalistic performances, Spanish monolingual speakers preferred a distance-oriented system in a side-by-side condition, a person-oriented system in a face-to-face condition and both a person-oriented and a distance-oriented system in a face-to-back condition. Coventry et al. (2008) provide experimental evidence that hearer position impacts on the use of the three terms, and interacts with distance.

Norwegian is a two-term system. Traditionally, the demonstrative pronouns *denne* and *den* have been considered to reflect the contrast between proximal (*denne*) and distal (*den*) object locations (Faarlund et al., 1997). However, the modern colloquial language uses a spatial adverb [*her* (*here*) and *der* (*there*)] as a reinforcement of both *denne* (proximal) and *den* (distal), thus yielding the so-called complex demonstrative forms *den/denne her* (*this here*) (Johannessen, 2006). This possibility comes to suggest that the form *den*, originally assumed to be distal, has evolved into a neutral form rather than signaling distance (Halmøy, 2016). This is further confirmed by the possibility of combining *den* with the distal adverb *der* (*there*), with *den der* meaning "that one over there." Adverbs denoting location have been the source of reinforcing expressions in several languages world-wide. Furthermore, when a demonstrative adverb is used adnominally, it usually does not function as a modifier of the noun, but rather as a reinforcement of the co-occurring demonstrative determiner. Vindenes (2018) argues that speaker strategies that are used to achieve joint attention are particularly important mechanisms in the (diachronic) process of reinforcement of demonstratives, also evidenced in the Modern Norwegian situation. While Spanish has been studied experimentally, to our knowledge there is no such research on Norwegian.

Dixon (2003) points out that a three-term system of demonstratives might convey either a relative distance (i.e., near, mid and far) or relate to the participant (i.e., near the speaker, near the hearer, near neither), but also to height, stance, visibility as well as elevation and movement (Diessel, 1999; Breunesse, 2020). Other parameters affecting the choice of demonstratives may refer to perspective-taking (e.g., for Turkish, Küntay and Özyürek, 2005), sociocentric proximity (Stevens and Zhang, 2013, 2014; Peeters et al., 2015), semantic features (Rocca et al., 2019), ownership, visibility, and familiarity of referent (Coventry et al., 2014), and proximity/distance of referent in relation to both speaker and hearer (i.e., Spanish, Catalan, and Japanese, Diessel, 1999; Jungbluth, 2003; Coventry et al., 2008).

Given these considerations, the difference between the Spanish and Norwegian adnominal/pronominal demonstrative systems mainly lies in the morpho-lexical choice of demonstrative term, and the number of such terms, while both systems might equally well reflect other semantic distinctions, as documented in extant research.

In the current study we were interested in the extent to which a subsequently acquired two-term system (Norwegian) might impact on the original three-term L1 system (Spanish) in adult language users. Our predictions were that closed-class systems of the deictic type are not easily attrited. However, we did expect subtle deviations from the native Spanish system in terms of specific distinctions (e.g., distance magnitude), and we expected this effect to be attributable to length of stay in Norway. In line with Coventry et al. (2008) we also expected position of hearer to influence participants' responses.

MATERIALS AND METHODS

Participants

Participants in the experimental group (*Spanish Living in Norway*, henceforth (SLiN)) were 20 adult native speakers of Spanish who had lived in Norway for work or study on average 110,4 months. 2/3rds of the SLiN participants had attended language courses or had experience from Norwegian education, while 1/3rd indicated that they had learned Norwegian naturalistically. Twelve participants rated their level of proficiency in Norwegian as advanced-to-near native, and only two assessed their level as beginners, which reflects advanced knowledge of Norwegian. In addition, all participants (with one exception) stated that they used both languages equally on a daily basis, with some prevalence for Norwegian. They were recruited via various channels, social media, university networks and via social contact. All participants provided signed informed consent prior to the study. Approval for the study and for collecting and storing the data was obtained from the Norwegian Da ta Protection Service (NSD). All SLiN participants had had their first exposure to Norwegian [Age of Arrival (AoA)] after age 20 years. For this reason, we used length of stay as predictor in the analyses.

The control group [*Spanish Living in Spain*; henceforth (SLiS)] comprised $N = 30$ ($MA = 23.5$; $SD = 5.88$; female = 18) native speakers of Castilian Spanish recruited at Universidad de Islas Baleares. Approval for the study and for collecting and storing the data was obtained from Comité de Ética de la Investigación (Universidad de Islas Baleares), and the School of Psychology Ethics Committee at the University of East Anglia as part of a bigger cross-linguistic study. All participants were matched for socioeconomic and educational background. The speakers who volunteered to take part in this study and, therefore, did not get any economic compensation for participation, were residents in Spain at the time of testing.

Stimuli

Participants were tested with the Spanish version of the (Spatial) Memory game (Gudde et al., 2018). The memory game paradigm is a behavioral procedure to explore the relationship between language, spatial memory, and object knowledge and has already been widely used in cross-linguistic research. In two different versions of the paradigm, spatial language use and memory for object location are tested under different, experimentally manipulated conditions. The current study employed only the spatial language use version of the paradigm. Participants were tested in naming markers placed on a table at different distances from the participant (= speaker). In one set up the experimenter (= hearer) was seated next to the participant, and in another, opposite to the participant. We elicited the production of demonstratives by locating six circular plastic disks on top of a conference table. The disks were 6 cm wide and presented different sketched images (see **Figure 1**). The experimenter located the disks on top of 12 colored dots equally distributed on the table (320 * 80 cm, see **Figure 2**). The table was covered by a black cloth. We used the following 6 locations to locate the disks: 25, 50, 150, 175, 275, and 300 cm.

Participants in the study were instructed to use *este*, *ese,* and *aquel* for the Spanish version of the experiment and *den her* (*this here*) and *den der* (*this there*), for the Norwegian version.

Procedure and Design

During the experiment, the participants sat at the table (within 3 cm distance), in front of the line marked by the colored dots (40 cm). The experimenter sat either laterally or frontally with respect to the participant. We instructed the participant to memorize the position of the disks that the experiment was locating on top of the dots. To help the memorization process, (s)he had to use a bimodal production: gestural and verbal. Every time the experimenter sat after locating the disk, the participant had to point at the disk (i.e., gestural performance), without standing up or touching the table. In addition, the participant had to produce a sentence consisting of three elements (i.e., verbal performance): a demonstrative, the color and the image in the disk (i.e., this/that red moon). Every time the participant performed the gestural and verbal production, the experimenter stood up to locate the subsequent disk on the list. The trials presented random breaks with memory questions regarding the last position of one or more disks. The total amount of trials was 36 per participant divided in two sub-sessions of eighteen trials each. On eighteen trials the experimenter sat next to the participant [laterally and on the remaining eighteen trials opposite the participant (frontally)]. We counterbalanced the order of presentation of the stimuli, the locations of the discs on the dots, as well as the position of the experimenter to avoid any effect of order.

The whole session, from welcoming to debriefing was conducted in the language of testing by the experimenter.

FIGURE 1 | Images of the disks. From left to right, the disks presented the following images: a green star, a black cross, a red moon, a yellow triangle, an orange square and a blue heart.

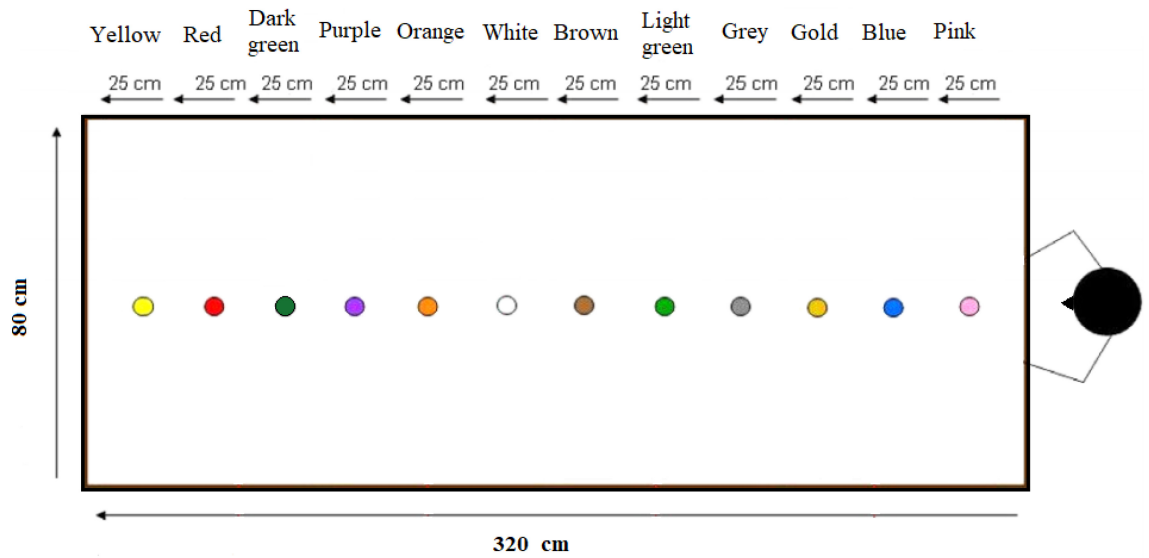


FIGURE 2 | Experiment set up. We used six positions: pink (1st position at 25 cm), blue (2nd position at 50 cm), brown (3rd position at 150 cm), white (4th position at 175 cm), red (5th position at 275 cm) and yellow (6th position at 300 cm). The space could be divided in three subspaces depending on the participants' arm reach: one peri-personal space, within participants' arm reach, and two extra-personal subspaces, out of reach.

For the purposes of indirect comparison, we also tested a group of adult Norwegian native speakers living in Norway (N = 23; MA = 23; SD = 2.87; female = 11) which was part of a bigger cross-linguistic study (Coventry, in preparation). Approval for the study was obtained by the University of East Anglia. The participants had similar educational and socio-economic backgrounds.

ANALYSIS AND RESULTS

Descriptives Before Merging the Data

The Spanish Living in Spain (SLiS) group used the three terms according to distance from speaker regardless of position of hearer. Thus, the proximal term was used exclusively to name the two closest distances (25 and 50 cm), the distal term was used exclusively to name the two outmost distances (275 and 300 cm), while the medial (third) term was used for the medial positions (150 and 175 cm). This was not the case for the Spanish Living in Norway (SLiN) group, whereby the most prevalent term used was the medial term (*ese*) regardless of distance from speaker/hearer at a total of 420 times (58.3%). Thus, overall, the Spanish Living in Norway used *ese* more than those living in Spain (58.3 vs. 38.3%), with minimal reduction in *este* (27.8 vs. 32.1%) and a notable drop in the use of *aquel/aquella* (13.9 vs. 29.1%) (see **Figure 3** and **Table 1**).

In the *Spanish Living in Norway* (SLiN) group there were also 29 occasions when participants used *este* in the 275 and 300 cm positions. These were seen both when the listener was side-by-side or opposite, against zero occurrences of *este* in the *Spanish Living in Spain* (SLiS) group in the 275 and 300 cm positions. Examination of the data showed that 23 of the 29 uses of *este* at 275 or 300 cm were attributable to two individuals (11 times and 12 times apiece), four other individuals used it once, and one further individual used it twice. In line with the hypothesis about time spent living in Norway as a predictor for different use of Spanish demonstratives, the use of *este* at 275 or 300 cm was tabulated alongside time living in Norway. Initial inspection of the data suggests longer exposure to the L2 measured in terms of length living in Norway was not associated with this different use of *este* by these two individuals (note: the median time living in Norway for the whole sample is 84 months, min 3 months, max 444 months) (see **Figure 4** and **Table 2**).

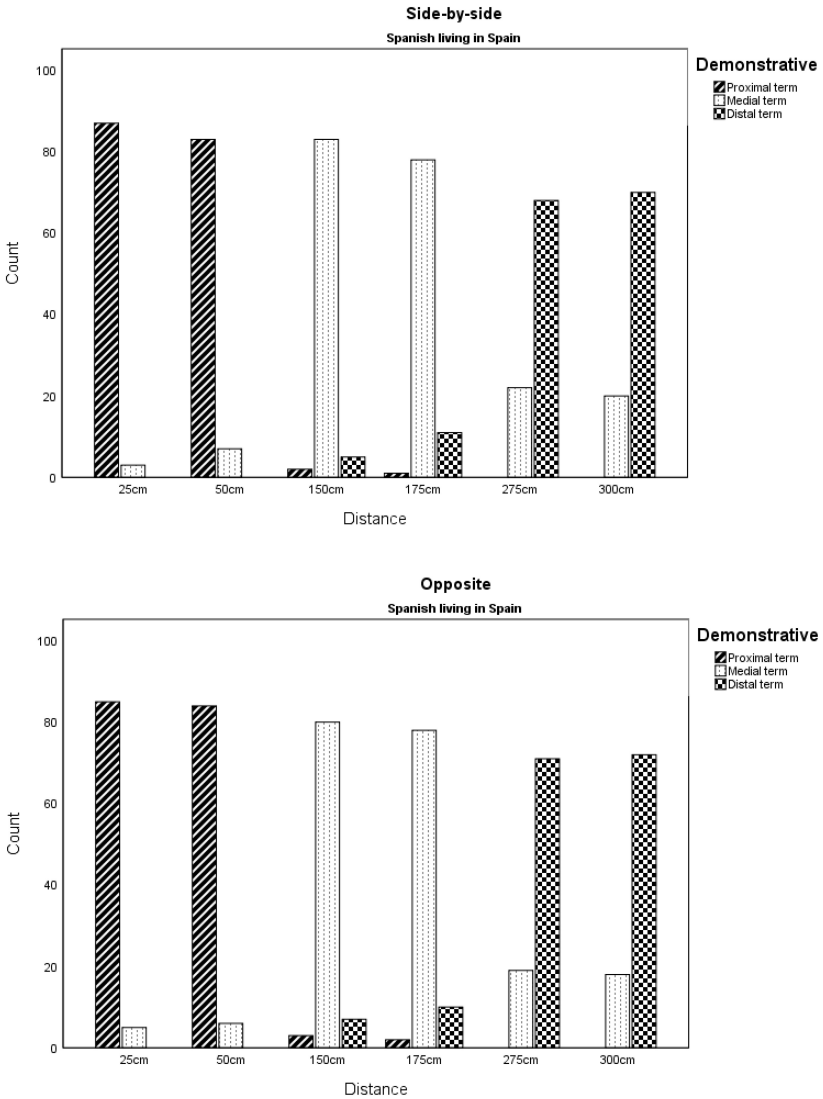


FIGURE 3 | Demonstratives by distance and hearer position for Spanish speakers living in Spain.

TABLE 1 | Demonstratives by distance and hearer position for Spanish speakers living in Spain.

			Demonstrative			Total
			Proximal term	Medial term	Distal term	
Side-by-side	Distance	25 cm	87	3	0	90
		50 cm	83	7	0	90
		150 cm	2	83	5	90
		175 cm	1	78	11	90
		275 cm	0	22	68	90
		300 cm	0	20	70	90
	Total		173	213	154	540
Opposite	Distance	25 cm	85	5	0	90
		50 cm	84	6	0	90
		150 cm	3	80	7	90
		175 cm	2	78	10	90
		275 cm	0	19	71	90
		300 cm	0	18	72	90
	Total		174	206	160	540

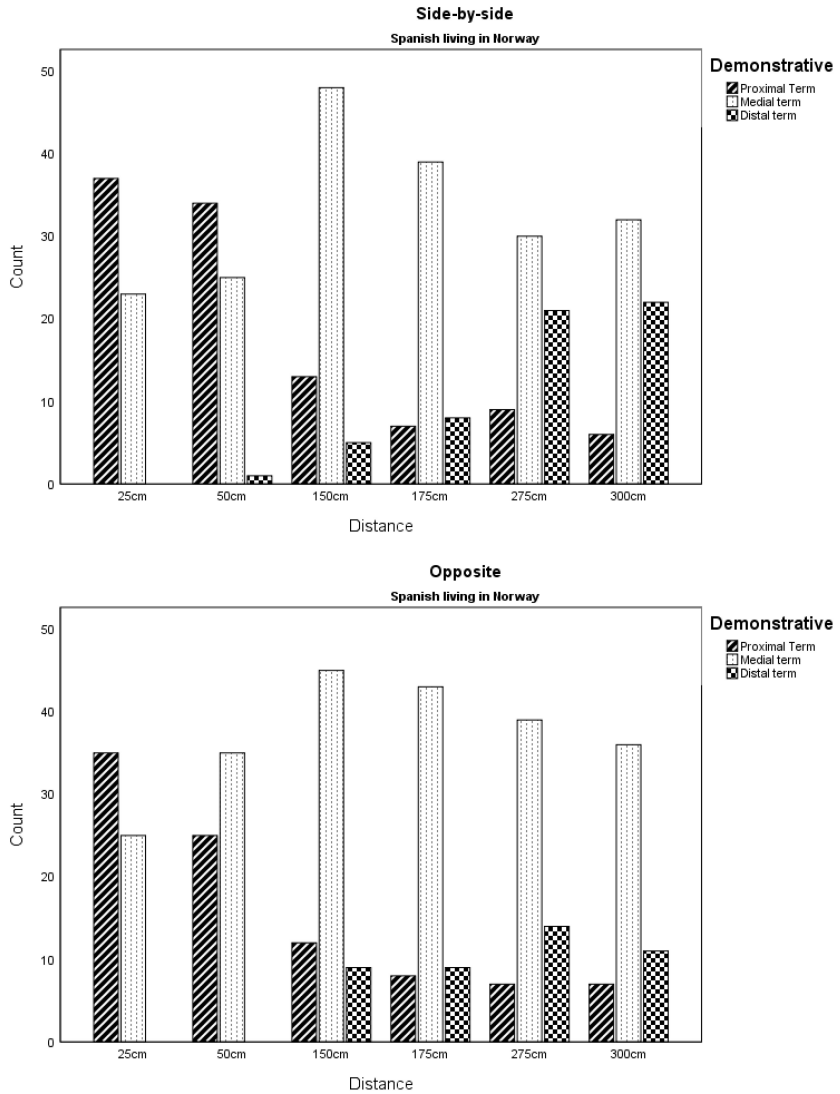


FIGURE 4 | Demonstratives by distance and hearer position for Spanish speakers living in Norway.

TABLE 2 | Demonstratives by distance and hearer position for Spanish speakers living in Norway.

			Demonstrative			Total
			Proximal term	Medial term	Distal term	
Side-by-side	Distance	25 cm	37	23	0	60
		50 cm	34	25	1	60
		150 cm	13	48	5	66
		175 cm	7	39	8	54
		275 cm	9	30	21	60
		300 cm	6	32	22	60
	Total		106	197	57	360
Opposite	Distance	25 cm	35	25	0	60
		50c m	25	35	0	60
		150 cm	12	45	9	66
		175 cm	8	43	9	60
		275 cm	7	39	14	60
		300 cm	7	36	11	54
	Total		94	223	43	360

TABLE 3 | Model 1b—Demonstrative by distance and hearer position with two levels (language and individual).

Fixed coefficients

		Coefficient	Std. error	*t*	Sig	Exp (Coefficient)	95% Confidence interval for Exp (Coefficient)	
							Lower	Upper
Medial term	Intercept	−6.642	1.1986	−5.541	< 0.001	0.001	< 0.001	0.014
	Opposite	0.063	0.2636	0.241	0.810	1.066	0.635	1.787
	Distance 300 cm	9.431	0.9063	10.405	< 0.001	12462.911	2106.910	73721.315
	Distance 275 cm	9.185	0.8745	10.503	< 0.001	9753.152	1754.807	54207.652
	Distance 175 cm	9.790	0.8272	11.836	< 0.001	17863.048	3526.983	90470.660
	Distance 150 cm	9.077	0.7851	11.563	< 0.001	8754.758	1877.356	40826.462
	Distance 50 cm	1.458	0.7261	2.008	0.045	4.297	1.034	17.851
Distal term	Intercept	−3.383	1.1417	−2.963	0.003	0.034	0.004	0.319
	Opposite	0.462	0.2475	1.867	0.062	1.587	0.977	2.580
	Distance 300 cm	8.488	0.6499	13.060	< 0.001	4855.323	1357.167	17370.124
	Distance 275 cm	8.060	0.6050	13.324	< 0.001	3166.541	966.709	10372.288
	Distance 175 cm	6.015	0.5407	11.123	< 0.001	409.463	141.781	1182.525
	Distance 150 cm	5.107	0.4721	10.818	< 0.001	165.188	65.441	416.969
	Distance 50 cm	0.746	0.3490	2.139	0.033	2.110	1.064	4.183

Reference: Proximal term, side-by-side hearer position, 25 cm distance.

TABLE 4 | Model 2d—Demonstrative by language, distance and hearer position with one level (individual).

Fixed coefficients[a]

Demonstrative		Coefficient	Std. error	*t*	Sig.	Exp (Coefficient)	95% Confidence interval for Exp (Coefficient)	
							Lower	Upper
Distal term	Intercept	−1.306	0.4852	−2.691	0.007	0.271	0.105	0.702
	Opposite	0.355	0.2124	1.670	0.095	1.426	0.940	2.163
	Distance 300 cm	3.271	0.4962	6.592	< 0.001	26.343	9.953	69.719
	Distance 275 cm	2.937	0.4589	6.400	< 0.001	18.862	7.668	46.399
	Distance 175 cm	3.269	0.4635	7.053	< 0.001	26.297	10.594	65.272
	Distance 150 cm	2.695	0.3920	6.874	< 0.001	14.799	6.860	31.923
	Distance 50 cm	0.617	0.3177	1.943	0.052	1.854	0.994	3.457
	Language	−16.820	324.8142	−0.052	0.959	49.6E-9	1.057E-284	232.5E + 267
	Language*300 cm	32.791	449.8260	0.073	0.942	174.1E + 12	< 0.001	
	Language*275 cm	33.090	449.7033	0.074	0.941	234.9E + 12	< 0.001	
	Language*175 cm	17.587	324.8152	0.054	0.957	43.4E + 6	9.242E-270	204.1E + 282
	Language*150 cm	16.974	324.8150	0.052	0.958	23.5E + 6	5.011E-270	110.5E + 282
	Language*50 cm	−0.575	459.4176	−0.001	0.999	0.563	< 0.001	
Third term	Intercept	−17.483	402.8104	−0.043	0.965	2.554E-08	< 0.001	
	Opposite	−0.007	0.2336	−0.030	0.976	0.993	0.628	1.570
	Distance 300 cm	18.624	402.8104	0.046	0.963	122.5E + 6	< 0.001	
	Distance 275 cm	18.351	402.8104	0.046	0.964	93.3E + 6	< 0.001	
	Distance 175 cm	17.447	402.8104	0.043	0.965	37.8E + 6	< 0.001	
	Distance 150 cm	16.399	402.8104	0.041	0.968	13.2E + 6	< 0.001	
	Distance 50 cm	12.284	402.8114	0.030	0.976	216.2E + 3	< 0.001	
	Language	13.421	402.8103	0.033	0.973	674.3E + 3	< 0.001	
	Language*300 cm	1.741	509.0141	0.003	0.997	5.704	< 0.001	
	Language*275 cm	2.128	508.9057	0.004	0.997	8.402	< 0.001	
	Language*175 cm	−7.771	402.8112	−0.019	0.985	< 0.001	< 0.001	
	Language*150 cm	−7.242	402.8110	−0.018	0.986	0.001	< 0.001	
	Language*50 cm	−11.716	402.8117	−0.029	0.977	8.166E-06	< 0.001	

Reference: Proximal term, side-by-side hearer position, 25 cm distance, Spanish in Spain.
Link function: Generalized logit[a].

Regression Models

For the analysis we carried out multilevel regression models which allow for the inter-related variance within all responses within a level, such as correlations within the responses of one individual, and possibly within the responses of individuals of one language compared to another. These variances are reported in the Random effect part of the model. The independent predictor variables are reported through the Fixed Effects. The models are all multinomial with LOGIT link, with the following three reference categories: the proximal term, side-by-side hearer position, and 25 cm distance.

We ran 2 models. Model 1a and 1b had language as level 1 (variety of language, i.e., Spanish Living in Norway and Spanish) and ID (individuals) as level 2. The two fixed effect predictors were position of hearer and distance. The two-way interaction of position of hearer x distance was not significant in Model 1a [$F(10, 1776) = 1.082$, $p = 0.372$], and was thus removed for Model 1b, which was the final model for the two level with interaction MLM analysis. Model 1b [$F(12, 1786) = 49.379$, $p \leq 0.001$] correctly predicted 89.6% of demonstratives, with significant fixed effects for distance [$F(2, 1786) = 59.201$, $p \leq 0.001$) and position of hearer [$F(2, 1786) = 3.426$, $p = 0.033$]. However, running model 1 showed that the amount of variance explained by language (level 1) was non-significant (Z = 0.562, $p = 0.574$ for medial and Z = 0.579, $p = 0.563$ for distal), though the variance accounted for by individuals within each language (level 2) was significant ($Z = 3.836$, $p \leq 0.001$ for medial and Z = 3.994, $p \leq 0.001$ for distal). For this reason, we amended the model to a one level model with just the variance within individuals' responses accounted for as a "level" in Model 2 (see **Table 3**).

Model 2a, 2b, 2c and 2d had language as a predictor and ID (individuals) as the only level. The three fixed effect predictors were language, position of hearer and distance. All interactions are first entered and then higher order interactions removed if not significant. The three-way interaction of language x position of hearer x distance was non-significant in Model 2a [$F(10, 1752) = 0.396$, $p = 0.949$], and was removed for Model 2b, then the non-significant two-way interaction position of hearer x distance [$F(10, 1762) = 0.781$, $p = 0.648$] was removed for Model 2c, and then the non-significant two- way interaction position of hearer x language [$F(2, 1772) = 0.573$), $p = 0.573$] was removed for Model 2d, which is the final model for the single level with interaction MLM analysis. Model 2d [$F(24, 1774) = 24.745$, $p < 0.001$ correctly predicted 86.7% of demonstratives correctly with significant fixed effects for distance [$F(10, 1774) = 16.881$, $p < 0.001$] and for the language x distance interaction [$F(10, 1774) = 21.456$], and not significant for language [$F(2, 1774) = p = 0.994$] and for position of hearer [$F(2, 1774) = 2.798$, $p = 0.061$]. The variance accounted for by level 1 (individuals within each language) was significant ($Z = 3.044$, $p = 0.002$ for medial and $Z = 3.075$, $p = 0.002$, for distal) (see **Table 4**).

In a separate model we analyzed only the data from the Spanish Living in Norway group, in order to assess the effect of time spent (i.e., exposure to the L2) in Norway on their performance. Time spent in Norway was entered as a random effect, and turned out to be highly non-significant ($p = 0.926$).

The Norwegian native speaker group was not included in the multilevel regression models due to lack of comparable number of dependent variables (two vs. three deictic terms). The descriptive data from that group, nevertheless, revealed an overwhelming use of the distal term [*den der* (that (over) there)] for all positions (689 times, 83.2%), except for the closest distances (25 and 50 cm) (139 occurrences, 16.8%), which were named by the proximal term *den her* (this here).

DISCUSSION AND FINAL REMARKS

In the current study, we expected the group of Spanish native speakers living in Norway to perform comparably to the control group of native speakers living in Spain. This was driven by theoretical accounts and hypotheses of language attrition, which is assumed to affect less robust systems first, leaving early acquired, basic and more robust systems relatively intact (Jakobson, 1941; Keijzer, 2010). This main hypothesis was not borne out. We saw a dramatic difference in the use of the three terms available in Spanish between the two groups. While the SLiS group used the three terms according to classical descriptions of the language, and previous experimental research (Coventry et al., 2008), the SLiN group saw a dramatic drop of the distal term (*aquel*), combined with an overwhelming use of the medial (third) term *ese*. The latter was used across the board for all experimental distances, and even in place of *este* for the closest object locations, with an equal number of *este* and *ese* already for the 50 cm distance. The regression analysis in Model 2d further confirmed the difference between the two groups of speakers through the significant language x distance interaction.

These results suggest that *ese* is becoming a neutral deictic term appropriate for referring to all possible locations of the referent with regard to the deictic center. This is true for Spanish native speakers who have moved to another country (Norway), which features a deictic system different from the Spanish one. Interestingly, this convergence on a two-term system, whereby the proximal term (*este*) is reserved for locations in the immediate vicinity of the speaker, and a second, neutral term (*ese*), is used to refer deictically to other and further locations beyond this one, is highly reminiscent of the results from the native Norwegian group (see also Coventry, in preparation). Two possible accounts present themselves. One possibility is that the observed change in deictic term use is the result of cross-linguistic transfer, leading to, sometimes irreversible, changes in the L1 language system, i.e., attrition (Cook, 2003; Köpke and Schmid, 2004). However, bi-directional influence of the two languages of the bilingual has been recognized in all traditions studying language learning and processing. Thus, the current results can also be attributed to the effects of bilingual language usage (Grosjean, 1992; Kroll and Bialystok, 2013). Following Schmid and Köpke (2007), we believe that the two perspectives are reconcilable and

not mutually exclusive. It is thus possible that the observed results are attributable to a bilingual system of mapping perceptual space onto the native language (Spanish), primarily reflected in language use, and as a result of daily practice of a second language. Indeed, recent studies on attrition in Spanish speakers exposed to English document that attrition effects may be partly reversible when speakers are re-immersed in the original L1 community (Chamorro et al., 2016; Chamorro and Sorace, 2019). These findings indicate that bilingual grammars are dynamic systems which reflect sensitivity to frequency of use. It may be further speculated that it is not the grammar itself that shows irreversible changes in first-generation speakers, but rather access to the grammar and the flexibility to map linguistic labels to referents in context. Since mapping between demonstrative form and contextual features which impact on deictic use requires cognitive effort, bilinguals may not always be in a position to do the appropriate mapping (Sorace, 2011, 2016, 2020). This may result in simplification and overuse of the most neutral or explicit form which fits a wider range of referential contexts, indicative of adaptive changes as a result of bilingual exposure (Sorace, 2016).

Simplification has been documented in other domains of first-generation language use. For example, the study by Tsimpli et al. (2004) provides evidence of attrition of subject pronouns in native speakers of Italian, a null-subject language after prolonged exposure to English. This study shows a selective simplification of the original system with inappropriate extension of the explicit form, in parallel with evidence from L2 speakers of such languages. Research on adult and child bilingual speakers of two null-subject languages of the same type found the same over-extension of the overt pronoun (Malgaza and Bel, 2006; Bonfieni et al., 2019).

Our results further suggest that deictic referential systems may "shrink" over time, and under pressure from bilingual language exposure, when certain perceptual distinctions are no longer systematically encoded in the respective terms. This is evidenced by diachronic changes in many Indo-European languages, whereby three-term systems evolve into two-term systems (Frei, 1944; Lyons, 1999; Manolessou, 2002; Vulchanova and Vulchanov, 2011). Interestingly, the Spanish living in Norway group appear to have converged on the medial (third) term (*ese*) as a distance-neutral term appropriate for reference to all types of locations relative to speaker, even including the peri-personal space, where the proximal term *este* is in close competition with this neutral term. Thus, at the 25 cm location, *este* was used a total of 72 times, against 48 for *ese*, while at 50 cm the two terms are already used equally often (59 vs. 60). Yilmaz and Schmid (2018) attribute the subsequent changes in L1 attrition exactly to an initial process of competition between items. Furthermore, a similar development has been attested also diachronically in the history of Bulgarian where the neutral term has come to replace the proximal one over time, subsequently becoming grammaticalized as an article (Vulchanova and Vulchanov, 2010, 2011). From a psycholinguistic and diachronic perspective, however, an open question remains whether to treat phenomena of this type as just a simplification or rather as a re-organization in the mapping of form to meaning, whether irreversible or dynamic.

Surprisingly, in a separate analysis run only on the SLiN group, time spent in Norway was highly non-significant. This finding is unexpected given the role of length of stay in host country, which is typically used as an important inclusionary criterion in attrition research. However, it is consistent with an account of deictic term use as driven by universal cognitive principles and parameters, rather than language-specific constraints and lexical encoding (Coventry et al., 2014), as well as with the changes documented in language diachrony discussed above. Furthermore, given these results, and the prevalence of two-term systems in the survey in Diessel (2005, 2013), it may be stipulated that three-term systems are less stable than two-term systems, by lexicalizing more, and more subtle distinctions.

An interesting finding in the current study is the absence of impact of position of hearer. Results for term use did not differ significantly between the two experimental conditions, and between the two groups of participants, also confirmed by the lack of significant effect of position of hearer. This result is unexpected against the semi-naturalistic performance results and account provided in Jungbluth (2003) and in Jungbluth and Da Milano (2015), where face-to-face deictic reference was dictated by a person-oriented system. Also, on that account, speakers are expected to differ as a result of the face-to-face constellation on use of the distal term *aquel*, but not on the proximal one (*este*). However, the native Speakers Living in Spain (SLiS) in our study used an equal number of distal terms between the two conditions for the relevant distance locations (275 and 300 cm). The current results contradict also the findings in Coventry et al. (2008), where position of hearer impacted on the use of the deictic terms available in Spanish, and interacted with distance. In the current study, the interaction between position of hearer and distance was non-significant, as was the interaction with language, for both groups of speakers. Given that no other differences with this earlier study of Spanish were evident in our results, and the descriptive data in both studies are highly consistent, we attribute the current finding to a methodological difference. Coventry et al. (2008) found a main effect of position of hearer only for the proximal term *este*, and an interaction with distance again only for *este*, whereby use of *este* was affected exclusively in the intermediate object positions at 100, 125, and 150cm. In the current design these positions were not named by participants, except for the 150 cm distance, and thus no data were correspondingly included in the analyses, explaining why this subtle interaction was not documented. If anything, we see a reduced use of proximal *este* in the 50 cm object location (56 vs. 41%), against an increase of *ese* (41 vs. 58%) when the hearer is seated opposite the participant, and only in the Spanish Living in Norway group, consistent with their overall preference for *ese*.

Overall, the current results indicate that peri-personal space is an important parameter in the mapping of perceptual space onto language, and are, as such, consistent with extant research and ideas on deictic demonstrative use (Coventry et al., 2014; Caldano and Coventry, 2019; Peeters et al., 2020). Thus, across both groups of Spanish participants in the study, as well as the Norwegian native group, locations closest to the speaker (25 and 50 cm), and within arm length's reach, were primarily associated with use of the respective

proximal terms. The differences between groups arose first with respect to reference to locations outside of this region. The finding that the Spanish native speakers living in Norway are converging on a relatively simpler system, based on a proximal term (*este*), and a neutral term (*ese*) which is used for all other locations, further confirms this idea. These results are consistent with, and further support, Diessel (2014) suggestion that spatial specifications are still relevant for the semantic analysis of demonstratives.

The current study fills a gap in research on deictic use under conditions of immersive exposure to a second language, and specifically, on possible changes the L1 deictic reference system can undergo under bilingual pressure.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Norwegian Data Protection Service (NSD), Comité de Ética de la Investigación (Universidad de Islas Baleares), and the School of Psychology Ethics Committee at the University of East Anglia. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MV and VV managed the data collection in Norway and prepared the manuscript. PG-F managed the data collection in Spain and provided the description of the Spanish demonstrative system. JC ran the statistical analyses. MV, VV, JC, and PG-F edited the manuscript for submission. All authors contributed to the article and approved the submitted version.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions grant agreement No 676063.

ACKNOWLEDGMENTS

We would like to thank Stefan Williams and Valeria Serrano for assistance with participant recruitment and the data collection in Norway.

REFERENCES

Anderson, S. R., and Keenan, L. E. (1985). "Deixis," in *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*, ed. T. Shopen, (Cambridge: Cambridge University Press), 259–308.

Bonfieni, M., Branigan, H., Pickering, M., and Sorace, A. (2019). Language experience modulates bilingual language control: the effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychol.* 193, 160–170. doi: 10.1016/j.actpsy.2018.11.004

de Bot, K., and Weltens, B. (1991). "Recapitulation, regression, and language loss," in *First Language Attrition*, eds H. Seliger, and R. Vago, (Cambridge: Cambridge University Press), 31–51. doi: 10.1017/cbo9780511620720.003

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.

Breunesse, M. (2020). *Demonstratives in Space and Discourse: A Synchronic and Diachronic Analysis*. PhD thesis. Jena: Friedrich-Schiller-Universität.

Caldano, M., and Coventry, K. R. (2019). Spatial demonstratives and perceptual space: to reach or not to reach? *Cognition* 191:103989. doi: 10.1016/j.cognition.2019.06.001

Chamorro, G., and Sorace, A. (2019). "The Interface Hypothesis as a framework for studying L1 attrition," in *The Oxford handbook of language attrition*, eds B. Köpke, and M. Schmid, (Oxford: Oxford University Press).

Chamorro, G., Sorace, A., and Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism Lang. Cogn.* 19, 520–532. doi: 10.1017/s1366728915000152

Clark, E. V. (1978). "From gesture to word: on the natural history of deixis in language acquisition," in *Human growth and development*, eds J. S. Bruner, and A. Garton, (Oxford: Oxford University Press), 85–120.

Cook, V. (2003). "The changing L1 in the L2 user's mind," in *Effects of the Second Language on the First*, ed. V. Cook, (Clevedon: Multilingual Matters), 1–18. doi: 10.21832/9781853596346-003

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach. Near-far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., et al. (in preparation). Semantic universals of space: spatial demonstrative systems across languages.

Diessel, H. (1999). *Demonstratives. Form, Function and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2005). "Distance contrasts in demonstratives," in *World Atlas of Language Structures*, eds M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie, (Oxford: Oxford University Press), 170–173.

Diessel, H. (2013). "Distance contrasts in demonstratives," in *The World Atlas of Language Structures Online*, eds M. Dryer, and M. Haspelmath, (Leipzig: Max Planck Institute of Evolutionary Anthropolog).

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Lang. Linguist. Compass* 8, 116–132. doi: 10.1111/lnc3.12066

Dixon, R. M. W. (2003). Demonstratives: a cross-linguistic typology. *Stud. Lang.* 27, 61–112. doi: 10.1075/sl.27.1.04dix

Faarlund, J. T., Lie, S., and Vannebo, K. I. (1997). *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.

Frei, H. (1944). Systèmes de déictiques. *Acta Linguist.* 4, 111–129. doi: 10.3917/admed.154.0111

Grosjean, F. (1992). "Another view of bilingualism," in *Cognitive Processing in Bilinguals*, ed. R. J. Harris, (Amsterdam: North Holland), 51–62. doi: 10.1016/s0166-4115(08)61487-9

Gudde, H., Griffiths, D., and Coventry, K. R. (2018). The (spatial) memory game: testing the relationship between spatial language, object knowledge, and spatial cognition. *J. Vis. Exp.* 132:56495. doi: 10.3791/56495

Halmøy, M. (2016). *The Norwegian Nominal System. A Neo-Saussurean Perspective*. Berlin: De Gruyter Mouton.

Hottenroth, P.-M. (1982). "The system of local deixis in Spanish," in *Here and There: Cross-linguistic Studies on Deixis and Demonstration*, eds J. Weissenborn, and W. Klein, (Amsterdam: John Benjamins Publishing Company).

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksell.

Johannessen, J. B. (2006). "Just any pronoun anywhere? Pronouns and "new" demonstratives in Norwegian," in *A Festschrift for Kjell Johan Sæbø: In Partial Fulfilment of the Requirements for the Celebration of his 50th Birthday*, eds A. Grønn, D. Haug, and T. Solstad, (Oslo: Solstad, Grønn & Haug), 91–106.

Jungbluth, K. (2003). "Deictics in the conversational dyad: findings in Spanish and some cross-linguistic outlines," in *Deictic Conceptualization of Space, Time and Person*, ed. L. Friedrich, (Amsterdam: John Benjamins), 13–40. doi: 10.1075/pbns.112.04jun

Jungbluth, K., and Da Milano, F. (2015). *Manual of Deixis in Romance Languages*. Berlin: Walter de Gruyter.

Keijzer, M. (2007). Last in first out?: an investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada. *Toegepaste Taalwetenschap Artikelen* 78, 131–139. doi: 10.1075/ttwia.78.12kei

Keijzer, M. (2010). The regression hypothesis as a framework for first language attrition. *Bilingualism Lang. Cogn.* 13, 9–18. doi: 10.1017/s1366728909990356

Kroll, J. F., and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *J. Cogn. Psychol.* 25:799170. doi: 10.1080/20445911.2013.799170

Köpke, B. (2007). "Language attrition at the crossroads of brain, mind, and society," in *Language Attrition. Theoretical perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert, (Amsterdam: John Benjamins), 9–37. doi: 10.1075/sibil.33.03kop

Köpke, B., and Schmid, M. S. (2004). "First language attrition: the next phase," in *First language attrition: Interdisciplinary perspectives on methodological issues*, eds M. S. Schmid, B. Köpke, M. Keijzer, and M. L. Weilemar, (Amsterdam: John Benjamins), 1–43. doi: 10.1075/sibil.28.02kop

Küntay, A. C., and Özyürek, A. (2005). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal. *J. Child Lang.* 33, 303–320. doi: 10.1017/s0305000906007380

Levinson, S. C. (2004). "Deixis and pragmatics," in *The Handbook of Pragmatics*, eds L. Horn, and G. Ward, (Oxford: Blackwell), 97–121.

Levinson, S. C. (2018). "Introduction: demonstratives: patterns in diversity," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira, (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781108333818.002

Levinson, S. C., Cutfield, S., Dunn, M. J., Enfield, N. J., and Meira, S. (2018). *Demonstratives in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Lyons, C. (1999). *Definiteness*. Cambridge: CUP.

Lyons, J. (1977). *Semantics*, Vol. 2. Cambridge: Cambridge University Press.

Malgaza, P., and Bel, A. (2006). "Null subjects at the syntax-pragmatics interface: evidence from Spanish interlanguage of Greek speakers," in *Proceedings of GASLA 2006*, eds M. G. O'Brien, C. Shea, and J. Archibald, (Somerville, MA: Cascadilla Press), 88–97.

Manolessou, I. (2002). The evolution of the demonstrative system in Greek. *J. Greek Linguist.* 2, 119–148. doi: 10.1075/jgl.2.05man

Peeters, D., Hagoort, P., and Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition* 134, 64–84. doi: 10.1016/j.cognition.2014.10.010

Peeters, D., Krahmer, E., and Maes, A. (2020). A conceptual framework for the studyof demonstrative reference. *PsyArXiv* Preprints Available online at: https://psyarxiv.com/ntydq/ (accessed April 28, 2020),

Rocca, R., Tylén, K., and Wallentin, M. (2019). *This* shoe, *that* tiger: semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLoS One* 14:e0210333. doi: 10.1371/journal.pone.0210333

Schmid, M. S., and Köpke, B. (2007). "Bilingualism and attrition," in *Language attrition: Theoretical perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert, (Amsterdam: John Benjamins), 1–8. doi: 10.1075/sibil.33.02sch

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguist. Approaches Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor

Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguist. Approaches Biling.* 6, 669–684. doi: 10.1075/lab.15055.sor

Sorace, A. (2020). L1 attrition in a wider perspective. *Second Lang. Res.* 36, 203–206. doi: 10.1177/0267658319895571

Stevens, J., and Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: an event-related potential study. *J. Neurolinguist.* 26, 31–45. doi: 10.1016/j.jneuroling.2012.02.005

Stevens, J., and Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: a cross-language study of spatial demonstratives. *J. Neurolinguist.* 30, 27–47. doi: 10.1016/j.jneuroling.2014.03.003

Tsimpli, T., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *Int. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601

Vindenes, U. (2018). Cyclic renewal of demonstratives. *Stud. Lang.* 42, 641–668. doi: 10.1075/sl.17047.vin

Vulchanova, M., and Vulchanov, V. (2010). "An article on the rise: contact-induced change and the rise and fall of N-to-D movement," in *Continuity and Change in Grammar*, eds A. Breitbarth, C. Lucas, D. Willis, and S. Watts, (Amsterdam: John Benjamins), 335–354. doi: 10.1075/la.159.16dim

Vulchanova, M., and Vulchanov, V. (2011). "An article åvolving: the case of old bulgarian," in *Grammatical Change. Origins, Nature, Outcomes*, eds A. Garrett, D. Jonas, and J. Whitman, (Oxford: Oxford University Press), 160–178. doi: 10.1093/acprof:oso/9780199582624.003.0008

Yilmaz, G., and Schmid, M. (2018). "First language attrition and bilingualism: adult speakers," in *Bilingual Cognition and Language: The State of the Science Across Its Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice, (Amsterdam: John Benjamins Publishing Company), 225–250. doi: 10.1075/sibil.54.11yil

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 December 2020
doi: 10.3389/fpsyg.2020.543549

Common Ground in Demonstrative Reference: The Case of Mano (Mande)

*Maria Khachaturyan**

Helsinki University Humanities Programme, University of Helsinki, Helsinki, Finland

That demonstratives often have endophoric functions marking referents outside the physical space of interaction but accessible through cognition, especially memory, is well-known. These functions are often classified as independent from exophoric ones and are typically seen as secondary with respect to spatial deixis. However, data from multiple languages show that cognitive access to referents functions alongside of perceptual access, including vision. Cognitive access is enabled by prior interactions and prior familiarity with the referents. As a result of such interactions, the interlocutors share a great deal of knowledge about the referents, which facilitates reference to objects in the interactive field. The centrality of common ground in reference to an object at the interactive scene challenges the often assumed classification of demonstrative reference into exophoric and endophoric. I illustrate this idea throughout the paper by using first-hand data from Mano, a Mande language of Guinea. Adding another argument in favor of viewing demonstrative reference as a social, interactive process, the Mano data push the idea of salience of non-spatial parameters further and emphasizes the importance of short and long-term interactional history and cultural knowledge both for the choice of demonstratives in exophoric reference and for the structuring of the demonstrative paradigm.

Keywords: deixis, common ground, reference, ethnography, interaction, interactional history, corpus, Mande languages

OPEN ACCESS

Edited by:
Harmen Bart Gudde,
University of East Anglia,
United Kingdom

Reviewed by:
Terra Edwards,
Saint Louis University, United States
Sonja Gipper,
University of Cologne, Germany

***Correspondence:**
Maria Khachaturyan
maria.khachaturyan@helsinki.fi

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *17 March 2020*
Accepted: *23 October 2020*
Published: *17 December 2020*

Citation:
Khachaturyan M (2020) Common Ground in Demonstrative Reference: The Case of Mano (Mande). Front. Psychol. 11:543549. doi: 10.3389/fpsyg.2020.543549

INTRODUCTION

It is generally accepted that the exophoric reference to objects in the physical space of interaction is more basic than the endophoric reference used to track referents in discourse or in reference to discourse itself (Diessel, 1999; Levinson et al., 2018). Much scholarly attention has been directed to the exophoric use using a targeted elicitation methodology (i.e., Coventry et al., 2014; Wilkins, 2018). In such studies, interaction participants are only given the very lean and abstract characteristics of "speaker" and "addressee." Yet such an approach obscures the fact that participants in

Abbreviations: ADR, addressee; ASSOC, associative plural; ATT, attention drawing; BKGR, backgrounding marker; COND, conditional; CONJ, conjunctive; COP, copula; DEM, demonstrative; DIST, distal; EMPH, emphatic; EXI, existential; FR, French; GER, gerund; H, high tone; INDEF, indefinite; INF, infinitive; INJ, interjection; INT, intensifier; IPFV, imperfective; JNT, conjoint; NEG, negative; PL, plural; POSS, possessive; PRF, perfect; PROH, prohibitive; PST, past; REFL, reflexive; REL, relativizer; SBJV, subjunctive; SG, singular.

real-life speech events come to any particular interaction with a set of expectations and background knowledge. This fact challenges the standard classification of demonstrative functions that contrasts exophoric and endophoric uses (on that point, see also Agha, 1996)[1]. Indeed, a referent physically present at the interactive scene often belongs both to the deictic sphere of interaction and to the non-deictic sphere of common ground which includes the participants' mutual knowledge, beliefs and suppositions (Clark et al., 1983). As argued by Coventry et al. (2014, p. 49), "the perception of space is not constrained solely by the characteristics of the physical environment, but is mediated by high-level knowledge about the objects being perceived." In this paper, by using first-hand, to a large extent previously unpublished data from Mano (Mande, Guinea), I extend this important conclusion by arguing that such mixed endophoric-exophoric uses cannot be accounted for unless one sees interactants as social actors and interaction as loaded with history (Hanks, 1990, 2005).

Let us compare two examples from Mano. In (1) the speaker is sitting side by side with the addressee on a sofa. The addressee is reading a book, occasionally using sticky notes to markup pages. The package of sticky notes referred to in (1) by using the demonstrative *tɔ́ɔ̃* is placed between the interlocutors. It has, perhaps, never been seen before in the household in question (it is not named but is referred to by a 3sg pronoun) and has not been discussed previously. The interlocutors were not attending the referent in a joint activity: in fact, the addressee's attention is focused on the book rather than on the sticky notes. For this reason the speaker makes a pointing gesture, which helps her secure the addressee's visual attention to the referent with the first attempt.

(1)
áyē *tɔ́ɔ̃* *lɛ́ɛ̀* *lō* ɲɛ̄-ɛ̀ ɲɛ́ɛ̀?
3SG.EMPH **DEM.ATT** 3SG.IPFV go:IPFV finish-GER isn't.it

"This one is going to finish, right?" [Mano, own fieldnotes].

In (2) a young street vendor is addressing two persons passing by about 1.5 m away on a motorbike. The two motorbike riders are engaged in a conversation, so they were not attending the referent before it was mentioned. The vendor is suggesting they buy popcorn balls, a very popular snack found on many corners of Mano villages and towns, something the interlocutors are surely used to and expect to see. For this reason no visual attention is required. Moreover, precisely because she expects the addressees to be familiar with the referent (which is also marked with the possessive pronoun kà "your") and because it is difficult to draw the addressees' visual attention, the speaker chooses the *wɛ̄* demonstrative.

(2)
kà ***kpɛ̀ɛ́*** ***wɛ̄*** *ɓ́* *wɛ̄!*
2PL.POSS **corn** **DEM1** buy DEM1

"Your popcorn (lit.: this corn of yours), buy it!" [Mano, own fieldnotes].

The choice between the demonstratives *tɔ́ɔ̃* and *wɛ̄* is not motivated by the differences in physical distance between the speaker and the object: in both cases the object is located within the arm's reach (on a sofa next to the speaker or on a stand). The position in the interactional sequence also does not appear to matter: both (1) and (2) concern first mentions of referents in a given interaction. Instead, the choice is motivated by the degree of prior familiarity with and the existence of shared knowledge about the referent. The Mano exophoric demonstratives tɔ̀ɔ̃ (and *dìɪ̰̄*) indicate that additional visual attention coordination is required in order to identify the referent and for this reason they are typically accompanied by deictic pointing, as in (1). In contrast, the demonstratives *wɛ̄* (and *yā*), which may be used exophorically, as in (2), or endophorically, do not encode such a requirement. Although they may be used for establishing joint visual attention, they are also (and even more frequently) used in situations where such visual attention coordination is already established or where it is impossible or unnecessary, when the speaker has reasons to believe that the referent is salient enough for the purposes at hand. In these situations, referent identification is primarily based on the interlocutors' mutual knowledge of the referent, which is derived from prior interactions or broader cultural knowledge, as in (2).

INTERACTIONAL PERSPECTIVE ON DEMONSTRATIVE REFERENCE

Demonstratives are linguistic expressions serving "to coordinate the interlocutors' joint focus of attention" on a reference object (Diessel, 2006, p. 464). The physical co-presence of interlocutors is understood as the prototypical property of interactional settings, and coordination of attention on objects in space is considered the primary function of demonstratives, primordial in phylogeny (Tomasello, 2008), best described in the literature (Levinson et al., 2018) and also seen as the source of further functional development and grammaticalization of non-spatial, endophoric functions (anaphora, discourse deixis, recognitional function, see Diessel, 1999). Even synchronically, anaphora is sometimes seen as a metaphoric extension of space (Anderson and Keenan, 1985).

A considerable volume of ethnographic and experimental research on demonstratives has shown that spatial distinctions cannot be reduced to mere physical distance. Natural, artificial, and culturally imposed boundaries (rivers, valleys, island boundaries; walls; family spaces, boundaries of the village) also contribute to the conceptualization of the proximity vs. distance contrast (as in *this village*, which can cover quite an extensive space, Margetts, 2018, see also Hanks, 1990; Enfield, 2003). In

[1]Diessel (1999) divides endophoric demonstrative use into four subfunctions: anaphora (*I saw a guy yesterday, that guy was clearly following me*), cataphora (*let me tell you this simple trick*), recognitional (as in Internet memes: *that feeling when*), and discourse deictic [*the moral of that story* (just been told)]. Levinson (2004) proposes a different classification, dividing demonstratives into two use-types, deictic (including exophoric, but also discourse deictic) and non-deictic, which includes recognitional, cataphoric and anaphoric uses plus the emphatic use (*that son of a bitch!*).

other words, in formulation and interpretation of demonstrative reference, the speakers engage with a great deal of knowledge about the social world and the space they inhabit is not merely physical but interactional. Thus, in Lao, the choice of demonstrative in a two-part system is determined by the position of the referent with respect to the interactionally defined here-space of the speaker (Enfield, 2003).

In addition to the knowledge about the space of interaction, in real-world interactional settings the referents themselves do not appear as timeless artifacts, as is often the case in experimental studies, but are loaded with history. On one level, the establishment of joint attention and referent identification are parts of an interactive process (Hindmarsh and Heath, 2000; Eriksson, 2009; Etelämäki, 2009). Several languages have been attested where demonstratives encode different stages in that process. Thus, in addition to the speaker-centric spatial distinction between a proximal and a distal demonstrative, another core semantic distinction in the Turkish demonstrative paradigm is whether the referent is in the addressee's focus of attention (Küntay and Özyürek, 2006). Such cases concern the history of a given interaction within several interactional turns (see section "Common Ground and Interactional History").

On a further level, as argued by Clark et al. (1983), reference resolution can be predicated on mutual knowledge built in long-term local histories of interaction, stretching beyond the interaction at hand. Thus, as the authors argue, the local history of interaction plays a role when a speaker points to a group of men and says "That is what George will look like very soon": the reference is interpretable upon a condition that something has been said between the two interlocutors about George, e.g., that he is gaining or losing too much weight, and that one of the men in the group looks overweight or underweight. Thus, meaning and reference are established in time, as part of incremental building of common ground within and across interactional encounters (Deppermann, 2018; Harjunpää et al., 2021).

A further level of complexity arises when interaction participants are not seen as merely speaker and addressee engaged in interaction, but as cultural beings who, by virtue of their membership in a particular collectivity (community of practice, speech community), operate with a great deal of common knowledge. Such culturally shared knowledge, which is (presumably) available to most if not all members of a given collectivity, became apparent as a determiner in an experiment discussed in Clark et al. (1983), where the subjects were shown a photograph of President Reagan and David Stockman, then the director of the Office of Management and Budget. When asked, "You know who this man is, don't you?" the overwhelming majority of subjects understood "this man" to be Reagan, not Stockman, who was much less known. Furthermore, interaction participants are also social actors engaged in specific partially scripted social activities. Such activities presuppose a specific optic through which some objects are seen and a specific way they are referred to Hindmarsh and Heath (2000). In particular, even when objects are referred to for the first time in a given interaction, they may be partially anticipated. For example, in Yucatec Maya, a shaman wrapping up a medicine for his patient may refer to it using a non-immediate deictic, despite the fact that the referent is immediately accessible, which would in other contexts warrant the use of an immediate deictic. This use, as Hanks (2005) explains, is in part because the referent, a medicine, is presented as mutually known and expected in the context of shamanic practice and such uses are typically covered by a non-immediate deictic.

A given referential act is thus part of different mutually constituting levels. On the one hand, it belongs to the level of the temporarily unfolding interactional process of reference resolution, which in turn is part of a longer-term interactional history involving the same communicating individuals. On the other hand, these individuals participate in communication not only as communicating agents, but also as social agents occupying different positions in social fields (Bourdieu, 1990) routinely dealing with and talking about specific kinds of artifacts. In other words, any given interaction is *embedded in* (Hanks, 2005), and is an instantiation of, a social field whereby the properties and the relationships between the positions in the social field are projected onto a very general structure of the interactional space. Thus, the properties and relationships in the triad shaman–patient–medicine is projected onto a given triad of speaker, addressee and referent and motivates the choice of the referential expression. Both kinds of embedding, embedding in interactional history and in social fields, transform referents in the interactive space from physical to social artifacts known to the interactants: while short- and long-term interactional history provides situated knowledge to given participants, the social field provides more general knowledge available to wider collectivities participating in the same social field.

The following discussion is organized in the following way. Section "Mano Demonstrative System" presents four Mano demonstratives as they are used in naturally occurring referential acts. Section " Demonstratives ***kpɛ́í*** and *dı̰ā̰*" offers some basic morphosyntactic information on the four Mano demonstratives. Section "Demonstratives wɛ̄ and yā" presents the functions of the demonstratives *tɔ̀ɔ̄* and *tɛ́ dı̰ā̰* in more detail. Section "Demonstratives wɛ̄ And yā" is dedicated to the demonstratives *wɛ̄* and *yā*, their endophoric (discourse reference, anaphora and recognitional function) and exophoric functions, as well as the contrasts between the two markers. Section Semantics and Pragmatics of Mano Demonstratives is an interim summary where I disentangle the semantic and pragmatic components of the meaning of Mano demonstratives. Section "Common Ground and Reference Resolution in Interaction" presents the demonstratives *wɛ̄* and *yā* in a broader interactional context which provides the interaction participants with knowledge about referents and motivates the use of *wɛ̄* and *yā*. Section "Common Ground and Interactional History" illustrates the interactional process of referent identification and the role of idiosyncratic mutual knowledge built in local interactions, in particular, between friends and family members. Section "Common Ground and Social Fields" deals with knowledge proper to specific social fields, namely the domain of ritual practice ("traditional" and Christian) and specificities of demonstrative reference proper to these domains. I discuss the findings in section "Discussion" and make my conclusions in section "Conclusion."

MANO DEMONSTRATIVE SYSTEM

Mano (*mááwè*) is a Southern Mande language spoken by 305,000 people in Liberia (Ethnologue[2]) and, according to different estimations, by 66,000 (Guinean census performed in 2014, Bah and Bangoura, 2017) and 95,000 (Ethnologue) in Guinea. A grammar of Mano can be found in Khachaturyan (2015), and for a typological portrait of the language, see Khachaturyan (2020a). The demonstrative system is subject to dialectal variation; for a preliminary account, see Khachaturyan (2018a). Despite the widespread multilingualism, Mano is well-preserved and well-transmitted to children; (quasi)-monolingual repertoires have also been attested (Khachaturyan and Konoshenko, forthcoming).

Mano, just like other Mande languages, is a largely isolating language. It has a fixed S-Aux-O-V-X word order, where Aux is an auxiliary inflected for person and number and agreeing with the subject, and X is any post-verbal argument expressed by a postpositional phrase. With three tonal levels, it has quite a large number of tonal morphemes but a relatively small inventory of segmental morphemes. Thus, besides very minimal derivational morphology, the only two inflectional nominal forms are the low-tone construct form (CSTR), appearing on the head of noun phrases with specific kinds of preposed dependents, and a high-tone form (H) used with demonstratives. Definiteness is not grammaticalized in the language, and although certain grammatical markers take on the functions of marking definiteness, they are never obligatory (Khachaturyan, 2020b).

This paper is based on the data from the Central Guinean dialect of Mano, Maa (máá), drawn from a corpus of recordings of spontaneous speech collected by the author during more than 15 months of fieldwork. All the examples taken from the Mano oral corpus are marked with MOC. Some examples come from fieldnotes [fieldnotes]: these are utterances that I overheard and noted and then asked the consultant to comment on them and, if necessary, correct. A minor fraction of examples are elicited (el.). All elicited utterances were contextualized and discussed with the primary language consultant. Whenever applicable, square brackets [] indicate exact discourse context preceding or following the utterance in question, parenthesis () indicate a summary of the preceding or following context or provide other textual commentaries. No systematic questionnaire study of exophoric use has been conducted, which represents a major limitation of the present study. However, since the focus of the paper is the role of common ground in reference resolution, some of which is acquired in interaction between specific individuals (inhabitants of the same village, or family members), the observational data are adequate for the analysis.

The Maa dialect of Mano (I will use Mano as a shortcut in the subsequent discussion) has four adnominal demonstratives used for exophoric reference: *tɔ́ɔ̃*, *dɪ̀ɪ̰̀*, *wɛ̄*∼*ɓɛ̄*∼wāā and *yā*∼*ā*∼*yāā*. Pronominal demonstratives are formed by adding *tɛ́* or *yé* (or only *yé*, in case of *tɔ́ɔ̃*) to the demonstrative stem: *tɔ́ɔ̃*∼*yétɔ́ɔ̃*, *tɛ́dɪ̀ɪ̰̀*, *tɛ́ā*∼*yéā* and *tɛ́wɛ̄*∼*yéwɛ̄*. They are assumed to be extensionally equivalent to the respective pronominal demonstratives and are occasionally mentioned in the paper (ex. 16). The demonstrative adverbs used in the language are: *zèè̄* "here," *dɪ̀ɪ̰̀* "there, distal" (which *dɪ̀ɪ̰̀* derives from), and *yī* "there, anaphoric." They are not discussed in this article.

The demonstrative *dɪ̀ɪ̰̀* is obligatorily preceded by the marker of attention alignment *tɛ́*∼*lɛ́*. For the demonstratives *wɛ̄* and *yā* the marker is optional. The demonstrative *tɔ́ɔ̃* does not allow the marker of attention alignment, so combinations like *tɛ́ tɔ́ɔ̃* or *lɛ́ tɔ́ɔ̃* do not occur in the corpus and are disallowed in elicitation. This is most likely because it historically derives from a fusion between *tɛ́* and the demonstrative *wɛ̄*[3]. As is shown in the discussion of the semantics and pragmatics of these markers, the additional attention alignment effort is what distinguishes the typical contexts of use of the demonstratives *tɔ́ɔ̃* and *dɪ̀ɪ̰̀*, on the one hand, and *wɛ̄* and *yā*, on the other.

Mano also has the demonstratives *kílíwɛ̄* and *kílíā*, which are used exclusively for reference tracking. Given their limited scope and the fact that they are not used for exophoric reference, I will not discuss them further. In addition, Mano has a marker of bridging, *à*, which, in contrast with the demonstratives, is situated prenominally and is not part of the demonstrative paradigm (Khachaturyan, 2020b).

Table 1 presents the Mano demonstrative forms. The five forms (excluding the free variants) differ in the domain of use (exophoric only) for *tɔ́ɔ̃* and *dɪ̀ɪ̰̀*; exophoric and endophoric (and simultaneously exophoric and endophoric) for *wɛ̄* and *yā*; and endophoric only for *kílíɓɛ̄* and *kílíā*. Further contrasts between the two pairs of forms, *tɔ́ɔ̃* and *dɪ̀ɪ̰̀*, on the one hand, and *wɛ̄* and *yā*, on the other, will be discussed in the following sections. The contrast between *kílíɓɛ̄* and *kílíā* should be an object of further investigation.

Demonstratives *tɔ́ɔ̃* and *dɪ̀ɪ̰̀*

The demonstrative *tɔ́ɔ̃* is used exclusively in the exophoric function to refer to objects present at the interactive scene; the act of reference is typically accompanied by a pointing gesture. In (3), the speaker is telling about his motorbike accident and is showing schematically on the ground where he was when the car hit him and how the car approached him.

(3)
*bon, à gà dōó gbɛ̀-ɛ̀ lɛ̄ **lɛ̀** tɔ́ɔ̃ mɔ̀.*
well[FR] 3SG leg one put-GER COP **place** **DEM.ATT** on

[3]Fusion is commonly present in Mano, especially in fast speech. However, *tɔ́ɔ̃* is used even in very slow and articulate speech; moreover, the fusion resulting in change of vowel quality (*tɛ́* + *wɛ̄* = *tɔ́ɔ̃*) is not so frequently attested in Mano. For these reasons, I consider *tɔ́ɔ̃* to be a separate demonstrative marker, rather than a combination of *tɛ́* and *wɛ̄*.

TABLE 1 | Adnominal demonstratives in Mano.

Exophoric	*tɔ̀ɔ̄*
	dɪ̀ɪ̰̀
Exophoric/endophoric	*wɛ̄* ∼ *ɓɛ̄*
	yā∼ *ā*∼ *yāā*
Endophoric (anaphora)	*kíllíɓɛ̄*
	kíllíā

[2]https://www.ethnologue.com/language/mev/23

"[Here is the asphalted part of the road.] So one of its wheels is here (pointing to the ground, at the edge of the asphalted part). [He is behind me, he looks like he's going to stop (at the side road)]" (MOC).

The demonstrative is speaker-anchored, which is evidenced by situations where the referent is invisible and inaccessible to the interlocutor, being relatively far away and in a different room (4).

(4)

mīī	***tɔ́ɔ̄***	*à*	*yímɛ̀*	*pɛ́nɛ̄ɛ̄*!
person	**DEM.ATT**	3SG	beat	today

(Referring to his son who has just entered the room where he is sitting a man shouts outside to his wife:) "This guy (the boy), scold him today!" [fieldnotes].

The proximity of the referent to the speaker is flexible and not limited to the peripersonal space: while in (3), the object is within arm's reach, in (4) it is about two meters away. My Mano interlocutors often introduce new people to me by pointing and saying *mīī* tɔ́ɔ̄ "this person" is such and such. The persons introduced can be at a considerable distance from me. What matters is that they are clearly visible and easily identifiable.

The demonstrative *dɪ̀ā̱* is used very rarely and I have only a few instances in my notes. It derives from a fusion of the deictic adverb *dɪ̀ī̱* "there" with the marker *ā*[4]. Similarly to tɔ́ɔ̄, the demonstrative *dɪ̀ā̱* is used in the exophoric function, typically with a pointing gesture. While the preferred situation is where the object is visible, *dɪ̀ā̱* can occur with invisible objects that the speaker can locate with certainty (5, 6).

(5)

là	***sɔ̄***	***lɛ́***	***dɪ̀ā̱,***	*kà*	*sí*
3SG.POSS	**cloth**	**ATT**	**DEM.ATT.DIST**	2PL.SBJV	take
ká	*nū*	*à*	*ká*!		
2PL.CONJ	come	3SG	with		

(The speaker is sitting on the floor, pointing to a basket on the opposite side of the room, about three meters away, asking her grandchild to bring her the clothes of another grandchild). "Those clothes of his, you (pl.) take them and bring them!" [fieldnotes].

(6)

mɛ́ŋ̀	***tɛ́***	***ē***	***dɪ̀ā̱,***	*lɛ̄*
something	**ATT**	**3SG.REFL**	**DEM.ATT.DIST**	3SG.EXI
ké,	*kàkò*	*lō*	*pɛ́nɛ̄ɛ̄,*	
like.this	1PL.SBJV	go:IPFV	today	
kó	*ló*	*à*	*vùò*	*ɓō-à.*
1PL.CONJ	go	3SG	greeting	do-GER

[4]Given that *dɪ̀ā̱* is always used with a marker of attention alignment, which is also used to introduced relativized NPs (Khachaturyan and Ozerov, in preparation) and that *ā* is used to frame the right edge of relative clauses, *dɪ̀ā̱* is interpretable both as an adnominal determiner "that X" and as part of a complex utterance of the type "X that is there."

(An imaginary discussion between two inhabitants of the same village). "That guy over there (in the village, distance undetermined, may be visible, preferred interpretation, or invisible) is such (in such a state), let's go today, let's go and greet him" [MOC].

For both tɔ́ɔ̄ and *dɪ̀ā̱* the distance between the speaker and the referent can vary. There is an overlap in the distance measures in tɔ́ɔ̄ and *dɪ̀ā̱*: in (4) with tɔ́ɔ̄ and (5) with *dɪ̀ā̱* the distance is roughly the same. See also (7a), (7b) illustrating that both demonstratives are acceptable in certain contexts.

(7)

a.

yɔ̄	***tɔ́ɔ̄***	*lɛ̄*	*nɛ́ŋ̀n ɛ̀ŋ̀.*
wine	**DEM.ATT**	3SG.EXI	sweet

Situation A (The interlocutors sit one facing the other on opposite sides of the room, about 3 m away. The speaker is holding a can of beer in his hand). "This beer is good."

Situation B (Same as above. The addressee is holding a can of beer in his hand. The speaker points to it and says). "This beer is good."

Situation C (The interlocutors sit side by side. Either of them holds a can of beer in his hand). "This beer is good."

b.

yɔ̄	***lɛ́***	***dɪ̀ā̱***	*lɛ̄*	*nɛ́ŋ̀n ɛ̀ŋ̀.*
wine	**ATT**	**DEM.ATT.DIST**	3SG.EXI	sweet

Situation A (The interlocutors sit one facing the other on opposite sides of the room, about 3 m away. The addressee is holding a can of beer in his hand). "That beer is good."

Situation B (The interlocutors sit side by side, the beer is located on the other side of the room, about 3 m away). "That beer is good."

*The utterance is ungrammatical in a situation where the speaker is holding a can of beer in his hand [el.].

Just like tɔ́ɔ̄, *dɪ̀ā̱* is a speaker-centered marker, since the location of the addressee does not affect its use (7a, 7b). In contrast with tɔ́ɔ̄, *dɪ̀ā̱* is never used to refer to objects in the peripersonal space (7b) and can be used with objects significantly further away from the speaker (6). Crucially, the spatial setting, the ongoing activity and the purpose of pointing may matter (although more examples are needed to confirm this). In (5), with *dɪ̀ā̱*, the object is located on the other side of the room. The speaker is sitting on the floor, the object is clearly out of reach and she is instructing her granddaughter to fetch the object—clothes—for her so that she could dress her newborn grandson. Thus, *dɪ̀ā̱* is used with objects which are not immediately accessible to the speaker by being outside of her engagement area defined as "the place which is, at moment t, the conceived site of a person's currently dominant manual and attentional engagement" (Enfield, 2003, p. 89). In contrast, tɔ́ɔ̄ is neutral in that regard: it can be used with referents both within (3, 7a, Situation A, see also 26.1 below) and outside (4, 7b, Situation B) the engagement area.

Demonstratives wɛ̄ and yā

The demonstrative *wɛ̄*, which also has the variants ɓɛ̄ and *wāā*, and *yā*, which has the variants *ā* and *yāā*, can be

used in all functions suggested by Himmelmann (1996) and Diessel (1999). They are very common in endophoric functions: referring to discourse itself (section "Discourse Reference"), reference tracking and anaphora (section "Anaphora"), or the recognitional function where the speaker assumes that the referent is identifiable for the interlocutor without prior mention (section "Recognitional Function")[5]. Both demonstratives can also be used in the exophoric function, referring to objects present at the interactive scene (section "Exophoric Function: Indexing Familiarity"). They fulfill very similar functions to each other, and are very frequently interchangeable; I will gloss them as DEM1 and DEM2, respectively. I will begin with endophoric functions and then continue with the exophoric one, in which they contrast not only with each other, but also with the demonstratives *tɔ́ɔ̃* and *dḭ̀ā̰* analyzed in the previous section.

Discourse Reference

The discourse referential function is not so frequently found in texts, but both demonstratives can be used in that function. The choice between them seems to be a matter of personal and/or dialectal preferences.

(8)

ɓá		*gèá*	*ɓà*	*né*	*lɛ̀ɛ̄,*
2SG.CONJ>3SG		say.COND	2SG.POSS	child	ADR
ŋwɔ́	*yé*	*tɔ́ɔ̃*	*kɛ̄*		
problem	REL	DEM.ATT	do		
é	*ī*	***wéè***	***yā***	*ɓɛ̀lɛ̄*	*yà*
3SG.CONJ	2SG	**speech**	**DEM2**	respect	place
ā,	*kɛ̄*	*āà*	*ī*	*kɛ̄*	*gbɔ́kò.*
BKGR	then	3SG.PRF	2SG	do	big

"If you say to your child: do this thing so that he pays respect to **that speech** of yours, (if he does so) then (it means that) he has honored you" [MOC].

(9)

mīī	*lɛ́*	*ā*	*gèē*	*ī*	*lɛ̀ɛ̄?*
person	ATT	3SG.PST>3SG	say	2SG	ADR
ɛ́ɛ́	***ŋwɔ́***	***wɛ̄***	*ɓā*	*sí*	*ī*
and	**problem**	**DEM1**	2SG.PST>3SG	take	2SG
dìè	*gé?*				
INT	stomach				

"[Pons Pilate said to Jesus: Man, is it you who are the king of Jews? And then he asked] Was it someone who told it to you? Or did you invent **that issue** (that you are the king of Jews) yourself (lit.: took **that problem** from your own stomach)?" [MOC].

Anaphora

In the anaphoric function, both demonstratives are widely used and seem to be interchangeable. *Yā* and *wɛ̄* were attested in all speech genres available in our corpus, both monological ones (folktales, 10 and 11) or conversations (26).

[5]My corpus lacks examples of the cataphoric reference fulfilled with demonstratives.

(10)

kélé	*nì*	*lɛ́*	*āà*	*dà*	*ɓɛ̄,*
shed	ASSOC	ATT	3SG.PRF	fall	BKGR
tó	***kélé***	*ɓɛ̄*	*ɓàá*	*là*	*kɛ̄.*
so	**shed**	**DEM1**	2SG.NEG>3SG	surface	do

"Even the shed that has fallen, **that shed**, you don't repair its roof" [MOC].

Example (11) is taken from a story about three hunters. The prior mention of the same referent with a 3pl pronoun occurred in the preceding clause.

(11)

ō	***yààkā***	***yā***	*wáà*	*lò*	*wà*	*lúú*
3PL	**three**	**DEM2**	3PL.JNT	go:JNT	3PL.POSS	bush
píé	*kɛ̄-è.*					
to	do-GER					

"[The story I want to tell concerns three hunters... Every month they go and hunt together very well]. **The three of them**, they went hunting" [MOC].

Recognitional Function

Both *yā* and *wɛ̄* are particularly common in the recognitional function when they are used to refer to objects not present at the interactive scene, but are accessible via the common ground of the interlocutors (on that function in Mano, see Khachaturyan, 2019). Thus in example (12) both *tòò* "tomorrow" and *ɓū* "rice" refer to entities made recognizable by a prior arrangement. "Everybody knows that I have to go tomorrow to my field to work," the speaker told me when I asked her to comment on her usage of the demonstrative in (12).

(12)

tòò	***yā,***	*kóò*	*lō*	***ɓú***	***yā***
tomorrow	**DEM2**	1PL.IPFV	go:IPFV	**rice:H**	**DEM2**
mɛ̀-è.					
beat-GER					

"**That tomorrow**, we will beat **that rice**" [fieldnotes].

(13)

ɓà	***ordinateur***	***wɛ̄,***	*à*	*sí*	*wɛ̄!*
2SG.POSS	**laptop[FR]**	**DEM1**	3SG	take	DEM1

(An imagined conversation, where the speaker advises his addressee not to forget to take her laptop on the trip). "**This laptop of yours**, take it!" [el].

Exophoric Function: Indexing Familiarity

In the corpus *wɛ̄* does not seem to show any clear preference. Example (2 above) was used with a referent close by, about 1.5 m away, while example (14) was used with a referent further away.

(14)

ɓà	***lɛ̀***	***wɛ̄***	*gɛ̀*	*è*
2SG.POSS	**field**	**DEM1**	COP.DEICT	3SG.SBJV
lō	*wɛ̄.*			
go:IPFV	EXI			

(An old man is showing, with a pointing gesture, to his daughter-in-law the placement and the direction of expanse of a field that he offered her and her husband). "Here is **this field of yours** (stretching to some 100 meters), it goes like this" [fieldnotes].

In (15), a woman instructs her brother-in-law to burn the feathers of a duck she is going to cook. The fireplace is some eight meters away from where she is sitting and is hidden behind a shed. I did not take a proper note of the position of her interlocutor.

(15)

Pèé	*í*	ɲɛ̄-ɛ́	*gbāā*,
P.N.	2SG.CONJ	finish:COND-COND	now>BKGR
í	***tié***	***yā***	*mɛ́ŋ̀*
2SG.CONJ	**fire**	**DEM2**	something
kɛ̄,	*ɓá*	*mɛ́ŋ̀*	*kɛ̄!*
do	2SG.CONJ>3SG	something	do

"Pe, when you have finished, do the thing with **that fire**, do the thing!" [fieldnotes].

In (16) the speaker is sitting at a table with his friend, eating dinner (with hands, as is customary among Mano). His wife approaches him from behind asking whether he has seen the charger to her phone, which she is holding in her hand. His response is given in (16). His hands are busy with eating, but he does not even need to point, he merely takes a quick look at her phone as he knows the model very well.

(16)

non-non-non,	*à*	*dò*	*wɛ́ī.*	***tɛ́ā***
no[FR]	3SG	INDEF	NEG.COP>there	**DEM2**
là	*chargeur*	*dò*	*wɔ́*	
3SG.POSS	charger[FR]	INDEF	COP.NEG	
kō	*kɛ̀lɛ̀*	*zèē.*		
1PL	hand	here		

"[Do you(pl.) have a charger here?]" "No, we don't. That (thing), we don't have its charger here."

The demonstratives *wɛ̄* and *yā* may be used when either the speaker or both interlocutors do not see the referent. In (17), the speaker is riding on a motorcycle with the addressee and reminds his addressee to take the laptop, among other things, from a charging station, which is still out of sight.

(17)

ordinateur	*nì*	*wɛ̄!*
laptop[FR]	ASSOC	DEM1

"The laptop (we are approaching the charging station)!" [fieldnotes].

Similarly, when I was discussing example (4) with my language consultant and asked what demonstrative form would be chosen if the boy were outside the house with his mother and the speaker inside, the consultant suggested the demonstrative *yā*, instead of *tɔ́ɔ̃* (18).

(18)

míí	***yā*,**	*à*	*yímɛ̀*	*pɛ́nɛ̄ɛ̄!*
person:H	**DEM2**	3SG	beat	today

(A boy went into a puddle and came home dirty. He is outside the house with his mother, whom the father is addressing from inside the house). "That guy, scold him today!" [el].

Both *wɛ̄* and *yā* can be used to attract the addressee's attention to the referents which were not discussed in the prior discourse (2). And yet, in contrast with *tɔ́ɔ̃* and *dɪ̰ā́*, typically the speaker expects some existing familiarity with the referents, even when the demonstratives are used exophorically. In (15) the instruction of the speaker is very vague, "do the thing with the fire," which means that they had already discussed the issue or that the addressee is used to those kinds of chores. In (14), there had clearly been some prior discussion of the field in the family. The old man is just showing his daughter-in-law where the field is before they start some bush clearing work. In (2), the speaker is a street vendor selling a very widespread Mano snack—popcorn balls. Prior familiarity with the object is what allows *wɛ̄* and *yā* to be used with invisible objects or objects to which it is difficult to draw the interlocutor's attention if (s)he is busy with other things: such as riding a motorbike (2). Likewise, the speaker may be limited in her or his capacity to clearly point: because she is cooking (15) or eating (16), but pointing is typically not essential in reference retrieval. The demonstratives *wɛ̄* and *yā* are especially common in the anaphoric and recognitional functions, which rely on the cognitive accessibility of the referents alone without any clues from the physical context.

wɛ̄ Vs. yā

In the real-life examples provided above, there is no clear tendency for the distance between the referent framed with the demonstratives *wɛ̄* and *yā* and the deictic center. In elicitation, however, objects framed with *wɛ̄* are presented as close to the speaker, while objects framed with *yā* are presented as further away. In (19a), repeated from 17, the speaker reminds his addressee to take the laptop, among other things, from a charging station. They were approaching it on a motorcycle and were already rather close, at the entrance to the town, so the speaker used *wɛ̄*. A contrasting example (19b), which would have been used had they been further away on the road, is with *yā.*

(19)

a.

	ordinateur	*nì*	*wɛ̄!*
	laptop[FR]	ASSOC	DEM1

"The laptop (we are approaching the charging station!)" [fieldnotes].

b.

	ordinateur	*nì*	*yā!*
	laptop[FR]	ASSOC	DEM2

"The laptop! (don't forget to pick it up when we pass by)" [el].

A similar contrast in the degree of familiarity may also affect the use of *wɛ̄* and *yā*. In (20), the demonstrative *wāā* (a variant of *wɛ̄*) is used to refer to a woman that the addressee has just met for the first time at the local hospital, so she is highly salient in the context. In contrast, the woman's husband, whom the addressee has never met but whose existence she may very well infer, given that the woman in question had just given birth to a child, is framed with the demonstrative *yā*. Both referents are out of sight and were not talked about in the prior conversation, but the woman is more familiar to the addressee than her husband. Note that here the speaker takes the addressee's perspective in evaluating the referent's relative familiarity.

(20)

léé	***wāā***,	*léé*	*ɓā*	*gɛ̀*
woman:H	**DEM1**	woman:H	2SG.PST>3SG	see
yɔ̀ɔ̀tɔlɔ̀	*pàà*	*wāā,*		
doctor	at	BKGR		
à	***dɛ̄***	***yā***,	*Dɔ̀mà*	*zíé*	*lɛ̄.*
3SG	**husband**	**DEM2**	P.N.	uncle	COP

"That woman, that woman that you saw at the hospital, that husband of hers, he is Doma's uncle" [fieldnotes].

In Mano, physical accessibility and engagement affects the use of *tɔ́ɔ̄* vs. *dḭ̀ā̰*. The difference between *wɛ̄* and *yā* is not yet clear from the data, but it is possible that a similar contrast is at play where engagement is seen in a more abstract way as a sphere of ownership, control, familiarity or mental preoccupation.

The objects referred to in (21) are expected to be served to the speaker by the addressee in the situation that the utterance describes. Therefore, although the referents are known to both parties in the interaction, which motivates the recognitional function, they belong to the sphere of the assumed control and possession of the speaker, so *wɛ̄* is chosen over *yā*.

(21)

kɛ̄	*ékɛ̀*	*ó*	*nū*	*à*	*gèē*	*à*
so.that	PROH	3PL.CONJ	come	3SG	say	3SG
lɛ̀ɛ̄,	*nū*	***yɔ̄***	*ɓɛ̄*	*ká,*		
ADR	come	**wine**	**DEM1**	with		
nū	***pɛ̄***	***yé***	*ɓɛ̄*	*ká,*	*ē*	*wàà.*
come	**thing**	**REL**	**DEM1**	with	3SG.PST	enter

(A man's mother and father died, but he did not have money to organize their funerals). "So that people don't come saying: "Bring this wine, bring this thing" (the food and the drinks that invitees at a funeral are expected to be served), he ran away" [MOC].

Example (22) is from my notes of my consultant's children commenting on pictures in a comic book. The children at that time were not fluent in French, so they could not read what was actually written in the word balloons and instead staged an imagined conversation between the book's characters. The characters played with the referent (the ball) together and had equal access to the information in question.

(22)

kóò	*lō*	***dèèkpō***	***yā***	*ŋwɛ́ŋ̀lɔ̀ɔ̄*
1PL.IPFV	go:IPFV	**ball**	**DEM2**	question
kɛ̄-ɛ̀.				
do-GER				

(Children were playing with a ball and accidentally threw it to the other side of the neighbor's fence. They are deciding among themselves what to do with the ball). "We will ask about **that ball**" [fieldnotes].

Example (23) is taken from a conversation between relatives, two sisters-in-law, but it is the addressee who is more knowledgeable about the whereabouts of the referent, her children.

(23)

māē,	*ŋ̄*	*kɛ̄*	*à*	*gèē-pɛ̀lɛ̀,*	
1SG.EMPH	1SG.PST	be	3SG	say-INF	
ɓà	***nɔ́ɔ̄ɓé***	***vɔ̀***	***yāā***	*séŋ́*	*ō*
2SG.POSS	**child.PL**	**PL**	**DEM2**	all	3PL.EXI
nū-pɛ̀lɛ̀.					
come-INF					

(A woman is talking to her sister-in-law, who came to celebrate the New Year with her, having brought only some of her children with her). "As for me, I thought that all of **those children** of yours were coming" [MOC].

Thus, the contrast between the demonstratives *wɛ̄* and *yā* implies the contrast in the engagement with the referent, where *wɛ̄* covers the engagement sphere of the speaker and *yā* is used for the common sphere or the addressee's sphere. The contrast emerges from prior interactions and expectations that the interlocutors have: who owns and controls what (children, food served to a guest), what business is a matter of common concern, and what is taken as a personal matter.

Semantics and Pragmatics of Mano Demonstratives

As argued above, the contrast within the two pairs of demonstratives, *tɔ́ɔ̄* and *dḭ̀ā̰*, and *wɛ̄* and *yā*, is motivated by the factor of the engagement sphere: *dḭ̀ā̰* is used for referents outside the engagement sphere of the speaker and *tɔ́ɔ̄* is neutral in that regard, while *wɛ̄* is used for referents within the engagement sphere of the speaker and *yā* is neutral. In the case of the former pair, engagement is understood in the sense of Enfield (2003) as an area of physical activity. In the case of the latter pair, engagement is seen in a more abstract way as an area of one's expertise, familiarity or control (see Evans et al., 2018).

The referents of the noun phrase framed with *tɔ́ɔ̄* and *dì̱ā̱* are objects that, as a rule, were not mentioned in the discussion immediately prior to the act of reference. Usually there is extra work needed to establish joint visual attention to the referent. The attention management marker *tɛ́*∼*lɛ́* that is obligatorily used with *dì̱ā̱* and that is likely embedded in the form of *tɔ́ɔ̄* informs the addressee that she needs to align her attention with a non-trivial referent (Khachaturyan and Ozerov, in preparation). Gesture becomes a key means of securing joint attention and establishing reference and usually accompanies noun phrases with the demonstratives *tɔ́ɔ̄* and *dì̱ā̱*.

In contrast, the common feature of all uses of *wɛ̄* and *yā* listed above is that the referents are easily identifiable given the common ground of the interlocutors. And yet they are also compatible with deictic gesture, as shown in (14), and, more importantly, with the marker of attention alignment. Example (24) is taken from a spontaneous translation of Luke 9:35, where God announces that Jesus is his son and is chosen by him. Note that God is speaking from a cloud, which complicates reference resolution and triggers the use of the attention management marker. However, given the unusual circumstances of the referential act, no pointing is possible, so neither *tɔ̀ɔ̄* nor *dì̱ā̱* is possible in this context.

(24)

mí	***tɛ́***	***yā***,	*lɛ̄*	*ŋ̀*	*nɛ́*
person:H	**ATT**	**DEM2**	3SG.EXI	1SG.POSS	child

ká.
with

"**That person**, he is my Son" [MOC].

The attention management marker may be used to accompany *wɛ̄* or *yā* when the referent is already in joint attention, but additional attention needs to be brought to it, as in emotional evaluations. Example (25) is taken from an explanation of the Bible episode where Jesus preaches in a synagogue in Nazareth, the town where he grew up. The Jews present in the synagogue know him well and are surprised that the "gracious" words are said by a man of such modest descent—the son of Joseph and Mary. Note that everyone is already attending to Jesus ("the eyes of everyone in the synagogue were fastened on him," Luke 4:20, NIV). The attention management marker that the prayer leader employs in his explanation of the situation is used in the expression of surprise—similar to the emphatic use studied by Levinson (2004)—rather than to overcome the difficulty of attention alignment[6].

(25)

mí	***tɛ́***	***wɛ̄***,	*ŋ̀ŋ́gɔ̀*	*Josef*	*gbē*
person:H	**ATT**	**DEM1**	isn't.it	Joseph	son

lɛ́	*wāā*,	*ŋ̀ŋ́gɔ̀*	*Marie*	*gbē*	*lɛ́*	*ō!*
ATT	DEM1	isn't.it	Mary	son	ATT	INJ

[6]The same attention management marker is also used to form focus constructions (see ex. 9, was it SOMEONE who told you, or you made it up?). On the relationship between attention management and focus, see Khachaturyan and Ozerov (in preparation).

TABLE 2 | Semantics of Mano demonstratives.

yā	DEM
wɛ̄	DEM, within speaker's engagement sphere
tɔ́ɔ̄	DEM, attention drawing
dì̱ā̱	DEM, attention drawing, outside speaker's engagement sphere

"That person, isn't he JOSEPH'S SON, isn't he MARY'S SON!" [MOC].

Thus, being-part-of-common-ground is not an invariant meaning of *wɛ̄* and *yā*, despite its frequent occurrence in natural demonstrative use. Instead, these demonstratives can be argued to have a general indicating function, DEM (Enfield, 2003). The use of a demonstrative in that function "presupposes that an addressee can know what it is referring to" (Enfield, 2003, p. 86). In contrast, *tɔ́ɔ̄* and *dì̱ā̱*, in addition to the indicating function, have the semantic function of drawing attention to a non-trivial referent. That *wɛ̄* and *yā* are often used to indicate that the referent is part of the common ground is a pragmatic inference (Levinson, 2000): "the use of a semantically less specific or "weaker" form (given that a semantically more specific or "stronger" form is an option in the same grammatical context) implies the converse of the stronger form, yet without semantically encoding it" (Enfield, 2003, p. 86). In other words, because the speaker chooses not to use the semantically specific attention-drawing markers *tɔ́ɔ̄* and *dì̱ā̱*, the addressee infers that extra work of attention alignment is not needed and that the referent is likely already available to her by virtue of the common ground she shares with the speaker. Yet, the inference can be overridden by an explicit use of the attention drawing marker *lɛ́*. **Table 2** summarizes the invariant semantics of the four Mano demonstratives from the least to the most specific.

COMMON GROUND AND REFERENCE RESOLUTION IN INTERACTION

Common Ground and Interactional History

A further layer of complexity arises when referential acts are seen not in isolation but as embedded in interactional sequences. As interaction unfolds within a given encounter and across encounters, more knowledge about referents, including those present at the interactive scene, becomes mutually available to the participants. The simplest case of mutual knowledge built in interaction and indexed by a demonstrative is anaphora. Indeed, the demonstratives *yā* and *wɛ̄*, which are commonly used for reference tracking in monological texts (see section "Anaphora"), are also used in conversations for reference tracking across speech turns. In 26, the two interaction participants are engaged in cooking. In (26.1) the speaker A draws her addressee's attention to the fish, which has a lot of bones. As joint attention is established, the speaker B uses *yāā* in the anaphoric function to confirm and elaborate on A's observation (26.2).

(26)

(26.1) Speaker A ***kpa̰á*** ***tɔ́ɔ̄***, *gɛ́nɛ́* *lāā*
fish DEM.ATT bone 3SG.EXI>3SG
yí *tōn̄tōn̄* *ká* *ō*
in too.much with INJ
"**This fish**, there are too many bones in it!"

(26.2) Speaker B ***gɛ́nɛ́*** ***yāā*** *lāā* *yí*
bone DEM2 3SG.EXI>3SG in
ē *kélékélé* *ká*
3SG.REFL small.PL with
"**Those bones**, they're all small in it. [MOC]"

Example 26 is a "lean" case, where knowledge about referents is available from the interaction setting and discourse immediately preceding the use of the demonstratives. Section "Recognitional Function" presented further cases where some already available mutual knowledge was necessary for referent identification. The next three examples illustrate cases where referent identification is based on mutual knowledge which is assumed by the speaker but negotiated in interaction.

In (27), a woman (Speaker A) is helping her sister-in-law (Sister B) cook a festival dish, namely rice with mixed protein, fish and duck. Poulty is a more typical protein to be put in such a dish, so Speaker A is surprised they are adding fish and assumes that Speaker B is doing so to offer some food to her mother, who residing in a village called Gou and is known to be a fish lover. And yet it was another person who asked to prepare the fish. The person was first identified by Speaker B as "that woman" with the *ā* demonstrative (a variant of *yā*) in the recognitional function (27.4), and then Speaker A made sure they are talking about the same person by using a proper name, Maria (27.5).

(27)

(27.1) Speaker A *òó* *gèē* *zénīí* *kā*
3PL.NEG>3SG say again 1PL.EXI
ló-pɛ̀lɛ̀ *Gúù* *kɛ̀ɛ̀?*
go-INF Gou isn't.it
"Don't they say you are going to Gou?"

(27.2) Speaker B *dēn̄* *lɛ́* *ā*
who ATT 3SG.PST>3SG
gèē *kílī?*
say like.this
"Who said that?"

(27.3) Speaker A *dēn̄* *là* *kpa̰á* *lɛ̄* *yī?*
who 3SG.POSS fish COP there
"Whose fish is it?"

(27.4) Speaker B ***léé*** ***ā***
woman **DEM2**
"**That woman**"

(27.5) Speaker A *Màrìà?*
"Maria?"

(27.6) Speaker B *ŋ̀ŋ̀*
"Yes"

(27.7) Speaker A *lɛ̀ɛ́* *wìì* *ɓɛ̀lɛ̀ɛ̀?*
3SG.NEG meat eat.Q
"She doesn't eat meat?"

(27.8) Speaker B *ŋ̀ŋ̀*
"Yes"

In addition to mutual knowledge, the engagement sphere factor structures local interactions, where a personal concern is put on the table and then taken up by the addressee as shared or, in contrast, a shared concern is projected and then recognized and validated by the addressee as hers. The next two examples illustrate that. In the elicited example (28) the speaker presents a referent as an object of his personal concerns and uses the demonstrative *wɛ̄*, which encodes the speaker's engagement area, while the addressee, ratifying the shared recognizability and at the same time conveying some additional information, uses the demonstrative *yā*.

(28)

(28.1) Speaker A *ŋ̀ŋ́* *yí* *dɔ̄*
1SG.NEG>3SG interior know
ékɛ̄á ***gɔ̰̄*** ***wɛ̄***
if man DEM1
lɛ́ɛ̀ *lō* *nū-à* *zèē*
3SG.IPFV go:IPFV come-GER here
ā.
BKGR
"I don't know whether **this guy** (I am expecting) is going to come."

(28.2) Speaker B ***gɔ̰̄*** ***yā*** *ē* *nū* *zèē*
man.H DEM1 3SG.PST come here
"**The man** came." [el.]

In (29), the sequence is inverse. The two speakers chat about several things, including a small eggplant plantation of one of their relatives, which keeps an elderly aunt of Speaker A busy (who is also the mother-in-law of Speaker B). By using the demonstrative *yā* and framing the issue as shared knowledge (29.1), A attempts to elicit a confirmation from her interlocutor that she follows what is being talked about. She receives feedback with the demonstrative *wɛ̄*, which indexes that the speaker recognizes the referent and includes it into her personal sphere (29.2).

(29)

(29.1) Speaker A *tàtí* *Gɔ̀níá* *là* ***kɔ̀nɛ̄***
auntie[FR] Gonia 3SG.POSS eggplant
ŋwɔ̀ ***yā*** *kɛ̀lɛ́ɛ̀*
problem:CSTR DEM2 isn't.it

yē *láā* *kɛ̄*
3SG.EMPH ATT>3SG.PST>3SG do
ɛ̄? *là* ***pílé***
BKGR 3SG.POSS heap
ŋwɔ̀ ***yāā?***
problem:CSTR DEM2
"It's **that eggplant business** of the auntie from Gonia that did it (so that the old woman is busy in the field), isn't it? **That garden business** of hers?"

(29.2) Speaker B ***là*** ***pílé*** ***ŋwɔ̀***
3SG.POSS heap problem:CSTR
wɛ̄.
DEM1
"**this garden business of hers**"
[MOC]

Common Ground and Social Fields

The previous section illustrates that a typical referential act is embedded in an interactional sequence of reference resolution. At the same time it is embedded in the longer-term history of interaction between the given participants, allowing them to have access to mutual knowledge. Referential acts are also part of partially scripted social activities taking place in social fields where participants occupy particular positions with relations of power and reciprocity. As argued by Hanks, because of this embeddedness, the interactive space defined by deixis is "invested with much more specific values and relationships whose interpretation turns not on deixis," but on a particular field (Hanks, 2005, p. 194). In particular, there is often domain-specific knowledge involved that the participants share even when a particular group of interactants has never communicated before.

The following example is an excerpt from a highly scripted type of discourse, a benediction ritual which is part of a traditional name-giving ceremony. The speaker, a classificatory nephew performing the benediction, utters a sequence of blessings to a newborn boy framed in the conjunctive verbal form ("let him be such," "let such a thing happen to him"). The public responds by repeating the end of each token phrase of the benediction in the habitualis form ("he is such," "such a thing happens to him"). The speaker refers to two abstract qualities (growing force, good intelligence) and one physical (a shining thing between the thighs, meaning well-functioning reproductive organs). Crucially, he refers to all three with the *wɛ̄* demonstrative because these are typical things to wish to a boy. The consistent use of the same demonstrative and the same tense forms endows the interaction with a rhythmical, routinized structure characteristic of the ritual context. At the same time, given that the boy is also present during the ritual, the reference to these qualities—especially the physical one—has an exophoric dimension.

(30)

benediction	***ī***	***fàŋá***	***wɛ̄***	*é*	*tɛ̀nɛ̀*!
	2SG	**force**	**DEM1**	3sg.conj	climb
	"let **this force of yours** grow!"				
response	*lɛ́ɛ̀*	*tɛ̀nɛ̀*!			
	3SG.IPFV	climb:IPFV			
	"it grows!"				
benediction	***ī***	***kílí***	***yīè***	***wɛ̄***	*é*
	2SG	**intelligence**	**good**	**DEM1**	**3SG.CONJ**
	nū!				
	come				
	"let **this good intelligence of yours** come!"				
response	*lɛ́ɛ̀*	*nū*!			
	3SG.IPFV	come:IPFV			
	"it comes!"				
benedication	*nàá,*	*mā*	*gèē*	*gə̰̄*	*lɛ́* *ī*
	man	1SG.PST>3SG	say	man.H	ATT 2SG
	ká,	*nàá,*	*mā*	*gèē,*	
	with	man	1SG.PST>3SG	say	
	ī	***gbáá***	***fɛ̀ŋ***	***wɛ̄***	*é* *ɓí!*
	2SG	**thigh**	**between**	**DEM1**	**3sg.conj** shine
	"man, I said, you are a boy, man, I said, let **this thing between your thighs** shine!"[7]				
response	*lɛ́ɛ̀*	*ɓī!*			
	3SG.CONJ	shine:IPFV			
	"it shines!"				

In some cases, the use of the demonstratives *wɛ̄* and *yā* does not only index the common ground and the routinized properties of interactions in a particular field, but also contributes to shape the context of interaction as a distinct social field with a presupposition of shared knowledge. Thus, in oral Bible translations as they are performed by Mano priests and prayer leaders, many noun phrases contain the demonstratives *wɛ̄* and *yā*. They are often used in first mentions of objects and places beyond the utterance context and perform the recognitional function. Many of these referents, however, are fairly exotic and cannot be assumed to be known by the community of Mano Catholics, such as the Horeb mountain in (31). Instead of indexing the shared knowledge of the referents in question, these deictic markers project it in a performative fashion. Because of the dialogic orientation of recognitional deixis, as a consequence of projection of recognizability, the speaker (a ritual specialist) and the addressees (the congregation) emerge as knowledge-sharing co-insiders. This, in turn, contributes to a performative creation of a community of co-insiders—a religious community sharing religious knowledge (Khachaturyan, 2019).

(31)

ē	*nū*	*là*	*tòlòpɛ̀*		*vɔ̀*	*yā*
3SG.PST	come	3SG.POSS	domestic.animal		PL	DEM2
ká	***yɛ̀í***	***kpóŋ***	***yā***			
with	**savannah**	**border**	**DEM2**			
yí	*Horeb,*	*ɛ́ɛ́*	*ē*	*kɛ̄*	*wálà*	*là*
in	Horeb	and	3SG.PST	be	God	3SG.POSS
tɔ̀ŋ						
mountain						
yāā	*ŋwíí*	*ká.*				
DEM2	top	with				

"He came with those domestic animals of his at **that border of the savannah**, at Horeb, that was a top of **that mountain** of God's."

French source: "Il mena le troupeau au-delà du désert et parvint à la montagne de Dieu, à l'Horeb."

NIV: "... and he led the flock to the far side of the wilderness and came to Horeb, the mountain of God."

Both (30) and (31) are examples of language use with an unusual participation framework where the speaker is a ritual specialist. Everyday family and village life is no less scripted than the fields of religious practice and is filled with routinized activities. It is the position of the wife of the elder brother that allows the speaker in (15) to give orders to her brother-in-law with minimal referential expressions ("do the thing with the fire"). Similarly, the seller of popcorn balls recurs to a recognizable marketing formula, "buy these X of yours" (2).

DISCUSSION

The relation of proximity is a function of the natural and social carving of the physical space. Such spatial divisions contribute to forming the notion of wider physical accessibility, which accounts for the use of speaker- or addressee-anchored forms (Burenhult, 2018). Furthermore, proximity is in certain cases a function of the bodily engagement of the interlocutors with the object and physical access to it. The notion of peripersonal space, which can be extended if the speaker uses tools (such as a stick) is in certain languages a better predictor of the choice of the demonstrative form than exact distance (Coventry et al., 2008).

As shown in examples (4) and (5) from Mano, however, the referent located at a similar distance outside the peripersonal space (about 2–3 m) may be framed with the marker *dìā̱* if the speaker intends to physically engage with the object but cannot reach it, or with the proximal marker *tɔ́ɔ̃* if mere pointing and identification is enough for the current purposes. If the speaker is busy with some chores and her hands are occupied, so that she cannot point (15) or if the addressee is busy with some tasks and cannot attend to pointing gestures (2), his attention can be called for by framing the referent as if it was invisible. Thus, the purpose of referent identification and the activity which it is part of motivate the choice of deictic marker. The engagement area, which is defined as "the place which is, at moment *t*, the conceived site of a person's currently dominant manual and attentional engagement" (Enfield, 2003, p. 89) and which shifts depending on the interactional setting and the interlocutor's current activity, matters for the choice of demonstrative reference sometimes more than physical distance *per se*. In Mano, the engagement area contrasts the demonstratives *tɔ́ɔ̃* and *dìā̱*, where the latter is situated outside the engagement area and the former is neutral in that regard. Engagement in a more abstract sense as an area of one's ownership, control, or concern appears to contrast the markers *yā* and *wɛ̄*, where the latter marks objects within the speaker's engagement area and the former is neutral.

Attention focus management as an evolving interactive process which gets reflected by the choice of the demonstrative form has been recognized in much recent literature. A particularly well-known case is Turkish, mentioned in section "Introduction." In Jahai, the addressee-centered marker *ton* is used when the addressee's attention is already focused on the referent (Burenhult, 2003). Tiriyó is another language where the addressee's attention focus, not physical distance, influences the choice between (two proximal) markers (Meira, 2018). In Mano, the demonstrative *dìā̱* is always used with a marker of attention alignment *lɛ́* or *tɛ́*. While the demonstrative *tɔ́ɔ̃* never combines with such a marker, it likely derives historically from a merger with *tɛ́*. Rather than a means to express attention alignment across attempts at securing the addressee's attention focus, the function common to both demonstratives is to underscore that the addressee needs to do some extra work to identify the referent; here, pointing at visible objects which often accompanies *tɔ́ɔ̃* and *dìā̱* is a means to secure attention.

Demonstratives *wɛ̄* and *yā* can also be used with the attention alignment markers *lɛ́* and *tɛ́* when there is some difficulty in referent identification (God speaking from a cloud and hence not being able to point) or there is some emphatic attention alignment. Thus, while *tɔ́ɔ̃* and *dìā̱* can be characterized as +ATTENTION ALIGNMENT, *wɛ̄* and *yā* are not -ATTENTION ALIGNMENT, but rather neutral and have the most general indicating function proper to demonstratives as a class (Enfield, 2003) as their primary meaning. Yet, in speech events they are often interpreted as marking referents that do not require additional attention alignment as they are already part of the common ground of the interlocutors. This interpretation arises from a pragmatic inference whereby the use of a semantically weaker term implies the opposite of the semantically stronger term that the speaker chose not to use. A similar contrast between a general indicating demonstrative and a demonstrative that indexes referents which require additional coordination between the speaker and the addressee is also attested in Yurakaré (Gipper, 2017).

The pragmatic function of the demonstratives *wɛ̄* and *yā* of marking common ground in reference to objects present at the interactive scene is very frequently observed in interaction. Because these demonstratives do not have visual attention alignment as a necessary component of their semantics, *wɛ̄* and *yā* are used in a variety of endophoric functions. Moreover, the markers *wɛ̄* and *yā* grammaticalize into generalized clause-final markers used explicitly to mark backgrounded information. Thus, *wɛ̄* is also used in imperative clauses when the request or invitation is highly expected in the given context (see also the utterance final *wɛ̄* in 2 and 13); both *wɛ̄* and *yā* are used as subordinate clause markers (see examples 8 and 10; Khachaturyan, 2018b).

The interlocutors' common ground is difficult to assess in experimental settings and is much more rarely discussed as a parameter for the demonstrative choice. And yet it seems to be more basic than some other interactional parameters discussed in the literature. In particular, it often underlies the rationale for the choice of demonstrative for invisible referents and overrules the visibility parameter *per se*, as in Yélî Dnye (Levinson, 2005, 2018). Common ground seems to be a contrast for the "invisible" forms in Quileute (Andrade, 1933), where one set of markers denotes referents which are known to the addressee and another the referents which are known to both parties. It is because the referents are cognitively accessible that they are identifiable while being invisible. (In)visibility *per se* is not encoded in Mano but is a contextual factor that favors the use of the demonstratives *wɛ̄* or *yā* that do not encode the attention coordination function and can be used to mark invisible, but cognitively available referents.

The way attentional and common ground distinctions are mapped into demonstrative systems varies from language to language. In Jahai, the same marker *ton* is used to confirm mutual attention to a referent, as well as in cases where referents form part of the common ground and for anaphoric reference (Burenhult, 2003). In Tiriyó, the marker used for referents already attended to can be used only for exophoric, and never for anaphoric reference (Meira, 2018). In Yélî Dnye, the common ground marker is used in the recognitional function, but not for anaphora, for which a dedicated marker exists (Levinson, 2018). In Yucatec Maya, the same non-immediate enclitic is used for recognitional and for anaphoric reference, as well as in exophoric uses where common ground is involved, but the contrast between the functions is expressed by a proclitic (Hanks, 2005). Finally, in Mano, the demonstratives *wɛ̄* and *yā* cover all functions in the endophoric domain and are used for exophoric-cum-endophoric reference when common ground appears to matter.

Common ground arises from interactional history and broader cultural knowledge. On the one hand, any referential act in natural language use is part of an interactional sequence and interactional history involving the current interlocutors. On the other hand, it belongs to the domain of social action occurring in social fields that endow the interlocutors with social roles. This double embedding (in terms of Hanks, 2005) makes referents mutually known to the interlocutors, and therefore, cognitively available and anticipated. Thus, cognitive accessibility becomes one of the factors determining the choice of a deictic marker in exophoric reference. In Yucatec Maya, it is likely the shared interactional history, which is responsible for the routinized nature of certain types of interactions, such as greetings or scoldings, that triggers the choice of the non-immediate deictic over the immediate deictic in speaker-proximal settings (Hanks, 1990, 2005). In Mano, the demonstratives *wɛ̄* and *yā* are systematically used to mark referents in particular routinized speech genres, such as benedictions.

An additional complication regarding common ground is that it is not a fixed artifact: it can be creatively shaped by individuals and negotiated in interaction. Wrongly assuming common ground may lead to failures in recipient design (Deppermann, 2015) and additional interactional work in referent identification (Khachaturyan, 2019). Creative common ground management may become a feature of certain registers, as I show in the example of the Catholic register, where the use of demonstratives frames some referents as known to the congregation, while there are reasons to doubt their universal recognizability. This register feature arguably has broader consequences for shaping the interactional context, since it concomitantly shapes the addressees, the Catholic congregation, as a community of knowledge-sharing co-insiders.

CONCLUSION

This paper is a first-hand ethnographic account of demonstrative reference in an under described language, Mano (Mande). It argues that in exophoric reference, the Mano demonstratives *tɔ́ɔ̄* and *dɪ̰ā̰* are contrasted with the demonstratives *wɛ̄* and *yā* in that the former index referents that require attention coordination for referent identification. In contrast, *wɛ̄* and *yā* are commonly interpreted, as a result of pragmatic inference, as the opposite of *tɔ́ɔ̄* and *dɪ̰ā̰* and index referents that do not require attention coordination being part of the common ground. They do not semantically encode common ground, however, having a more general referent identification function as their invariant semantics (Enfield, 2003; Gipper, 2017).

Common ground, including mutual knowledge of the referents, cannot be accounted for unless one studies demonstrative reference in natural use. Indeed, referential acts are part, first, of the immediate interactional sequence of referent identification and negotiation (Küntay and Özyürek, 2006), and second, in long-term interactional history involving given participants (Deppermann, 2018; Harjunpää et al., 2021). A further layer is the embeddedness of referential acts into the fabric of social action within particular social fields (Hanks, 2005). All these levels provide speech act participants with knowledge about referents which enable referent identification for objects both within and outside the interactive space. Thus, as argued by Agha (1996), Burenhult (2003), and most prominently Hanks (2005, 2011), cognitive access to referents functions alongside of perceptual access, which includes, but is not restricted to, spatial (visual) access. The centrality of common ground in reference to the objects at the interactive scene challenges the often assumed classification of demonstrative reference into exophoric and endophoric and, as a logical consequence of this, the primacy of exophoric reference at the level of the actual referential practice.

This article adds another argument in favor of viewing demonstrative reference as a social, interactive process (Peeters and Özyürek, 2016). It contributes to the empirical studies of (demonstrative) reference by bringing together the interactionist perspective and the sociological concept of field (Hanks, 2005) and by drawing on examples from distinct social domains of interaction, including everyday conversations and ritual discourse. The lack of a theoretical framework that would articulate the interlocking of interactional space, interactional history and social fields is a major shortcoming of this paper. A promising line of research which would support the development of such a framework would be a longitudinal study of (language) socialization within particular fields—how do people come to inhabit their social roles, know what they know and how is this process reflected in referential practice?

List of Languages Cited and Their ISO 639-3 Codes

Jahai, Austroasiatic [jhi]; Mano, Mande [mev]; Quileute, Chimakuan [qui]; Tiriyó, Cariban [tri]; Turkish, Turkic [tur]; Yélî Dnye, isolate [yle]; Yucatec Maya, Mayan [yua]; Yurakaré, isolate [yuz].

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available. The Mano Oral Corpus, on which some part of the research is based, will gradually be made publicly available at least in the anonymized and transcribed form and, current European GDPR allowing, also in the audio form. Some parts of the materials, namely, elicitation, can also be made available upon request. Fieldnotes, however, will not be made available since they contain sensitive information.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was based on fieldwork that was generously supported by the Fondation Fyssen, France, in 2015–2016 and the University of Helsinki, 2017–2018 and 2018–2019. In 2020, my work was supported by the Kone Foundation, Finland, application number 201907156.

ACKNOWLEDGMENTS

I gratefully acknowledge the warmth and hospitality of my Mano hosting family, in particular my primary language consultant, Pe Mamy, and his wife, Pola Kourouma. Without them, the ethnographic research would not have been so enjoyable. I thank Pe Mamy for his patience and for his insightful discussions about Mano. Different versions of this manuscript have benefited from comments from Denis Creissels, Katariina Harjunpää, Don Killian, and Anna Margetts. I thank the reviewers whose insightful comments helped me to significantly revise my argument. Last but not least, I thank William F. Hanks, who inspired me to look into Mano deixis.

REFERENCES

Agha, A. (1996). Schema and superposition in spatial deixis. *Anthropol. Linguist.* 38, 643–682.

Anderson, S. L., and Keenan, E. L. (1985). "Deixis," in *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Vol. III, ed. T. Shopen (Cambridge: Cambridge University Press), 259–308.

Andrade, M. J. (1933). "Quileute," in *Handbook of American Indian Languages*, Vol. 3, ed. F. Boas (New York, NY: J.J.Augustin), 149–292.

Bah, C. M., and Bangoura, M. A. (2017). *Rapport d'Analyse des Données du Troisième Recensement Général de la Population et de l'Habitation*. Conakry: Institut national de statistique.

Bourdieu, P. (1990). *The Logic of Practice*. Cambridge: Polity Press.

Burenhult, N. (2003). Attention, accessibility, and the addressee: the case of the Jahai demonstrative *ton*. *Pragmatics* 13, 363–379. doi: 10.1075/prag.13.3.01bur

Burenhult, N. (2018). "The Jahai multi-term demonstrative system: what's spatial about it?," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Cambridge: Cambridge University Press), 361–380. doi: 10.1017/9781108333818.018

Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *J. Verbal Learn. Verbal Behav.* 22, 245–258. doi: 10.1016/s0022-5371(83)90189-5

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near-far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

Deppermann, A. (2015). When recipient design fails: egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. *Gesprächsforschung* 16, 63–101.

Deppermann, A. (2018). "Changes in turn-design over interactional histories – the case of instructions in driving school lessons," in *Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources*, eds A. Deppermann and J. Streeck (Amsterdam: John Benjamins), 293–324. doi: 10.1075/pbns.293.09dep

Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489.

Enfield, N. J. (2003). Demonstratives in space and interaction: data from lao speakers and implications for semantic analysis. *Language* 79, 82–117. doi: 10.1353/lan.2003.0075

Eriksson, Ms (2009). Referring as interaction: on the interplay between linguistic and bodily practices. *J. Pragmat.* 41, 240–262. doi: 10.1016/j.pragma.2008.10.011

Etelämäki, M. (2009). The Finnish demonstrative pronouns in light of interaction. *J. Pragmat.* 41, 25–46. doi: 10.1016/j.pragma.2008.09.005

Evans, N., Bergqvist, H., and San Roque, L. (2018). The grammar of engagement I: framework and initial exemplification. *Lang. Cogn.* 10, 110–140. doi: 10.1017/langcog.2017.21

Gipper, S. (2017). Pre-semantic pragmatics encoded: a non-spatial account of Yurakaré demonstratives. *J. Pragmat.* 120, 122–143. doi: 10.1016/j.pragma.2017.08.012

Hanks, W. F. (1990). *Referential Practice. Language and Lived Space Among the Maya*. Chicago, IL: University of Chicago Press.

Hanks, W. F. (2005). Explorations in the deictic field. *Curr. Anthropol.* 46, 191–220. doi: 10.1086/427120

Hanks, W. F. (2011). "Deixis and indexicality," in *Foundations of Pragmatics*, eds W. Wolfram Bublitz and N. R. Norrick (Berlin: Mouton de Gruyter), 315–346.

Harjunpää, K., Deppermann, A., and Sorjonen, M.-L. (2021). Constructing the Chekhovian inner body in instructions: an interactional history of factuality and agentivity. *J. Pragmat.* 171, 158–174.

Himmelmann, N. P. (1996). "Demonstratives in narrative discourse: a taxonomy of universal uses," in *Studies in Anaphora*, ed. B. A. Fox (Amsterdam: John Benjamins), 205–254. doi: 10.1075/tsl.33.08him

Hindmarsh, J., and Heath, C. (2000). Embodied reference: a study of deixis in workplace interaction. *J. Pragmat.* 32, 1855–1878. doi: 10.1016/S0378-2166(99)00122-8

Khachaturyan, M. (2015). Grammaire du mano. *Mandenkan* 54, 1–252.

Khachaturyan, M. (2018a). A sketch of dialectal variation in Mano. *Mandenkan* 59, 31–56.

Khachaturyan, M. (2018b). "From recognitional function to topicalization and pragmatic backgrounding: discourse functions of demonstratives in Mano, Southern Mande," Paper Presented at the "*Discourse Functions of Demonstratives" Workshop, June 14-15*, (Norway: University of Oslo).

Khachaturyan, M. (2019). Knowing is belonging: recognitional deixis and emergence of common ground in religious conversion. *Signs Soc.* 7, 186–216. doi: 10.1086/702441

Khachaturyan, M. (2020a). A typological portrait of Mano, Southern Mande. *Linguist. Typol.* doi: 10.1515/lingty-2020-2050. [Epub ahead of print].

Khachaturyan, M. (2020b). Bridging caught in a crossfire: the marker of situated definiteness in Mano and language contact. *Lang. Afr.* 1. Forthcoming.

Khachaturyan, M., and Konoshenko, M. (in press). Assessing (a)symmetry in multilingualism: the case of Mano and Kpelle in Guinea. *Int. J. Biling.*

Khachaturyan, M., and Ozerov, P. (in preparation). Focus as a pragmatic interpretation of attention management: the case of Mano (Mande).

Küntay, A. C., and Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *J. Child Lang.* 33, 303–320. doi: 10.1017/S0305000906007380

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.

Levinson, S. C. (2004). "Deixis," in *The Handbook of Pragmatics*, eds L. R. Horn and G. Ward (Hoboken, NJ: Blackwell), 97–121.

Levinson, S. C. (2005). Living with Manny's dangerous idea. *Discourse Stud.* 7, 431–453. doi: 10.1177/1461445605054401

Levinson, S. C. (2018). "Yéli Dnye: demonstratives in the language of Rossel Island, Papua New Guinea," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Cambridge: Cambridge University Press), 318–342. doi: 10.1017/9781108333818.016

Levinson, S. C., Cutfield, S., Dunn, M. J., Enfield, N. J., and Meira, S. (eds). (2018). *Demonstratives in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Margetts, A. (2018). "Saliba-Logea: exophoric demonstratives," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Cambridge: Cambridge University Press), 257–281. doi: 10.1017/9781108333818.013

Meira, S. (2018). "Tiriyó: non-contrastive exophoric uses of demonstratives," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Cambridge: Cambridge University Press), 222–241. doi: 10.1017/9781108333818.011

Peeters, D., and Özyürek, A. (2016). This and that revisited: a social and multimodal approach to spatial demonstratives. *Front. Psychol.* 7:222. doi: 10.3389/fpsyg.2016.00222

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Wilkins, D. P. (2018). "The demonstrative questionnaire: "THIS" and "THAT" in comparative perspective," in *Demonstratives in Cross-Linguistic Perspective*, eds S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Cambridge: Cambridge University Press), 43–71. doi: 10.1017/9781108333818.003

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpsyg.2020.584231

Effects of Scale on Multimodal Deixis: Evidence From Quiahije Chatino

Kate Mesh[1]*, *Emiliana Cruz*[2], *Joost van de Weijer*[1], *Niclas Burenhult*[1,3] and *Marianne Gullberg*[1,3]

[1] Lund University Humanities Lab, Lund University, Lund, Sweden, [2] Department of Anthropology, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS-CDMX), Mexico City, Mexico, [3] Centre for Languages and Literature, Lund University, Lund, Sweden

As humans interact in the world, they often orient one another's attention to objects through the use of spoken demonstrative expressions and head and/or hand movements to point to the objects. Although indicating behaviors have frequently been studied in lab settings, we know surprisingly little about how demonstratives and pointing are used to coordinate attention in large-scale space and in natural contexts. This study investigates how speakers of Quiahije Chatino, an indigenous language of Mexico, use demonstratives and pointing to give directions to named places in large-scale space across multiple scales (local activity, district, state). The results show that the use and coordination of demonstratives and pointing change as the scale of search space for the target grows. At larger scales, demonstratives and pointing are more likely to occur together, and the two signals appear to manage different aspects of the search for the target: demonstratives orient attention primarily to the gesturing body, while pointing provides cues for narrowing the search space. These findings underscore the distinct contributions of speech and gesture to the linguistic composite, while illustrating the dynamic nature of their interplay.

Abstracts in Spanish and Quiahije Chatino are provided as appendices.

Se incluyen como apéndices resúmenes en español y en el chatino de San Juan Quiahije.

Son^{G} ktyi^{C} re^{C} in^{H}, ngyaq^{C} ska^{E} ktyi^{C} no^{E} nda^{H} son^{B} na^{F} nga^{J} no^{I} ngyaq^{C} lo^{E} ktyi^{C} re^{C}, ngyaq^{C} ran^{F} chaq^{E} xlya^{K} qo^{E} chaq^{F} jnya^{J} no^{A} ndywiq^{A} renq^{A} Kchin^{A} Kyqya^{C}.

Keywords: deixis, pointing, multimodality, indicating, demonstratives, Mesoamerica

OPEN ACCESS

Edited by:
Kenny Coventry, University of East Anglia, United Kingdom

Reviewed by:
David Peeters, Tilburg University, Netherlands
Mikkel Wallentin, Aarhus University, Denmark

***Correspondence:**
Kate Mesh
kate.mesh@humlab.lu.se

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 July 2020*
Accepted: *26 October 2020*
Published: *12 January 2021*

Citation:
Mesh K, Cruz E, van de Weijer J, Burenhult N and Gullberg M (2021) Effects of Scale on Multimodal Deixis: Evidence From Quiahije Chatino. Front. Psychol. 11:584231. doi: 10.3389/fpsyg.2020.584231

1. INTRODUCTION

Language users regularly *indicate* entities—that is, they reorient attention to particular spaces, and prompt a search for entities within those spaces. The act of indicating is performed with apparent ease, and yet it is strikingly intricate, often involving the combination of speech and gesture to manage attention. The complexity of indicating, and especially its multimodal character, have drawn interest in the cognitive sciences, with special consideration given to the combination of demonstrative expressions and deictic gestures. Yet studies of these two strategies have mainly explored their use in laboratory settings, asking how pointing and demonstratives are combined to indicate manipulable objects, often within or just outside of the speaker's and addressee's reach. As a result, we know surprisingly little about how demonstratives and deictic gestures are coordinated to manage attention in large-scale space and in actual usage. Here, we present a first study of multimodal indicating that takes into account the effect of scale, and focuses on multimodal

indicating in large-scale space in particular. We study this phenomenon in a naturalistic setting, considering how speakers indicate named places and participate in familiar direction-giving practices. Our study is performed with speakers of Quiahije Chatino, an indigenous language of Mexico in which multiple features of demonstrative use and pointing practice have already been documented, facilitating a closer study of their combination in multimodal indicating acts. We begin the paper by reviewing the theoretical and empirical background to research on indicating, and then contextualize the placement of our project in the Quiahije Chatino-speaking community, before turning to the current empirical study.

2. BACKGROUND

2.1. Elements of an Indicating Event

To *indicate* is to direct attention to something by creating a connection to it in space and/or time (Peirce, 1955; Clark, 1996, 2003, 2020). A typical act of indicating involves a *sender* (a speaker or signer, depending on the language modality), an *addressee* whose attention can be managed (cf. Burenhult, 2018), an object for their attention, variously called a *referent* or *target* (cf. Clark and Bangerter, 2004; Talmy, 2018), and crucially a spoken or embodied *sign* that evokes a delimited search domain in which the target can be found. Some indicating acts draw a connection to an imaginary target (Bühler, 1934; Levy and McNeill, 1992; Cooperrider, 2014; Stukenbrock, 2014; Rocca and Wallentin, 2020) or to a target present in speech rather than in the spatiotemporal context (Levy and McNeill, 1992). A common act of indicating—often deemed prototypical—draws attention to a concrete entity in the real-world space surrounding the sender and addressee: this kind of *exophoric* indicating will be our focus here (Fillmore, 1982; Diessel, 1999; Levinson, 2004; Fricke, 2014).

Spoken languages have a specialized set of signs for indicating—demonstrative expressions, such as English's *this* and *that*, *here* and *there*. In gesture and in sign languages, the same function is served by deictic movements including pointing (Kendon and Versante, 2003; Kita, 2003; Cooperrider and Mesh, in press). Both of these indicating behaviors manage an addressee's attention and delimit the search domain for the target along some dimension(s), such as direction or distance. Both behaviors also invoke other features that may further delimit the search domain, or characterize the participants' perceptual and attentional relationship to the target (Burenhult, 2003; Jungbluth, 2003; Küntay and Özyürek, 2006).

No matter the modality in which it is performed, indicating demands that an addressee attend to the intended target in the search domain. This task is facilitated if the addressee has a conception of the scale of the domain: an expression like *here* might evoke the space on a microscope slide or the expanse of a galaxy, and attention may be aimed quite differently in search domains at different scales.

Some investigations of spatial indicating have explicitly invoked the notion of scale, asking whether speakers have specialized strategies for indicating targets within their reach (cf. Kemmerer, 1999; Wilkins, 1999a; Coventry et al., 2008, 2014; Gudde et al., 2016), within delimited spaces where ongoing activities are taking place (Wilkins, 1999b, 2018), and at "expanded" scales, including "landscape scale" or "large-scale geographical space" (cf. Wilkins, 1999b, 2018; Bril, 2004; Ozanne-Rivierre, 2004; Burenhult, 2008; Schapper and San Roque, 2011). These studies are categorized into two types: research in highly controlled laboratory experimental settings, where the scales in question are typically encompassed within the space of a room, and elicitation studies that consider strategies across a greater range of scales, but report speaker intuitions rather than observed indicating behaviors. As a consequence, we know little about how people indicate targets at different scales in natural communication contexts.

2.2. Demonstratives and Scale

Demonstratives are a closed grammatical class of expressions specialized for indicating: they manage the addressee's attention by inviting a search for some target, and evoking a search domain in which the target can be found. They are deictic, relating the search domain to either of the speech act participants (speaker and addressee) or to the broader speech situation (see, e.g., Burenhult 2008, p. 100). To delimit the search space, demonstratives have traditionally been said to encode paradigmatic oppositions (Himmelman, 1996) of distance (Anderson and Keenan, 1985; Diessel, 1999, 2014; Dixon, 2003). An increasing number of studies finds that demonstrative oppositions are better characterized in terms of participants' shared knowledge and context, rather than in terms of distance (e.g, Laury, 1997; Enfield, 2003, 2018; Piwek et al., 2008; Jarbou, 2010; Peeters et al., 2015b; Peeters and Özyürek, 2016; Rocca et al., 2019), though distance has a role to play in shaping that context (cf. Burenhult, 2003, p. 365; 2018, p. 367).

Talmy (1988, p. 168–169) observes that demonstrative oppositions–whatever their semantic encodings–can operate at multiple scales. He provides an example in the sentences in (1):

(1) a. This speck is smaller than that speck
    b. This planet is smaller than that planet

This observation about the scalability of demonstrative oppositions occasions an empirical question: how do speakers employ demonstrative oppositions across scales? Much of the research on demonstrative use has investigated how speakers employ demonstrative oppositions in small-scale space, with targets in very close proximity to the deictic center. In contrast, we know little about the factors that influence demonstrative use when the search area for the target is at a larger scale, and when the target itself is likely to be larger.

2.3. Pointing and Scale

Pointing is the prototypical deictic gesture. Produced by extending an articulator to form or trace a line, a point invites the addressee to extend that line, conceptualizing a beam projected from the articulator and searching within that beam for an intended target (Kranstedt et al., 2006). Pointing is most often performed with the fingers, hand, and arm, and can take a variety of forms depending on how these articulators are configured to evoke a line (Kendon and Versante, 2003; Wilkins, 2003; Kendon, 2004; Hassemer and McCleary, 2018). Yet it is by no means limited to these articulators: a toss of the head, a jut of the chin, and/or funneling of the lips, combined with gaze in the

target direction, are common indicating gestures in a variety of cultures (e.g., Sherzer, 1973; Enfield, 2001; Mihas, 2017) and may be preferred over manual pointing in some contexts (Cooperrider et al., 2018).

Pointing conveys information not only about the direction of the target, but also about its distance. Some research studies have found that pointing is more likely to occur when the target is farther away, so that its very presence suggests a relatively distant target (Bangerter, 2004; Cooperrider, 2011, 2015). Moreover, the form of the point itself conveys target distance via the *far-is-up* strategy—the farther the target, the higher the pointing arm. This strategy has been attested in pointing across a variety of cultures (Kendon and Versante, 2003; Wilkins, 2003; Mesh, 2017),(Mesh, submitted) and has even been found in non-human primates (Gonseth et al., 2017), suggesting that it may be a fundamental schema for representing distance.

Research on the factors influencing pointing—both its presence and its form—has largely focused on points toward manipulable objects relatively near the deictic center, and visually accessible to both members of the speech dyad (but for work in which visibility is manipulated, see Peeters and Özyürek, 2016). Exceptions to this trend have considered points toward targets in large-scale space without making a comparison between pointing strategies across multiple scales (cf. Mesh, 2017) (Mesh, submitted). As a consequence, we know little about whether pointing strategies shift as the scale of the search domain—and often the scale of the target itself—changes.

2.4. Co-organization of Multimodal Indicating Strategies and Scale

Demonstrative expressions and pointing can be produced and interpreted individually, but are much more often performed together (Diessel, 2006). The semantic contributions of each behavior are distinct, as not all of the perceptual and geophysical dimensions that they invoke are shared (Haviland, 2003, 2009; Kendon and Versante, 2003; Kendon, 2004). Yet the two indicating behaviors jointly facilitate the narrowing of the search domain (Levinson, 2003; Wilkins, 2003; Diessel, 2012). When they are co-produced, demonstratives and pointing are tightly temporally coordinated (Levelt et al., 1985; Chu and Hagoort, 2014; Krivokapic et al., 2016), suggesting that they are planned and organized together in speech production. They are also neurocognitively interpreted jointly (cf. Peeters et al., 2015a), providing further evidence for their connection.

Research on multimodal indicating is still in its early stages, yet the work to date has decisively shown that pointing and demonstratives are more than merely connected in function—they are manifestly co-organized (Bangerter, 2004; Cooperrider, 2011). Whether the two behaviors are coordinated in the same way for indicating at different scales, however, is still unknown.

2.5. Demonstratives and Pointing in Quiahije Chatino

2.5.1. Setting: San Juan Quiahije, Oaxaca, Mexico

Quiahije Chatino is spoken by the ∼3,600 inhabitants of the San Juan Quiahije municipality in Oaxaca, Mexico (INEGI, 2010). It is a variety of Eastern Chatino, one of three Chatino languages classified in the Zapotecan branch of the Otomanguean language stock (Campbell, 2013). The language is characterized by an intricate morphophonological system, with both grammatical and lexical distinctions encoded tonally (Cruz, 2011).

The Quiahije variety of Chatino is notably vital: children are still acquiring it as their first language, even as many of the surrounding Chatino communities are undergoing rapid language shift to Spanish (Cruz and Woodbury, 2014; Villard and Sullivant, 2016). Nevertheless, many of the Quiahije community's oral traditions are not being transmitted to younger generations (cf. Cruz, 2014). Recognizing that their community runs the risk of losing its traditions, community members in Quiahije have begun working with elders to preserve local knowledge. Early projects have focused on knowledge about the landscape and in particular on place names and practices for giving route directions (Cruz, 2017). Expertise in this domain was common in the community as recently as one generation ago, as community members navigated the mountainous terrain in the southern Sierra Madre mountain range to reach neighboring communities and to conduct trade. At present, there are many community elders who can faithfully describe the contours of trade routes that take as many as 5 days to walk (Smith Aguilar, 2017). For these speakers to locate crucial landmarks along the route, two linked indicating behaviors are indispensable: demonstrative expressions and pointing gestures.

2.5.2. Demonstratives in Quiahije Chatino

Quiahije Chatino demonstratives are a closed and formally diverse class of five forms serving to indicate referents in relation to the deictic center. Four of the demonstrative forms are used for exophoric reference (i.e., reference to objects and entities in the real-world environment) and one form is used for discourse anaphoric reference. The exophoric demonstratives have been analyzed in terms of distance from and/or accessibility to the speech act participants (Cruz and Sullivant, 2012; Mesh, 2017). The preliminary analysis for the system is summarized in **Table 1**.

Discourse-givenness and/or discourse focus appear to influence the choice of the speaker-anchored proximal forms, while other features of their semantics appear to be shared. As a consequence, we discuss "speaker-anchored forms" broadly in this paper.

TABLE 1 | The Quiahije Chatino demonstrative system.

Demonstrative form(s)[a]	Gloss	Functional distinction
re^{C}/nde^{C}	dem:1	Speaker-anchored proximal
kwa^{J}	dem:2	Addressee-anchored proximal
kwa^{F}	dem:n	Unmarked/neutral
kanq^{G}	dem:d	Discourse anaphoric

[a] *We use a practical orthography to transcribe Chatino, rather than the International Phonetic Alphabet, and we represent the tone of each word using a superscripted letter. The orthography, including the letters assigned to each tone value, is presented in* ***Appendix 1****.*

All five demonstrative forms can occur as pronouns when preceded by the nominalizing particle *no*^{A}, as shown in Example 2a[1]. All five forms can also occur as adnominals (adjectives) when preceded by a noun or followed by a relational noun, as shown in Example 2b. The exophoric demonstrative forms can occur as adverbs, alone or preceded by the locative particles *ti*^{H} or *ri*^{H}, as shown in Example 2c.

(2) a. tlu^{C} la^{E} xneq^{C} no^{C} nde^{C}
big more dog NOM DEM:1
"that is the bigger dog"

b. xneq^{C} kwa^{F} tlu^{C} la^{E}
dog DEM:N big more
"that dog is big"

c. (ri^{H}) kwa^{F} ntyjaq^{B} xneq^{K}
LOC DEM:N sleep.3sg dog
"the dog sleeps (over) there"

The current functional description of the Quiahije Chatino demonstrative system is based on elicited speaker judgments. No research to date has investigated demonstrative function and demonstrative choice in Quiahije Chatino speakers during spontaneous discourse.

2.5.3. Pointing in San Juan Quiahije

Two forms of pointing are frequent in face-to-face interaction in Quiahije: the manual point and the chin point (a jut of the chin, optionally with pursed, extended lips). When using the hand and arm to point, Quiahije Chatino speakers have been shown to use the *far-is-up* strategy: the farther the target, the higher the pointing arm is raised[2]. Mesh (2017), (Mesh, submitted) analyzed video recorded interviews in which Quiahije Chatino speakers located landmarks near their homes and in the surrounding landscape, and found that speakers used the far-is-up strategy consistently when indicating targets with a distance range of 200 m to 107 km from the interview site. For this study, all targets were conceptualized as "at the landscape scale" and the notion of scale itself was not further explored. Chin pointing was not investigated, and to date there is no analysis of the contexts of use for chin pointing among Quiahije Chatino speakers.

3. RESEARCH QUESTIONS

Prior studies of demonstrative use and pointing in the Quiahije community have laid the groundwork for a more focused study of multimodal indicating in usage. Moreover, the central role of direction-giving in traditional community practices and the resurgence of interest in these practices through language revitalization projects in the community make such a study especially urgent.

For the current study, we pose the following research questions:

1. Does the distance of the indicated target influence:
   (a) the choice of demonstratives, across scales?
   (b) the presence of chin pointing, across scales?
   (c) the presence of manual pointing, across scales?
   (d) the form (height) of manual pointing, across scales?
2. Is there a relationship between demonstrative choice and use of pointing:
   (a) with the chin, across scales?
   (b) with the hand, across scales?

4. CURRENT STUDY

4.1. Methods

4.1.1. Participants

Data for the current study were drawn from interviews performed with eight native speakers of Quiahije Chatino (four female). Speakers were recruited by native speaker research assistants on the basis of their near-exclusive use of Chatino, though all participants showed at least some passive knowledge of Spanish (demonstrated outside of interviews, as participants heard questions posed by the first author in Spanish and responded in Chatino without waiting for interpretation). Interviews were performed in Quiahije Chatino by a native speaker of the language who is also fluent in Spanish, allowing for direct communication with the first author. Consent was obtained from all participants to use their research data, and many participants additionally gave permission to make their recorded image available to the public. Demographic information for all participants, including age, gender, languages used, and education level, is provided in **Table 2**.

4.1.2. Procedure

4.1.2.1. Interview Design

Participants took part in an interview performed in six preselected stops along a 2.2-km walking trail to the peak of kyqya^{C} kcheq^{B} ("Thorn Mountain"), a location of religious and cultural significance to Chatino people in and outside of San Juan Quiahije. During the interview, participants discussed the role of six preselected stops on the trail in the annual religious pilgrimage performed by members of multiple Chatino communities. They also identified ten towns of importance in the surrounding district, and four towns vital to trade with communities in the larger state of Oaxaca. In keeping with our large-scale theme, targets prompted in the interviews involved named places. Such targets represent a class of sizeable and spatially stable entities of high sociocultural salience and interactional significance, as well as obvious relevance at the landscape scale (cf. Blythe et al., 2016). They were therefore deemed particularly suitable for our purposes. The locations of the interview stops, and the places to be discussed in each

[1]In this and all examples to follow, the participant-anchored proximal forms are translated to English as *this* or *here* while the neutral demonstrative is translated as *that* or *there*. We provide these translations to facilitate understanding the sentence meaning, while nevertheless cautioning the reader that there is no translation equivalent for the Quaihije Chatino demonstratives in English.

[2]This strategy was also found in users of San Juan Quiahije Chatino sign language, a young sign language emerging in the Quiahije municipality (Mesh, 2017; Cooperrider and Mesh, in press).

interview, were selected to elicit indicating behaviors with search domains at three scales:

- **Activity:** participants anticipated, and later reviewed, each of the six stops along the 2.2-km walking trail.
- **District:** participants discussed six towns at distances between 1.2 and 11 km from the walking trail.
- **State:** participants discussed four towns/cities in the state of Oaxaca, at distances between 16 and 108 km from the walking trail.

Figure 1 presents the trail with the full set of 16 targets. Targets at the activity, district, and state scales are distinguished by the color and style of their placemarks. Each of the search domain scales can be defined as a span of distance from the speech dyad (i.e., the interviewer and participant). The scales can also be distinguished by the general characteristics of the search domains, as presented in **Table 3**.

All 16 targets were discussed at each stop, as participants (in the role of senders) were prompted to provide to the interviewer (in the role of addressee): (1) the name of the current trail stop

TABLE 2 | Participant information.

Participant No.	Gender	Age	L1	L2	Education
SJF13	F	67	Quiahije Chatino	Passive Spanish	None
SJF14	F	62	Quiahije Chatino	Passive Spanish	None
SJF15	F	63	Quiahije Chatino	Passive Spanish	None
SJF16	F	64	Quiahije Chatino	Passive Spanish	None
SJM11	M	56	Quiahije Chatino	Passive Spanish	Primary School
CM06	M	66	Quiahije Chatino	Passive Spanish	None
CM11	M	66	Quiahije Chatino	Passive Spanish	None
CM12	M	69	Quiahije Chatino	Spanish	Primary School

TABLE 3 | Scales for search domains, with their characteristics.

Scale	Distance (km)	General characteristics
Activity	<2.2	Area of ongoing interview activity, within the speech dyad and beyond
District	1.2–11	Local area of Chatino residence/identity
State	>11	Regional area of Chatino identity/trade

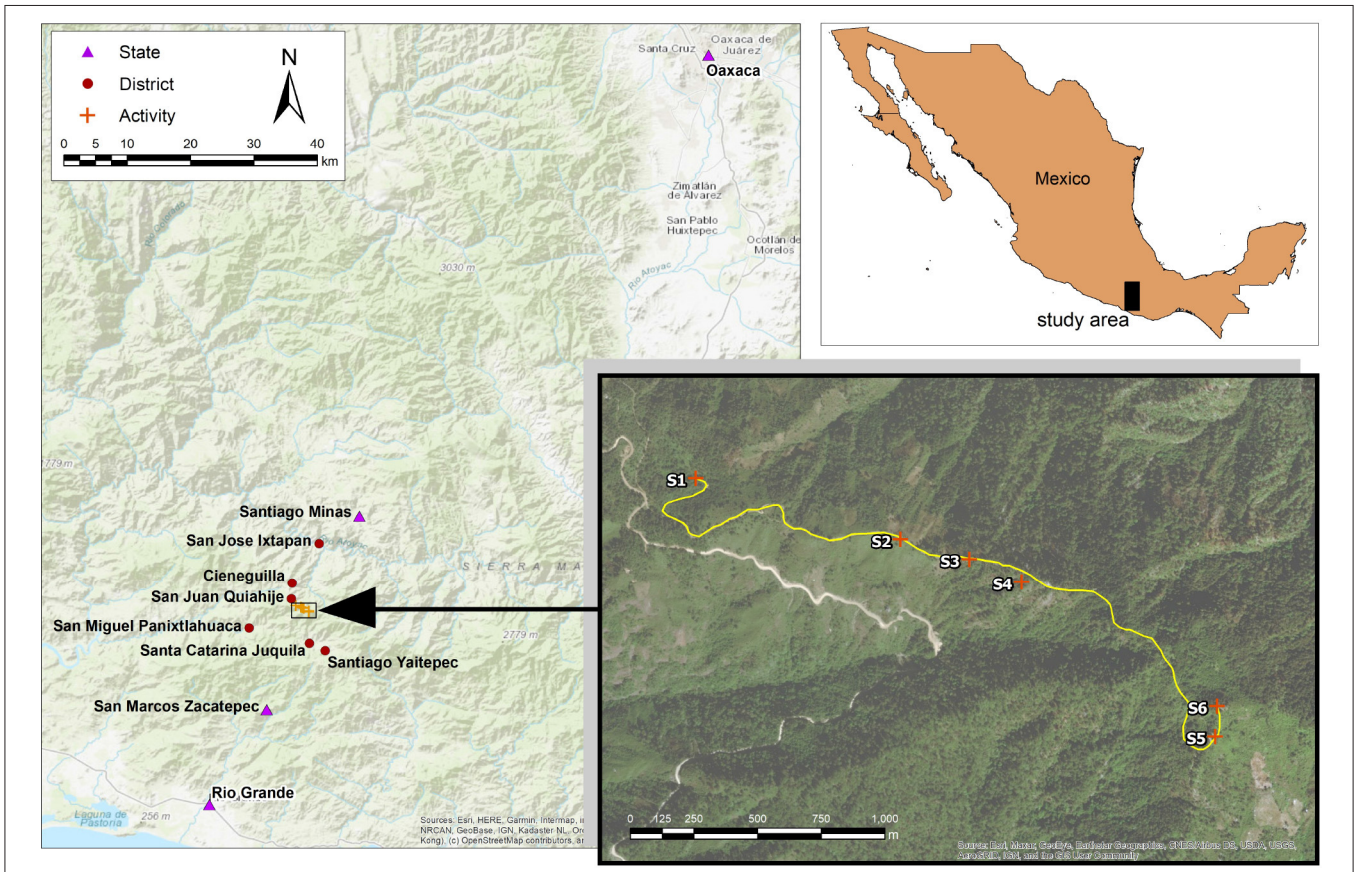


FIGURE 1 | Targets discussed by interview participants at six different locations along the trail.

and its origins, (2) the names of the towns visible from the stop and their origins, and (3) the rough direction of all targets (minus the current stop), as gauged from the current stop. The full interview protocol, including the walk to the peak of kyqya^{C} kcheq^{B} (but excluding the subsequent descent) required a total of 3 h, ∼60 min of which were spent performing interviews at the preselected stops along the route. The full interview protocol has been made available with the **Supplementary Material** for this paper.

4.1.2.2. Recording Procedure

All interviews were video recorded from two perspectives (giving front and side views of the participants) using Garmin Virb action cameras. The interviewer and participant each wore a head-mounted Røde HS2 headset microphone connected to a Røde Wireless Go transmitter with a receiver that attached directly to one of the action cameras. Digital video was recorded by the first author and a trained research assistant, neither of whom participated in the interview. Video was shot in MP4 format with a video mode of 1080p and a frame rate of 30 fps. The Virb action cameras additionally collected geoinformation, producing a GPX file containing the coordinates of each camera (collected at a rate of 10 Hz).

4.1.3. Dataset

Since each participant was recorded during interviews at six locations along the trail, the dataset consisted of a total of 48 video recordings. The recordings ranged in length from 2:59 to 8:24 min (M = 5:34 SD = 1:14). We excluded recordings from the first trail stop, treating the interviews recorded there as a training activity in which participants were familiarized with the task of indicating 16 targets. This left a dataset of 42 recordings for analysis.

4.1.4. Data Treatment and Coding

The audio tracks from both cameras were combined to produce a single integrated sound file in WAV format, and the video recording start times were synched, using Adobe Premier. The digital video and audio files were transcribed, translated and coded using frame-by-frame analysis, performed in the video annotation software, ELAN (2020).

For this study, the unit of analysis was the *indicating act*, defined as all behaviors—spoken and gestural—that occurred during a stretch of speech in which a demonstrative expression was used, ±1 s. Speech was used as the point of entry to the data: demonstrative expressions in all three formal contexts (pronominal, adnominal, adverbial) were identified, and the prosodic units in which they occurred were assigned to indicating acts. By definition, then, all indicating acts contained at least one demonstrative expression. An indicating act could additionally contain one or multiple pointing gestures, produced by jutting the chin or extending a hand/arm.

4.1.4.1. Speech

4.1.4.1.1. Transcription. Three research assistants (native speakers of Quiahije Chatino) with experience writing the language watched the video recorded interviews and identified every reference to our pre-selected targets. They identified the first three cases where a demonstrative expression was used to indicate each target. They then identified the *breath unit* surrounding each demonstrative expression, defined as the stretch of phonation visible on a waveform viewer, bounded on both sides by a lack of phonation (cf. Lieberman, 1967). They transcribed all talk in each breath unit, following the orthographic conventions of Cruz (2011), and produced a corresponding translation to Spanish (the language of communication between the research assistants and the first author).

4.1.4.1.2. Speech coding. Each indicating act was coded for the demonstrative form it contained. For those indicating acts that encompassed multiple demonstratives with the same target, only the first demonstrative was coded, to preserve the independence of the data points. If an indicating act contained two demonstratives with different targets, e.g., "Sour rock is *here* and Turkey Breast Rock is *there*," the speech was reanalyzed into two separate indicating acts, and each was assigned a code for demonstrative form. This resulted in a set of 883 indicating acts in total.

4.1.4.2. Gestures

4.1.4.2.1. Gesture identification. All gesture coding was performed with the audio switched off, and with transcriptions and translations hidden. This ensured that coders had no access to the content of the speech in the recordings.

Gestures with strokes that occurred inside the boundaries of an indicating act were identified, and assigned to the corresponding indicating act. To do this, the first author proceeded frame-by-frame, first identifying all manual gesture units (from the onset of a spatial excursion of the fingers, hand and/or arm to the assumption of a rest position) as well as head movements that might constitute a deictic chin point. The first author then identified the stroke phase within each of the identified manual gesture units. The boundaries of the stroke were identified via changes in the velocity of the hand movement (such as when the movement slowed, or stopped altogether in the case of static strokes) and/or changes in the handshape (cf. Kendon, 1972; Kita et al., 1998; Seyfeddinipur, 2006).

When no stroke boundary could be identified using these criteria, the stroke was identified in the frame(s) in which the articulators (fingers, hand, and/or arm) were at the point of fullest extension. Self-regulators (gestures touching body or face, cf. Ekman and Friesen, 1969) and gestures with a possible "pragmatic" function (such as conveying the speaker's epistemic stance toward their statement, often diagnosed via palm-up gesture forms, cf. Kendon, 2004; Müller, 2004) were excluded from analysis[3].

4.1.4.2.2. Gesture coding. Each gesture contained within an indicating act was coded as C (chin point), M (manual point), or

[3]Although palm-up gestures can and often do serve deictic functions, we chose to err on the side of caution when creating our form-based exclusion criteria, and removed all gestures with an upturned palm. Future research will examine the full range of deictic handshapes used by Chatino speakers.

CM (chin point and manual point). If multiple manual gestures or multiple chin points occurred within a single indicating act, the first token of each gesture type was coded. Of the 882 indicating acts identified, 68 contained a chin point, 416 contained a manual point, and 8 contained both pointing types.

One formational feature of the manual points was further coded: the elbow height of the arm during the articulation of each stroke was coded as low (below shoulder) or high (at or above shoulder). A first coder coded elbow height in the full dataset, while a second coder, assigning height values to a set of pre-identified strokes, coded one randomly selected video recording from each participant (∼17% of the dataset). We computed inter-rater reliability measures (Hallgren, 2012) using R version 3.6.1 (R Core Team, 2019) with the irr package (Gamer et al., 2019), and found that the two coders showed agreement in 93% of cases (Cohen's kappa = 0.85).

4.1.4.3. Target Distance

Each indicating act was assigned a distance measure, reflecting the geodesic distance in meters between the interview site and the target location. Geodata (latitude–longitude pairs stored in Garmin's proprietary GMetrix file format) were extracted from a single interview, and one representative latitude–longitude pair was identified at the approximate center of each of the interview stops along the trail. A latitude–longitude pair was also identified at the approximate center of each of the off-trail targets, allowing for the distance between the interview location and target location to be measured in a geographic information system (GIS).

Table 4 presents distance measures for each target discussed in the interviews. Targets are identified by their Chatino names, as well as by conventional placenames assigned by Spanish speakers (or, in the absence of these conventions, by a translation from Chatino to English). The reported distance range represents the minimum and maximum possible distances between the target and the stops along the trail.

TABLE 4 | Walking interview targets: scale, names, and distance range.

Scale	Name, Quiahije Chatino	Name, Translated	Distance (km)
Activity	ntenq^{F} tiyuq^{G}	Plain of the Spring	0–2.2
	tu^{C} kchi^{C}	Mountain Pass	0–1.4
	ke^{A} tiyeq^{B}	Sour Rock	0–1.2
	ke^{A} ku^{E} suq^{C}	Turkey Breast Rock	0–1.3
	lo^{A} si^{K} kyqya^{C} kcheq^{B}	Petition Monument	0–2.2
	kyqya^{C} kcheq^{B}	Peak of Thorn Mountain	0–2.2
District	kchin^{A}	San Juan Quiahije	1.3–3.4
	ntenq^{F}	Cieneguilla	3.6–5.2
	tqwa^{A} tyku^{E}	San José Ixtapan	10.4–11.0
	skwi^{E}	San Miguel Panixtlahuaca	7.9–9.4
	sqwe^{F}	Santa Catarina Juquila	4.9–6.3
	ke^{G} xin^{E}	Santiago Yaitepec	6.6–8.4
State	se^{A} na^{A} nya^{K}	Santiago Minas	16.8–17.1
	tsi^{C}	San Marcos Zacatepec	16.7–1.07
	ntenq^{F} tyku^{E} jlyu^{B}	Rio Grande	33.9–34.2
	lo^{A} ntqa^{B}	Oaxaca City	106.6–107.0

4.1.4.4. Data Exclusion

A total of 283 coded indicating acts were removed from the dataset for the following reasons. Indicating acts containing the discourse-anaphoric demonstrative *kanq*^{G} ($n = 120$) were excluded in order to narrow the dataset to cases of exophoric reference (i.e., reference to concrete, physical entities in the space around the speech dyad). Indicating acts containing the addressee-anchored proximal forms *kwa*^{J} and *kwa*^{E} ($n = 24$) were removed because their infrequent occurrence did not support a statistical analysis. Indicating acts containing demonstratives with multiple targets (e.g., "Sour Rock and Turkey Breast Rock are *there*") could not be assigned a single measure for target distance. This was also the case for indicating acts in which a single pointing gesture had multiple targets (because it accompanied a demonstrative expression with multiple targets, or because it extended across multiple demonstrative expressions with discrete targets). All such indicating acts ($n = 96$) were excluded. Indicating acts that contained speech with segment-by-segment route directions ($n = 38$) were excluded, since in these cases speakers often used demonstratives in set phrases roughly equivalent to *from there we go on*, and it is unclear whether these cases are comparable to other indicating speech. Indicating acts containing two pointing types ($n = 4$) were excluded to simplify the analysis.

After these exclusions, the dataset for analysis comprised a total of 601 indicating acts, all with an exophoric function. Of these, all contained a demonstrative, 35 contained an additional chin point, and 256 contained an additional manual point. Notably, after exclusions the dataset contained only three demonstrative forms: the speaker-anchored proximal forms *nde*^{C} or *re*^{C}, which we treat jointly in our analysis, and the unmarked/neutral form *kwa*^{F}. We hereafter refer to these forms as the "speaker-proximal" and "neutral" forms.

4.1.5. Data Analysis

The goal of the analysis was to test whether target distance influenced how multimodal indicating was performed within and across three scales—activity, district, and state.

We treated distance in two ways for our analyses. For descriptive tables, we subdivided the distance range within each scale into bins. For the activity scale, we created four bins of 0–541, 542–1,082, 1,083–1,623, and 1,624–2,206 m. For the district scale, we created four bins of 0–2,751, 2,752–5,502, 5,503–8,253, and 826–11,004 m. For the state scale, we created three bins spanning the distances where our targets clustered, with spans of 0–19,000, 19,001–36,000, and 36,001–108,000 m). This treatment of distance as categorical allowed us to present descriptive statistics in terms of the distribution of demonstrative forms and pointing use across distance categories.

For statistical analyses, we took a different approach. Within each scale, the actual distance values (in m) were rescaled to values from 0 (i.e., the minimal distance within the scale) to 1 (i.e., the maximal distance within the scale). This transformation

leaves the relative differences between the values within each scale intact. At the same time, it facilitates the comparison of the distance effects across the three scales because the estimated effects (regression coefficients) receive an equivalent interpretation (i.e., whether the change in occurrence of the outcome variable at the maximal distance within a scale compared with that at the minimal distance within a scale).

We performed six separate statistical analyses to answer the research questions. In the first four analyses, we looked at the effects of distance and scale on choice of demonstrative form (speaker-proximal or neutral), the presence of a chin point, the presence of a manual point, and the height of a manual point (low or high). Since we wanted to know whether the effect of distance varied across scales, we primarily looked at the interaction of these two predictors. If this interaction was significant, we tested the individual effects of distance within the activity, district, and state scales (simple main effects). If the interaction effect was not significant, it was removed from the analysis model, to see whether any of the remaining effects were significant.

In the final two analyses, we looked at the effect of choice of demonstrative on the presence of a chin point and on the presence of a manual point, again within each of the three scales. The procedure for testing these two effects was similar as the one for the first four analyses: We primarily looked at the interaction of choice of demonstrative and scale. If this interaction was significant, then we looked at the simple main effects of demonstrative within each of the three scales. Otherwise, we removed it from the analysis to see whether any of the other remaining effects was significant.

All six analyses were mixed effects logistic regression models with scale and distance (analyses 1–4) or scale and choice of demonstrative (analyses 5 and 6) as fixed factors, and participant as a random factor. In the results section, we provide estimates (*EST*), standard errors (*SE*), *z*-values, and *p*-values for the effects that are most relevant for the research questions. All *p*-values for simple main effects have been corrected for multiple comparisons (Dunnett's method). A list of the six regression models (fixed effects parts only) is given in **Appendix 3**. The analysis was performed in R version 3.6.1 (R Core Team, 2019) using the packages lme4 (Bates et al., 2015) and multcomp (Hothorn et al., 2008).

4.2. Results

4.2.1. Sample Description

The dataset consisted of 601 indicating acts, of which 235 had targets at the Activity scale, 238 had targets at the District scale, and 128 had targets at the state scale. Every indicating act contained one demonstrative expression—using a speaker-proximal form (re^C or nde^C) or a neutral form (kwa^F)—while 35 contained an additional chin point, and 255 contained an additional manual point.

Across participants, manual points and the two demonstrative forms were used to refer to all 16 targets, but chin points were used to refer to 11 targets only. Across targets, all participants used the two demonstrative forms and every participant used manual points at least six times and chin points at least twice. There was natural variation across targets and participants in the frequencies of demonstratives and indicating strategies. For example, manual points comprised 95% of the pointing gestures of some participants (with chin points accounting for the other 5%), while for other participants, manual points comprised 60% of their pointing gestures (with chin points accounting for the remaining 40%). The distribution of indicating strategies and indicating forms, across targets and across participants, is presented in **Appendix 2**.

4.2.2. Effect of Distance on Demonstrative Choice

Our research question (1a) asked whether the distance of the indicated target influences demonstrative choice, and whether this effect is found across multiple scales. The distribution of demonstrative forms across distance categories is presented in **Table 5**. This distribution suggested that participants were more likely to use a speaker-proximal demonstrative form when the target was closer to the speech dyad, but only for targets at the activity scale. The interaction between distance and scale was significant (see the **Appendix 3.1**), and subsequently pursued by an analysis of simple main effects (i.e., the effect of distance within each level of scale). A significant simple main effect of distance was found only in the activity scale: participants were more likely to use the speaker-proximal form when the target was relatively near to them on the trail and less likely to use this form as the distance to the activity scale targets increased ($EST = 2.832$, $SE = 0.520$, $z = 5.452$, $p = 0.000$). No effects were found at the district scale ($EST = 0.611$, $SE = 0.476$, $z = 1.282$, $p = 0.488$) or at the state scale ($EST = 0.255$, $SE = 0.406$, $z = 0.629$, $p = 0.896$).

4.2.3. Effect of Distance on the Use of Chin Pointing

Our research question (1b) asked whether the distance of the indicated target influences the use of chin pointing, and whether this effect is found across multiple scales. The distribution of chin pointing (vs. its absence) across distance categories is presented in **Table 6**. The descriptive results reflect the relatively small number of chin points produced in the study: only 35 in total. With such a small number of cases, we would be unlikely to find a strong relationship between the distance of the target and the use of chin pointing. The results of the analysis showed no significant joint effect of distance and scale, nor any significant effect of distance or scale when used as individual predictors (see **Appendix 3.2**).

TABLE 5 | Raw frequencies (with proportions in parentheses) of demonstrative forms across distance categories at the activity, district, and state scales.

Dist. Cat.	Activity				District				State			
	re^C/nde^C		kwa^F		re^C/nde^C		kwa^F		re^C/nde^C		kwa^F	
1	117	(0.85)	21	(0.15)	12	(0.75)	4	(0.25)	34	(0.54)	29	(0.46)
2	21	(0.46)	25	(0.54)	43	(0.61)	27	(0.39)	9	(0.38)	15	(0.62)
3	11	(0.35)	20	(0.65)	36	(0.59)	25	(0.41)	19	(0.46)	22	(0.54)
4	10	(0.50)	10	(0.50)	53	(0.58)	38	(0.42)				
SUM	159	(0.68)	76	(0.32)	144	(0.61)	94	(0.39)	62	(0.48)	66	(0.52)

4.2.4. Effect of Distance on the Use of Manual Pointing

Our research question (1c) asked whether the distance of the indicated target influences the use of manual pointing, and whether this effect is found across multiple scales. The distribution of manual pointing (vs. its absence) across distance categories is presented in **Table 7**. The descriptive results suggested that the distance of the target did not influence whether participants pointed at the district and state scales. Only in the activity scale did the descriptive data suggest an effect of distance: here it appeared that participants were more likely to use a manual point when the target was farther from the speech dyad. The interaction between distance and scale was significant (see the **Appendix 3.3**), and subsequently pursued by an analysis of simple main effects (i.e., the effect of distance within each level of scale). The analysis showed a significant main effect of distance only for the activity scale: participants were least likely to use the manual point when the target was nearest to the deictic center, and more likely to point with the hand as the distance to the activity scale targets increased ($EST = 1.606$, $SE = 0.487$, $z = 3.296$, $p = 0.000$). No effects were found at the district scale ($EST = -0.205$, $SE = 0.478$, $z = 0.428$, $p = 0.964$) or at the state scale ($EST = 0.157$, $SE = 0.418$, $z = 0.374$, $p = 0.975$).

4.2.5. Effect of Distance on the Elbow Height of Manual Points

Our research question (1d) asked whether the distance of the indicated target influences the form (height) of manual pointing, and whether this effect is found across multiple scales. The height values of manual pointing across distance categories are presented in **Table 8**. The descriptive results suggested that distance weakly influenced pointing height at the activity scale alone. The interaction between scale and height was significant (see **Appendix 3.4**) and pursued with an analysis of simple main effects. We found a marginally significant effect of distance within the activity scale: as targets increased in distance, participants were more likely to raise the elbow of the pointing arm at the activity scale ($EST = 1.985$, $SE = 0.907$, $z = 2.187$, $p = 0.084$). No effects were found at the district scale ($EST = 0.213$, $SE = 0.713$, $z = 0.299$, $p = 0.987$) or state scale ($EST = -0.420$, $SE = 0.676$, $z = -0.621$, $p = 0.899$).

Notably, the height of manual points appeared to shift between the scales, with low elbow predominating at the activity scale, and a high elbow at the district and state scales. To test this observation, we simplified the logistic regression model, using scale alone as a fixed factor and participant as a random factor. We found a significant main effect of scale: participants were more likely to point with a raised arm to targets at the district scale ($EST = 0.835$, $SE = 0.338$, $z = 2.468$, $p = 0.117$) and at the state scale ($EST = 1.784$, $SE = 0.431$, $z = 4.144$, $p = 0.000$), compared to the activity scale.

4.2.6. Relationship Between Demonstrative Form and Use of Chin Pointing

Our research question (2a) asked whether there is a relationship between demonstrative choice and use of pointing with the chin, and whether this effect is found across multiple scales. The distribution of chin pointing and demonstrative choice across distance categories is presented in **Table 9**. With just 35 observations of chin points in the dataset, we did not anticipate an analysis to reveal a strong relationship between the use of chin pointing and the choice of a speaker-proximal or distal demonstrative. The analysis showed no significant interaction between choice of demonstrative and scale (see **Appendix 3.5**).

4.2.7. Relationship Between Demonstrative Form and Use of Manual Pointing

Our research question (2b) asked whether there is a relationship between demonstrative choice and use of pointing with the hand, and whether this effect is found across multiple scales. The distribution of manual pointing and demonstrative choice across distance categories is presented in **Table 10**. The descriptive

TABLE 6 | Raw frequencies (with proportions in parentheses) of chin points at the activity, district, and state scales.

Dist. Cat.	Activity				District				State			
	Chin pt.		No		Chin pt.		No		Chin pt.		No	
1	7	(0.05)	131	(0.95)	2	(0.13)	14	(0.87)	0	(0.00)	63	(1.0)
2	1	(0.02)	45	(0.98)	5	(0.07)	65	(0.93)	4	(0.16)	20	(0.84)
3	4	(0.12)	27	(0.78)	4	(0.07)	57	(0.93)	3	(0.07)	38	(0.93)
4	1	(0.05)	19	(0.95)	4	(0.04)	87	(0.96)				
SUM	13	(0.05)	222	(0.95)	15	(0.07)	223	(0.93)	7	(0.05)	121	(0.95)

TABLE 7 | Raw frequencies (proportion) of manual points at the activity, district, and state scales.

Dist. Cat.	Activity				District				State			
	Manual pt.		No		Manual pt.		No		Manual pt.		No	
1	30	(0.22)	108	(0.78)	8	(0.50)	8	(0.50)	33	(0.52)	30	(0.48)
2	18	(0.39)	28	(0.61)	41	(0.59)	29	(0.41)	11	(0.46)	13	(0.54)
3	12	(0.39)	19	(0.61)	32	(0.52)	29	(0.48)	22	(0.54)	19	(0.46)
4	9	(0.45)	11	(0.55)	40	(0.44)	51	(0.56)	0	(0.00)		
SUM	69	(0.29)	166	(0.71)	121	(0.51)	117	(0.49)	66	(0.51)	62	(0.49)

TABLE 8 | Raw frequencies (proportion) of elbow height values for manual points at the activity, district, and state scales.

Dist. Cat.	Activity				District				State			
	Low		High		Low		High		Low		High	
1	21	(0.70)	9	(0.30)	4	(0.50)	4	(0.50)	8	(0.24)	25	(0.76)
2	11	(0.61)	7	(0.39)	14	(0.50)	24	(0.50)	1	(0.09)	10	(0.91)
3	5	(0.42)	7	(0.58)	12	(0.38)	20	(0.62)	7	(0.32)	15	(0.68)
4	3	(0.33)	6	(0.67)	16	(0.42)	22	(0.58)				
SUM	40	(0.58)	29	(0.42)	46	(0.40)	70	(0.60)	16	(0.24)	50	(0.76)

results suggested a relationship between the use of a manual point and the use of a speaker-proximal demonstrative form. The interaction between choice of demonstrative and scale was marginally significant (see **Appendix 3.6**). In addition, the removal of the interaction did not result in a significantly worse model ($\chi^2 = 3.549$, $df = 2$, $p = 0.170$). In this model without the interaction, both scale and demonstrative form showed a significant relationship with manual points. There were overall more manual points with *nde*^{C}/*re*^{C} than with *kwa*^{F} ($EST = -0.873$, $SE = 0.195$, $z = -4.469$, $p = 0.000$), and, compared to the activity scale, this effect was stronger at the district scale ($EST = 1.099$, $SE = 0.208$, $z = 5.262$, $p = 0.000$) and at the state scale ($EST = 1.262$, $SE = 0.251$, $z = 5.024$, $p = 0.000$), compared to the activity scale.

4.3. Discussion

4.3.1. Target Distance Influences Demonstrative Choice and Manual Pointing, Only in Activity Scale Space

For this study, we defined three scales for the search domains of indicating acts. The scales differed in their spatial extent and in other characteristics, as described in **Table 3**. We asked whether the factor of target distance would have an effect on multimodal indicating behaviors, and, if the effect were present, whether it would be the same across the three scales.

A prominent finding from this study is that distance had an effect on indicating behaviors at only one scale. The activity scale—the smallest scale in the study design—was the one at which participants showed sensitivity to target distance, both in their demonstrative choice and in their use and modulation of manual pointing.

4.3.1.1. Demonstrative Choice

Participants were significantly more likely to use a speaker-proximal demonstrative when activity scale targets were near them. As the distance to the target increased, so did the likelihood that participants would use the neutral demonstrative form. At the district and state scales, by contrast, the speaker-proximal and neutral forms were used with near-equal frequency: there was no significant effect of distance on the choice of demonstrative forms. We illustrate these findings below, with examples from the video data.

In Example 3, a participant stands at Petition Monument, the fifth stop on the kyqya^C kcheq^B trail. The final stop on the trail lies 120 m ahead, through a wooded path. The participant uses the speaker-proximal demonstrative to indicate the stop (3a). Later, the participant explains that she and the interviewer stopped at every landmark on the trail, and indicates the farthest one with the neutral demonstrative (3b).

(3) a. Activity scale search domain (target: peak of kyqya^C kcheq^B, distance = 120 m)

nde^C nga^J no^I nga^J kiqya^C
DEM:1 BE.3S.HAB RELVZ BE.3SG.HAB MOUNTAIN
kcheq^B BE.3S.HAB DEM:1 ADVZ
THORN DEM:1 ADVZ

"here it is, is Thorn mountain, here"
20191215_R09-P06, 00:00:17

b. Activity scale search domain (target: Plain of the Spring, distance = 2206 m)

qo^E kwiq^J no^A nga^J neq^C ykwaq^A
AND ALSO RELVZ BE.3S.HAB SWAMP INSIDE.3SG
neq^F-tiyuq^C kwa^F no^J nga^J
STOMACH.3SG-MARSH DEM:N RELVZ BE.3SG.HAB

"and also at the marsh, over there"
20191215_R09-P06, 00:05:00

In considering the study results for demonstrative choice, we take note of the contrast between the activity scale and the two larger scales operationalized for the study. We observe that our participants showed sensitivity to target distance in activity scale space in much the same way as participants in a variety of experimental studies have done when indicating targets in "interaction-scale" or tabletop space. We interpret this as evidence that the demonstrative oppositions long explored in smaller-scale space *can* scale up—though not without limit. Our results suggest that distance effects on oppositions at smaller scales disappear at larger ones.

It is noteworthy that the point at which the distance effect disappears in our study coincides with the outer perimeter of the hiking activity itself. This suggests that the participants' conception of the target as co-present with both sender and addressee in a shared activity space is central to the use of demonstrative oppositions—an explanation that has been favored in accounts of social factors driving demonstrative choice (especially Enfield, 2003, 2018). Our results at the activity scale suggest that distance exerts some influence on demonstrative choice, though this influence may well be conditioned, or even eclipsed, by other social-pragmatic factors.

TABLE 9 | Raw frequencies (proportion) of demonstrative forms and chin points at the activity, district, and state scales.

Dem. choice	Activity				District				State			
	Chin pt.		No		Chin pt.		No		Chin pt.		No	
nde^C/*re*^{C}	6	(0.04)	153	(0.96)	6	(0.04)	138	(0.96)	4	(0.06)	58	(0.94)
kwa^F	7	(0.09)	69	(0.91)	9	(0.10)	85	(0.90)	3	(0.05)	63	(0.95)
SUM	13	(0.06)	222	(0.94)	15	(0.06)	223	(0.94)	7	(0.05)	121	(0.95)

TABLE 10 | Raw frequencies (proportion) of demonstrative forms and manual points at the activity, district, and state scales.

Dem. choice	Activity				District				State			
	Manual pt.		No		Manual pt.		No		Manual pt.		No	
nde^C/*re*^{C}	50	(0.31)	109	(0.69)	82	(0.57)	62	(0.43)	41	(0.66)	21	(0.34)
kwa^F	19	(0.25)	57	(0.75)	39	(0.41)	55	(0.59)	25	(0.38)	41	(0.62)
SUM	69	(0.29)	166	(0.71)	121	(0.51)	117	(0.49)	66	(0.52)	62	(0.48)

4.3.1.2. Manual Pointing: Presence and Form

Distance influenced two aspects of manual pointing to targets at the activity scale. First, the distance of the target influenced the *presence* of a manual point: participants were more likely to point to targets using their hand when targets were farther from them. In addition, distance had a marginal influence on the *form* of the manual point: for targets farther away, participants were more likely to raise the pointing arm until the elbow was at the level of the shoulder or above.

In Example 4, the participant stands at the Petition Monument, the fifth stop on the kyqya^{C} kcheq^{B} trail. She indicates the nearest trail stop using a low elbow alongside the speaker-proximal demonstrative, *nde*^{C} (Example 4a, **Figure 2**). When describing the stop at Turkey Breast Rock 950 m away, she indicates the more distant location using a pointing gesture with a high elbow, alongside the neutral demonstrative *kwa*^{F} (Example 4b, **Figure 3**). Notably, this distant target is at a lower altitude than the speaker, making her pointing form interpretable only as an application of the far-is-up schema for encoding distance[4].

(4) a. *Activity scale search domain (target: Petition Monument, dist. = 120 m)*[5]

nde^{C} xka^{I} **ti^{A}** **re^{C}** kiqya^{C} lo^{E}
DEM:1 ONE ADVZR DEM:1 MOUNTAIN RELVZ
la^{E} ti^{J}…
SEE.3P.HAB ADVZR

"here, toward here, is the only other visible one…"
20191207_R07-P05, 00:05:17

b. *Activity scale search domain (target: Turkey Breast Rock, dist. = 950 m)*

ti^{H} kwa^{F} nga^{J} **kanq^{G}** **ni^{C}**
LOC DEM:N BE.3SG.HAB DEM:D NOW

"it's there, it is now"
20191207_R07-P05, 00:00:40

Here again, we see a pattern at our smallest distance scale that parallels observed patterns in "interaction-scale" space. In the laboratory and on the trail, participants are more likely to point to targets when they are farther away (cf. Bangerter, 2004; Cooperrider, 2015), and show sensitivity to the distance of the target by modulating the form of the point itself. Again, this sensitivity appears to be bounded. Neither the presence nor the form of a manual point appears to be influenced by distance beyond what, for our study, amounts to activity scale space.

4.3.2. Pointing Form Is a Cue to Scale Itself

There was also a strong effect of scale itself on the height of the pointing gestures. For targets at the activity scale, distance did prompt raising of the pointing arm, yet participants were more likely overall to point with a lowered arm. For district scale targets, participants were more likely to point with a raised arm, and likelier still to do so when indicating state scale targets. These findings are illustrated in the examples that follow.

FIGURE 2 | Activity scale [target: Peak of kyqya^{C} kcheq^{B}].

FIGURE 3 | District scale [target: ke^{A} ku^{E} suq^{C}].

In Example 5, the participant is standing at Sour Rock, the third stop on the kyqya^{C} kcheq^{B} trail. When asked about Plain of the Spring, an activity scale target, she indicates its location using the speaker-proximal demonstrative *nde*^{C} and a manual point with a low elbow (Example 5a, **Figure 4**). To locate Santa

[4]For further evidence that pointing form in San Juan Quihaije is influenced by target distance, and not by target altitude, see (Mesh, submitted).

[5]In this and all other examples in which speech and a pointing gesture overlap, the speech co-occurring with the stroke of gesture appears in boldface type.

FIGURE 4 | Activity scale [target: Plain of the Spring].

Catarina Juquila, a district scale target, she indicates the town using the proximal demonstrative *nde*^{C} and a point with a raised elbow (Example 5b, **Figure 5**). When indicating Rio Grande, a state scale target, she uses the neutral demonstrative *kwa*^{F} and a point with a raised elbow (Example 5c, **Figure 6**).

(5) a. Activity scale search domain (target: Plain of the spring)

kwiq^{J} niyan^{J} ndwa^{K} qya^{K} re^{C} no^{H}
ALSO STRAIGHT LIE.3SG.HAB DOWNHILL DEM:1 RELVZ
janq^{G} **ni^{C}**
PRO.3SG.INDEF NOW

"it's also straight down here, it is now"
20191130_R04-P03, 00:00:57

b. District scale search domain (target: Santa Catarina Juquila)
qin^{K} nde^{K} ti^{K} **ntqen^{J}** sqwe^{F}
NEAR DEM:1 ADVZR LIE.3SG.PROG JUQUILA

"here, closeby, is Juquila"
20191130_R04-P03, 00:02:40

c. State scale search domain (target: Rio Grande)
ntenq^{F}-tyku^{E}-jlyu^{B} **ni^{A}-ndwa^{B}** **kwa^{F}**
VALLEY-RIVER-BIG STRAIGHT-BE.3SG.PROG DEM:N
no^{H} **janq^{G}** **ni^{C}**
LIE.3SG.POT PRO.3SG.INDEF NOW

FIGURE 5 | District scale [target: Santa Catarina Juquila].

"Rio Grande, it's straight over there now"
20191130_R04-P03, 00:01:26

Importantly, the pattern we see with manual pointing is quite distinct from the pattern with demonstrative choice. At larger scales, the two demonstrative forms are used with near-equal frequency, suggesting that factors other than distance exercise a greater influence at those scales. By contrast, manual points are produced with a raised elbow significantly more often at larger scales. Thus, pointing form provides cues to the scale of the search domain in a way that demonstrative form does not.

4.3.3. Distance and Scale Influence How Demonstratives and Points Are Co-organized

One phenomenon that we investigated showed distance effects at all scales. This was the co-organization of demonstrative forms and pointing types.

When speakers used a demonstrative expression with a chin point, they showed a marginal preference for the neutral demonstrative form. This preference was influenced by target distance: the farther the target from the speech dyad, the stronger the preference. It was also influenced by target scale, as the trend was weaker at the largest of the study scales. In a notable contrast, when speakers paired a demonstrative with a manual point,

FIGURE 6 | State scale [target: Rio Grande].

FIGURE 7 | Dem. + chin point [target: Oaxaca].

they showed a strong preference for using a speaker-proximal demonstrative form. Again, this preference was influenced by target distance: the farther the target from the speech dyad, the stronger the preference. In this case, the preference grew stronger as the scale size increased. We illustrate these findings in the following examples.

In Example 6, the speaker is standing at the Petition Monument, the fifth stop on the kyqya^{C} kcheq^{B} trail. She indicates the city of Oaxaca, a state scale target, using a demonstrative with a chin point, and uses the neutral demonstrative form, *kwa*^{F} (Example 6a, **Figure 7**). When indicating the same state scale target using both a demonstrative and a manual point, she uses the speaker-proximal demonstrative form, *nde*^{C} (Example 6b, **Figure 8**).

(6) a. Demonstrative and chin point (target: Oaxaca)
ti^{C} qan^{E} janq^{G} ya^{A}-je^{K} no^{A} no^{A}
STILL GO.3SG.COMP PRO.3SG TRADE RELVZ RELVZ
MORE DEM:N PRO.3SG
MOREDEM:N PRO.3SG
"they were going for trade, more toward there"
20191207_R07-P05, 00:04:07

b. Demonstrative and manual point (target: Oaxaca)
nde^{C} ti^{J} jyaq^{F} nga^{J} janq^{G}
DEM:1 JUST CLF.MEASUREMENT BE.3SG.HAB PRO.3SG
ni^{C}
NOW
"it's right about there, now"
20191207_R07-P07, 00:03:22

At least one of the above patterns of co-organization has a parallel in smaller-scale space. In studies conducted in laboratory environments, speakers of Dutch and of American English showed a preference for pairing (speaker or speech dyad)-proximal demonstrative forms with manual points (Piwek et al., 2008; Cooperrider, 2011, 2015), though in the study where target distance was explored as a potential conditioning factor (Cooperrider, 2011, 2015) participants showed none of the sensitivity to distance that we see in our study results. In explaining the affinity of proximal demonstratives and pointing, Piwek et al. (2008) and Cooperrider (2011, 2015) focus on the contribution of the demonstrative to the multimodal indicating act, positing that the marked proximal form more "intensely" recruits the attention of the addressee in these constructions. Neither account is explicit about the role of the pointing gesture in these cases of more "intense" multimodal indicating.

Our study results provide a clue to the roles of both the demonstrative and the pointing gesture when they are coupled for more "intense" indicating. At the two largest scales operationalized for the study, we found that demonstratives ceased to participate in a distance-influenced oppositional paradigm, while pointing gestures remained informative about two dimensions of the search domain: its direction and distance. In exactly those contexts, we found the closest relationship between the speaker-proximal demonstratives and the manual

FIGURE 8 | Dem. + manual point [target: Oaxaca].

point. We propose, in line with Piwek et al. (2008) and Cooperrider (2011, 2015), that in this context the proximal demonstrative is indeed recruiting attention with greater intensity. We further suggest that the demonstrative is orienting visual attention not primarily to the target, but instead (and in some cases exclusively) to the more informative contribution of the speaker's gesturing body (for a similar suggestion, cf. Bangerter, 2004). Demonstratives have been shown to call visual attention to speaker's gestures that represent spatial features of a referent (such as its orientation in space, cf. Emmorey and Casey, 2001; Hegarty et al., 2005). Our findings suggest that demonstratives play a similar role in orienting speaker attention to the gesturing body, as well as the indicated target, during multimodal indicating.

If the speaker-proximal demonstratives draw attention to manual points, what role does the neutral demonstrative play alongside chin points? The picture is less clear here, simply because of the small number of data points we were able to collect and analyze for this study. Chin points have been proposed to occur with neutral demonstrative forms in contexts where the gesture is less informative (Enfield, 2001; Mihas, 2017; Cooperrider et al., 2018). In such contexts, the pointing gesture needs to provide few cues for delimiting the search domain, and the speaker may not expect the addressee to shift their full gaze to the gesturing body and it its attention-directing cues. This may well prompt the speaker to recruit the gaze of the addressee to the gesturing body less intensively. More research about the coordination of demonstratives with chin points will be necessary to further investigate this claim.

5. CONCLUSION

This study has systematically considered the influence of scale on multimodal indicating behaviors, a domain hitherto not investigated. By defining multiple scales within what has previously simply been described as "large-scale" or "geographic-scale" space, we have been able to distinguish between those patterns of indicating that are operational at all scales, and those that are constrained to usage in smaller-scale spaces.

Our first finding—that distance does not straightforwardly account for demonstrative choice, pointing use, or pointing form at larger scales—occasions the question of whether other factors may influence multimodal indicating across scales. More research is called for, in particular into such social-pragmatic factors as the attention of the speech act participants, and their conception of the target as being in or outside of a shared domain of activity.

Our second finding—that some features of the organization of demonstratives and points are present across all scales, and even stronger at larger scales—raises additional questions about how demonstratives and pointing gestures jointly function to manage attention. We have suggested here that manual pointing gestures are the most informative of the indicating behaviors when targets are in large-scale space and have proposed that demonstratives may recruit visual attention to manual and chin points primarily, allowing the points themselves to indicate the target location. This proposal prompts empirical questions about the gaze of the addressee in response to multimodal indicating. It also raises more fundamental questions about the sequencing of demonstratives and pointing gestures at indicating events across scales, as well as about the exact temporal alignment between the modalities, since any theory of their joint function relies on evidence from the temporal coordination of speech and gesture.

The combined findings have broader implications for research on the multimodality of language, as they underscore not only the distinct contributions of speech and gesture to the linguistic composite but also the dynamic nature of their interplay. In exploring how the scale of the search space influences the indicating event, we found yet another source of evidence for the intricate organization of multimodal expressions, and for the tailoring of that organization to specific contexts of language use.

DATA AVAILABILITY STATEMENT

Raw data underlying the conclusions made in this paper are publicly available in the Lund University Humanities Lab's Corpus Server (http://hdl.handle.net/10050/00-0000-0000-0004-1F68-A@view). Analytical materials, including interview protocols, manuals for data collection and coding, the resulting datasets, and the scripts used to perform the statistical analyses have been made available via the Texas Data Repository (https://doi.org/10.18738/T8/QHMQIY).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Swedish National Ethics Authority, Dnr

2019-04621. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Because much of the population under study for this research is not literate, written consent for study participation and data use was not obtained. Instead, we created video recordings of informed consent being given. Participants whose images appear in this paper gave informed consent for identifiable images of themselves to be published. The consent procedures were approved by the authorities of the San Juan Quiahije municipality, and their approval was recognized by the Swedish National Ethics Authority.

AUTHOR CONTRIBUTIONS

KM, MG, and NB contributed to the conception and experimental design of the study. KM and EC performed the experiments, collected the data, and helped to collate the data together with experimental assistants. JW and KM conducted the analyses. All authors contributed to the interpretation of the results and to the writing of the manuscript, and approved the final version of the manuscript for submission.

FUNDING

We gratefully acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 839074. NB acknowledges support from the Bank of Sweden Tercentenary Foundation, grant No. NHS14-1665:1.

ACKNOWLEDGMENTS

We warmly thank our interview participants Severa Canseco Baltazar, Severiana Baltazar Lorenzo, María Orocio Nicolas, María Apolonio Cortés, Ambrocio Baltazar Cruz, and three participants who preferred to remain anonymous. This work would not have been possible without Chatino speaking interviewers and transcribers: we thank research assistant María Elena Mendez Cortés for interviewing all participants and for transcription and translation, Claudia García Baltazar and Rosalía Baltazar Baltazar for transcription and translation, as well as Beatríz Baltazar Canseco, Tomás Cruz Cruz, and Julian Cruz Bautista for camera operation and note-taking during our walking interviews. We thank two reviewers for very helpful comments on a previous version of the manuscript; research assistant Nicolas Femia for reliability coding of gesture forms at Lund University; Giacomo Landeschi in the Lund University Humanities Lab for creating a map of our interview targets; and Henrik Garde, Peter Roslund, and Jens Larsson in the Lund University Humanities Lab for support with camera piloting, data cleaning, and data archiving. We gratefully acknowledge Lund University Humanities Lab and CIESAS (Centro de Investigaciones Superiores en Antropología Social) Mexico City.

Agradecemos sinceramente a las personas que participaron en nuestras entrevistas: Severa Canseco Baltazar, Severiana Baltazar Lorenzo, Maria Orocio Nicolas, Maria Apolonio Cortés, Ambrocio Baltazar Cruz y tres participantes que pidieron no ser nombrados. Este trabajo no hubiera sido posible sin el trabajo de los entrevistadores y transcriptores que hablan chatino: agradecemos a la asistente de investigación Maria Elena Mendez Cortes por entrevistar a todos los participantes y por la transcripción y traducción, a Claudia García Baltazar y a Rosalia Baltazar Baltazar por la transcripción y traduccón, así como a Beatriz Baltazar Canseco, a Tomás Cruz Cruz y a Julian Cruz Bautista por la operación de la cámara y la toma de notas durante nuestras entrevistas. También agradecemos el trabajo de nuestros colaboradores en Suecia. Agradecemos el apoyo del Laboratorio de Humanidades de la Universidad de Lund y al Centro de Investigaciones Superiores en Antropología Social (CIESAS-CDMX).

Nde^{C} tya^{F} wa^{G} xqwe^{F} qin^{J} nten^{B} no^{K} ntsaq^{F} qwa^{G} wra^{K} no^{K} ni^{E} chaq^{F} wa^{G} qin^{A}: Sebera^{J} Kanseko^{F} Baltasar^{F}, Seberyana^{J} Baltasar^{F} Lorenso^{F}, Liya^{G} Orosyo^{J} Nikolas^{F}, Liya^{B} Apoloniyo^{F} Korte^{F}, Ambrosiyo^{J} Baltasar^{F} Crus^{F} qo^{E} xna^{E} nten^{B} no^{J} ja^{A} la^{I} ngwa^{C} riq^{C} chaq^{G} kyaq^{G} ne^{G}. Jnya^{F} no^{A} nde^{C} ja^{A} la^{I} ngwa^{C} ran^{F} si^{K} no^{K} ja^{A} sqwi^{I} nten^{B} no^{K} ni^{K} chaq^{F} qo^{E} nten^{B} no^{K} nya^{B} chaq^{F} jnya^{J}, no^{A} ndywiq^{A} chaq^{F} jnya^{J}: tya^{F} wa^{G} xqwe^{F} qin^{J} nten^{B} no^{K} na^{E} son^{B}, Liya^{B} Lena^{J} Mende^{F} Korte^{F} chaq^{F} ni^{E} chaq^{F} qin^{J} tqa^{J} nten^{B} no^{K} ntsaq^{F} qo^{E} no^{A} nya^{B} ktyi^{K} chaq^{F} jnya^{J} qo^{E} ktyi^{C} chaq^{F} xlya^{K}, qin^{A} Kladiya^{J} Garsiqa^{J} Baltasar^{F} qo^{E} qin^{A} Rosaliya^{J} Baltasar^{F} Baltasar^{F} chonq^{G} chaq^{F} nya^{B} ktyi^{K} chaq^{F} jnya^{J} qo^{E} ktyi^{C} chaq^{F} xlya^{L}, kwiq^{J} kwan^{H} niyan^{J}, qin^{A} Beqatris^{F} Baltasar^{F} Kanseko^{F}, qin^{A} Toma^{H} Crus^{F} Crus^{F}, qo^{E} qin^{A} Jya^{B} Crus^{F} Batista^{J} no^{A} qne^{G} jnya^{F} chaq^{F} ylaq^{J} tykwan^{F} no^{A} nlyo^{E} kwen^{E} qo^{E} no^{A} nya^{B} ktyi^{C} wra^{K} no^{K} ni^{E} chaq^{F} wa^{G} qin^{A} nten^{B}. Kwiq^{J} kwan^{H} niyan^{J} tya^{F} wa^{G} xqwe^{F} qin^{J} nten^{B} no^{K} qne^{G} jnya^{F} ti^{H} Swesya^{J}. Tya^{F} wa^{G} xqwe^{F} qin^{J} tqa^{J} nten^{K} no^{K} qne^{J} jnya^{F} neq^{C} Laboratoryo^{J} qin^{J} qan^{A} xla^{K} Lund chaq^{F} nda^{F} yaq^{C} qwa^{G} qo^{E} qin^{J} nten^{B} no^{K} na^{H} son^{B} no^{A} nlo^{I} qin^{J} qan^{A} xlya^{K} no^{A} qne^{J} xqan^{H} nten^{B} (CIESAS), kchin^{A} Xyaq^{A}.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.584231/full#supplementary-material

REFERENCES

Anderson, S. R., and Keenan, E. L. (1985). "Deixis," in *Language Typology and Syntactic Description*, ed T. Shopen (Cambridge: Cambridge University Press), 259–308.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi: 10.1111/j.0956-7976.2004.00694.x

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Blythe, J., Mardigan, K. C., Perdjert, M. E., and Stoakes, H. (2016). Pointing out directions in Murrinhpatha. *Open Linguist.* 2, 1–28. doi: 10.1515/opli-2016-0007

Bril, I. (2004). "Deixis in nêlêmwa (New Caledonia)," in *Deixis and Demonstratives in Oceanic languages*, ed G. Senft (Pacific Linguistics), 99–127.

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache [Language Theory: The Representational Function of Language]*. Stuttgart: Fischer.

Burenhult, N. (2003). Attention, accessibility, and the addressee: the case of the Jahai demonstrative ton. *Pragmatics* 13, 363–379. doi: 10.1075/prag.13.3.01bur

Burenhult, N. (2008). Spatial coordinate systems in demonstrative meaning. *Linguist. Typol.* 12, 99–142. doi: 10.1515/LITY.2008.033

Burenhult, N. (2018). *The Jahai Multi-Term Demonstrative System: What's Spatial About It?* Vol. 14. Cambridge: Cambridge University Press.

Campbell, E. (2013). The internal diversification and subgrouping of Chatino. *Int. J. Am. Linguist.* 79, 395–420. doi: 10.1086/670924

Chu, M., and Hagoort, P. (2014). Synchronization of speech and gesture: evidence for interaction in action. *J. Exp. Psychol. Gen.* 143, 1726–1741. doi: 10.1037/a0036281

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). "Pointing and placing," in *Pointing: Where Language, Culture and Cognition Meet*, ed S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 243–261.

Clark, H. H. (2020). Anchoring utterances. *Top. Cogn. Sci.* doi: 10.1111/tops.12496

Clark, H. H., and Bangerter, A. (2004). "Changing ideas about reference," in *Experimental Pragmatics*, eds I. Noveck and D. Sperber (Basingstoke: Palgrave MacMillan), 25–49. doi: 10.1057/9780230524125_2

Cooperrider, K. (2011). *Reference in action: links between pointing and language* (Ph.D. thesis), University of California at San Diego, San Diego, CA, United States.

Cooperrider, K. (2014). Body-directed gestures: pointing to the self and beyond. *J. Pragmat.* 17, 1–16. doi: 10.1016/j.pragma.2014.07.003

Cooperrider, K. (2015). The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53, 632–656. doi: 10.1080/0163853X.2015.1094280

Cooperrider, K., and Mesh, K. (2020). "Pointing in gesture and sign: one tool, many uses," in *The Blossoming of Gesture in Language, eds A. Morganstern and S. Goldin-Meadow* (Berlin: Mouton de Gruyter).

Cooperrider, K., Slotta, J., and Núñez, R. (2018). The preference for pointing with the hand is not universal. *Cogn. Sci.* 42, 1375–1390. doi: 10.1111/cogs.12585

Coventry, K. R., Griffiths, D., and Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: describing and remembering object location. *Cogn. Psychol.* 69, 46–70. doi: 10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near–far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

Cruz, E. (2011). *Phonology, tone and the functions of tone in San Juan Quiahije Chatino* (Ph.D. thesis), The University of Texas at Austin, Austin, TX, United States.

Cruz, E. (2017). Documenting landscape knowledge in Eastern Chatino: narratives of fieldwork in San Juan Quiahije. *Anthropol. Linguist.* 59, 205–231. doi: 10.1353/anl.2017.0006

Cruz, E., and Sullivant, J. R. (2012). *Demostrativos próximos y distales: Un estudio comparativo del uso de demostrativos en el chatino de Quiahije y Tataltepec*. Retrieved from: https://islandora-ailla.lib.utexas.edu/islandora/object/ailla:274910

Cruz, E., and Woodbury, A. C. (2014). Finding a way into a family of tone languages: the story and methods of the Chatino Language Documentation Project. *Lang. Document. Conserv.* 8, 490–524. Available online at: http://hdl.handle.net/10125/24615

Cruz, H. (2014). *Linguistic poetics and rhetoric of Eastern Chatino of San Juan Quiahije* (Ph.D. thesis), The University of Texas, Austin, TX, United States.

Diessel, H. (1999). *Demonstratives: Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguist.* 17, 463–489. doi: 10.1515/COG.2006.015

Diessel, H. (2012). "Deixis and demonstratives," in *Semantics: An International Handbook of Natural Language Meaning*, eds C. Maienborn, K. Heusinger, and P. Portner (Berlin: De Gruyter), 2407–2432.

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Lang. Linguist. Compass* 8, 116–132. doi: 10.1111/lnc3.12066

Dixon, R. M. W. (2003). Demonstratives: a cross-linguistic typology. *Stud. Lang.* 27, 61–112. doi: 10.1075/sl.27.1.04dix

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi: 10.1515/semi.1969.1.1.49

ELAN (2020). *ELAN (Version 5.9) [Computer Software]*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from: https://archive.mpi.nl/tla/elan

Emmorey, K., and Casey, S. (2001). Gesture, thought and spatial language. *Gesture* 1, 35–50. doi: 10.1075/gest.1.1.04emm

Enfield, N. (2018). "Lao demonstrative determiners nii4 and nan4: an intensionally discrete distinction for extensionally analogue space," in *Demonstratives in Cross-Linguistic Perspective*, Vol. 14, eds S. Levinson, S. Cutfield, M. Dunn, N. Enfield, S. Meira, and D. Wilkins (Cambridge: Cambridge University Press), 72. doi: 10.1017/9781108333818.004

Enfield, N. J. (2001). 'Lip-pointing': a discussion of form and function with reference to data from Laos. *Gesture* 1, 185–212. doi: 10.1075/gest.1.2.06enf

Enfield, N. J. (2003). Demonstratives in space and interaction: data from Lao speakers and implications for semantic analysis. *Language* 79, 82–117. doi: 10.1353/lan.2003.0075

Fillmore, C. J. (1982). "Towards a descriptive framework for spatial deixis," in *Speech, Place, and Action: Studies in Deixis and Related Topics*, eds K. Buhler, R. J. Jarvella, and W. Klein (New York, NY: John Wiley & Sons), 31–59.

Fricke, E. (2014). "Deixis, gesture, and embodiment from a linguistic point of view," in *Body–Language–Communication*, eds C. Müller, A. Cienki, and E. Fricke (Berlin: Walter de Gruyter), 1803–1823.

Gamer, M., Lemon, J., Fellows, I., and Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1. Retrieved from: https://CRAN.R-project.org/package=irr

Gonseth, C., Kawakami, F., Ichino, E., and Tomonaga, M. (2017). The higher the farther: distance-specific referential gestures in chimpanzees (pan troglodytes). *Biol. Lett.* 13:20170398. doi: 10.1098/rsbl.2017.0398

Gudde, H. B., Coventry, K. R., and Engelhardt, P. E. (2016). Language and memory for object location. *Cognition* 153, 99–107. doi: 10.1016/j.cognition.2016.04.016

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor. Quant. Methods Psychol.* 8:23. doi: 10.20982/tqmp.08.1.p023

Hassemer, J., and McCleary, L. (2018). The multidimensionality of pointing. *Gesture* 17, 417–463. doi: 10.1075/gest.17018.has

Haviland, J. B. (2003). "How to point in Zinacantán," in *Pointing: Where Language, Culture and Cognition Meet*, ed S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 139–170.

Haviland, J. B. (2009). "Pointing, gesture spaces, and mental maps," in *Language and Gesture*, ed D. McNeill (Cambridge: Cambridge University Press), 13–46. doi: 10.1017/CBO9780511620850.003

Hegarty, M., Mayer, S., Kriz, S., and Keehner, M. (2005). The role of gestures in mental animation. *Spat. Cogn. Comput.* 5, 333–356. doi: 10.1207/s15427633scc0504_3

Himmelman, N. (1996). "Demonstratives in narrative discourse: a taxonomy of universal uses," in *Studies in Anaphora*, ed B. Fox (Amsterdam: John Benjamins), 205–254. doi: 10.1075/tsl.33.08him

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometric. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Instituto Nacional de Estadística, Geograf?a e Informática (INEGI) (2010). *INEGI Censo de Población y Vivienda* 2010. Retrieved from: https://www.inegi.org.mx/

Jarbou, S. O. (2010). Accessibility vs. physical proximity: an analysis of exophoric demonstrative practice in spoken Jordanian Arabic. *J. Pragmat.* 42, 3078–3097. doi: 10.1016/j.pragma.2010.04.014

Jungbluth, K. (2003). "Deictics in the conversational dyad," in *Deictic Conceptualisation of Space, Time and Person, Volume 112 of Pragmatics & Beyond New Series*, ed F. Lenz (Amsterdam: John Benjamins), 13–40. doi: 10.1075/pbns.112.04jun

Kemmerer, D. (1999). 'Near' and 'far' in language and perception. *Cognition* 73, 35–63. doi: 10.1016/S0010-0277(99)00040-2

Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds Seigman and B. Pope (New York, NY: Pergamon Press), 177–216. doi: 10.1016/B978-0-08-015867-9.50013-7

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kendon, A., and Versante, L. (2003). "Pointing by hand in Neapolitan," in *Pointing: Where Language, Culture and Cognition Meet*, ed S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 109–137.

Kita, S. (Ed.). (2003). *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Psychology Press.

Kita, S., van Gijn, I., and van der Hulst, H. (1998). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction, International Gesture Workshop*, eds I. Wachsmuth and M. Frölich (Bielefeld: Springer), 23–35. doi: 10.1007/BFb0052986

Kranstedt, A., Lücking, A., Pfieffer, T., Rieser, H., and Wachsmuth, I. (2006). "Deixis: how to determine demonstrated objects using a pointing cone," in *Gesture in Human-Computer Interaction and Simulation*, eds S. Gibet, N. Courty, and J. F. Kamp (Berlin: Springer), 300–311. doi: 10.1007/11678816_34

Krivokapic, J., Tiede, M., and Tyrone, M. E. (2016). "Speech and manual gesture coordination in a pointing task," in *Proceedings of Speech Prosody 2016*, eds J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux. doi: 10.21437/SpeechProsody.2016-255

Küntay, A. C., and Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *J. Child Lang.* 33, 303–320. doi: 10.1017/S0305000906007380

Laury, R. (1997). *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish. Studies in Discourse and Grammar*. Amsterdam: John Benjamins.

Levelt, W., Richardson, G., and La Heij, W. (1985). Pointing and voicing in deictic expressions. *J. Mem. Lang.* 24, 133–164. doi: 10.1016/0749-596X(85)90021-X

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.

Levinson, S. C. (2004). "Deixis," in *The Handbook of Pragmatics*, eds L. Horn and G. Ward (Oxford: Blackwell Publishing), 97–121. doi: 10.1002/9780470756959.ch5

Levy, E. T., and McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Process.* 15, 277–301. doi: 10.1080/01638539209544813

Lieberman, P. (1967). *Intonation, Perception, and Language*. (Research Monograph No. 38.) Cambridge, MA: MIT Press.

Mesh, K. (2017). *Points of comparison: What indicating gestures tell us about the origins of signs in San Juan Quiahije Chatino Sign Language* (Ph.D. thesis), The University of Texas at Austin, Austin, TX, United States.

Mihas, E. (2017). Interactional functions of lip funneling gestures: a case study of Northern Kampa Arawaks of Peru. *Gesture* 16, 432–479. doi: 10.1075/gest.00004.mih

Müller, C. (2004). "Forms and uses of the palm up open hand: a case of a gesture family," in *The Semantics and Pragmatics of Everyday Gestures*, Vol. 9 (Berlin: Weidler Berlin), 233–256.

Ozanne-Rivierre, F. (2004). "Spatial deixis in iaai (loyalty islands)," in *Deixis and Demonstratives in Oceanic Languages*, ed G. Senft (Canberra: Pacific Linguistics), 129–139.

Peeters, D., Chu, M., Holler, J., Hagoort, P., and Özyürek, A. (2015a). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* 27, 2352–2368. doi: 10.1162/jocn_a_00865

Peeters, D., Hagoort, P., and Özyürek, A. (2015b). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition* 136, 64–84. doi: 10.1016/j.cognition.2014.10.010

Peeters, D., and Özyürek, A. (2016). This and that revisited: a social and multimodal approach to spatial demonstratives. *Front. Psychol.* 7:222. doi: 10.3389/fpsyg.2016.00222

Peirce, C. (1955). *Philosophical Writings of Peirce*. New York, NY: Dover Publications.

Piwek, P., Beun, R.-J., and Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: evidence from deictic demonstratives in Dutch. *J. Pragmat.* 40, 694–718. doi: 10.1016/j.pragma.2007.05.001

R Core Team (2019). *R: A Language and Environment for Statistical Computing (Version 3.6.1)*. Vienna: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/

Rocca, R., and Wallentin, M. (2020). Demonstrative reference and semantic space: a large-scale demonstrative choice task study. *Front. Psychol.* 11:629. doi: 10.3389/fpsyg.2020.00629

Rocca, R., Wallentin, M., Vesper, C., and Tylén, K. (2019). This is for you: social modulations of proximal vs. distal space in collaborative interaction. *Sci. Rep.* 9:14967. doi: 10.1038/s41598-019-51134-8

Schapper, A., and San Roque, L. (2011). Demonstratives and non-embedded nominalisations in three Papuan languages of the Timor-Alor-Pantar family. *Stud. Lang.* 35, 380–408. doi: 10.1075/sl.35.2.05sch

Seyfeddinipur, M. (2006). *Disfluency: interrupting speech and gesture* (Ph.D. thesis), Radboud University, Nijmegen, Netherlands.

Sherzer, J. (1973). Verbal and nonverbal deixis: the pointed lip gesture among the San Blas Cuna. *Lang. Soc.* 2:117. doi: 10.1017/S0047404500000087

Smith Aguilar, A. E. (2017). *Paisaje ritual y fragmentación del territorio en San Juan Quiahije, Oaxaca* (Master's thesis), Universidad Nacional Autónoma de México, Mexico City, Mexico.

Stukenbrock, A. (2014). Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction. *J. Pragmat.* 74, 70–93. doi: 10.1016/j.pragma.2014.08.001

Talmy, L. (1988). "The relation of grammar to cognition," in *Topics in Cognitive Linguistics*, ed B. Rudzka (Amsterdam: John Benjamins Publishing), 165–206. doi: 10.1075/cilt.50.08tal

Talmy, L. (2018). *The Targeting System of Language*. Cambridge, MA: MIT Press.

Villard, S., and Sullivant, J. R. (2016). "Language documentation in two communities with high migration rates," in *Language Documentation and Revitalization in Latin American Contexts*, eds G. Pérez Báez, C. Rogers, and J. E. Rosés Labrada (Berlin; Boston, MA: De Gruyter Mouton), 273–304. doi: 10.1515/9783110428902-011

Wilkins, D. (1999a). "Eliciting contrastive use of demonstratives for objects within close personal space (all objects well within arm's reach)," in *Manual for the 1999 Field Season*, ed D. Wilkins (Nijmegen: Max Planck Institute for Psycholinguistics), 25–28.

Wilkins, D. (1999b). "The 1999 demonstrative questionnaire: "This" and "that" in comparative perspective," in *Manual for the Field Season 2001*, ed D. Wilkins (Nijmegen: Max Planck Institute for Psycholinguistics), 1–24.

Wilkins, D. (2003). "Why pointing with the index finger is not a universal (in sociocultural and semiotic terms)," in *Pointing: Where Language, Culture and Cognition Meet*, ed S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 171–215.

Wilkins, D. (2018). "The demonstrative questionnaire: "THIS" and "THAT" in comparative perspective," in *Demonstratives in Cross-Linguistic Perspective*, eds S. Levinson, S. Cutfield, M. Dunn, N. Enfield, S. Meira, and D. Wilkins (Cambridge: Cambridge University Press), 43–71. doi: 10.1017/9781108333818.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

RESÚMENES—GYAQ^{C}

Cuando los humanos interactúan en el mundo, a menudo orientan la atención de los demás hacia ciertos objetos mediante el uso de expresiones demostrativas y el uso de movimientos de la cabeza, la barbilla, o la mano para señalar dichos objetos. Aunque las estrategias de indicación se han estudiado con frecuencia en entornos de laboratorio, sabemos sorprendentemente poco sobre cómo se utilizan los demostrativos y el señalamiento para coordinar la atención en espacios a gran escala y en contextos naturales. Este estudio investiga cómo los hablantes del chatino de San Juan Quiahije, una lengua indígena de México, señalan y usan demostrativos para dar direcciones a lugares conocidos en múltiples escalas (actividad local, distrito, estado). Los resultados muestran que el uso y la coordinación de demostrativos y el señalamiento cambian a medida que crece la escala del espacio de búsqueda del objetivo. En escalas más grandes, es más probable que los demostrativos y el señalamiento se produzcan juntos, y las dos estrategias parecen manejar diferentes aspectos de la búsqueda del objetivo: los demostrativos orientan la atención principalmente al cuerpo que gesticula, mientras que señalar proporciona claves para reducir el espacio de búsqueda. Estos resultados sacan a luz las distintas contribuciones del habla y el gesto al compuesto lingüístico, al tiempo que ilustran el carácter dinámico de su interacción.

Qan^{K} no^{K} ndywenq^{J} qo^{E} xka^{I} ta^{A} la^{E} nten^{B}, tqa^{K} ti^{K} ntsanq^{B} qin^{K} chonq^{G} na^{F} no^{A} ndiya^{I} qo^{E} loga^{B} no^{K} nlo^{I} la^{J} ti^{J} qna^{G}. Ndiya^{I} tkwa^{J} niyan^{J} muru^{K} jnya^{K} qin^{A} nten^{B} chaq^{F} qne^{J} kasu^{K} qo^{E} jnyi^{A} qya^{H} ndiya^{A} na^{F} qo^{E} loga^{B}: ka^{C} chaq^{F} jlanq^{J} qo^{E} ntykwenq^{J} chaq^{F} (qo^{E} chaq^{F} niyan^{K} no^{K} nde^{C}: re^{C} kwa^{J} qo^{E} kwa^{F}) qo^{E} ka^{C} chaq^{F} qu^{B} xnyi^{K} in^{H} qo^{E} yanq^{C} an^{I} qo^{E} ta^{A} qo^{E} skwaq^{G} tqwan^{J} an^{I}. Lo^{A} ktyi^{C} no^{E} nde^{C} in^{H}, wa^{G} re^{C} ntsaq^{B} wa^{G} qwan^{K} niyan^{K} nlyaq^{I} no^{A} nga^{J} nten^{B} no^{K} ndywiq^{A} chaq^{F} jnya^{J} chaq^{F} qo^{E} qwan^{K} niyan^{K} ntqu^{K} qo^{E} ntsaq^{B} renq^{K} qan^{K} no^{K} jnya^{H} renq^{J} qin^{A} chaq^{F} jnyi^{J} qya^{H} qo^{E} ktsaq^{B} ndiya^{A} loga^{B} no^{K} nlo^{I} qin^{A} kchin^{A} tyi^{A}. Ntqu^{B} wa^{G} chaq^{F} nten^{B} nlo^{J} swi^{H} chaq^{F} qo^{E} qwi^{A} ska^{A} niyan^{J} qne^{E} qan^{K} no^{K} ntsaq^{B} qo^{E} ntqu^{B} ska^{K} na^{F} qo^{E} loga^{B}. Nlo^{J} swi^{H} chaq^{F} qo^{E} qwi^{A} ska^{A} niyan^{J} qne^{E} kanq^{G} chaq^{F} ntqu^{B} ran^{H} si^{K} ta^{A} tyjyuq^{A} nga^{J} qo^{E} ntqen^{A} loga^{B} no^{K} nda^{E} renq^{J} qo^{E} ndywiq^{I} renq^{A} chaq^{F} qin^{J} sqen^{A} no^{A} ntqen^{A} renq^{A}.

1. CHATINO ORTHOGRAPHY

1.1 Practical Orthography: Correspondences With the International Phonetic Alphabet

We use a practical orthography to transcribe Quiahije Chatino, rather than International Phonetic Alphabet (IPA). The consonant phonemes, in practical orthography and IPA, are as follows: bilabial p = [p], b = [nb], m = [m]; apicodental t = [t], d = [nd], ts = [ts],s = [s], n = [n], r =[ɾ], l = [l]; laminoalveolar ty = [t̻], ny = [n̻], ly = [l̻]; alveolopalatal ch = [ʧ], x = [ʃ], y = [j]; Velar k = [k], g = [ng]; labiovelar w = [w]; glottal q = [ʔ], j = [h]. The consonant phonemes, in practical orthography and IPA, are as follows: /i/, /u/, /e/, /o/, /a/. Where the IPA represents nasalized vowels with a diacritic, as in /ã/, we use a vowel followed by an n, as in /an/.

1.2 Transcription of Tone Values

Every word in Quiahije Chatino bears one tone. The tone is represented as a capital letter at the end of the word. The tone value assignments are: A = [Low], B = [High-Low], C = [Mid], E = [High], F = [Low-Mid], G = [Low-High], H = [Mid-superhigh], I = [Midhigh], J = [Mid-Low], K= [superhigh], L = [Low superhigh], and M = [superhigh Low].

2. DISTRIBUTION OF INDICATING STRATEGIES AND DEMONSTRATIVE FORMS, ACROSS TARGETS, AND ACROSS PARTICIPANTS

2.1 Demonstrative Forms (nde^{C}/re^{C} = Proximal, kwa^{F} = Neutral) by Target

Scale	Target, Quiahije Chatino	Target, Translated	nde^{C}/re^{C}	kwa^{F}
Activity	ntenq^{F} tiyuq^{G}	Plain of the Spring	21	24
	tu^{C} kchi^{C}	Mountain Pass	21	10
	ke^{A} tiyeq^{B}	Sour Rock	22	11
	ke^{A} ku^{E} suq^{C}	Turkey Breast Rock	22	17
	lo^{A} si^{K} kyqya^{C} kcheq^{B}	Petition Monument	30	6
	kyqya^{C} kcheq^{B}	Peak of Thorn Mountain	43	8
District	kchin^{A}	San Juan Quiahije	17	6
	ntenq^{F}	Cieneguilla	20	19
	tqwa^{A} tyku^{E}	San José Ixtapan	19	20
	skwi^{E}	San Miguel Panixtlahuaca	34	18
	sqwe^{F}	Santa Catarina Juquila	38	10
	ke^{G} xin^{E}	Santiago Yaitepec	16	21
State	se^{A} na^{A} nya^{K}	Santiago Minas	18	17
	tsi^{C}	San Marcos Zacatepec	16	12
	ntenq^{F} tyku^{E} jlyu^{B}	Rio Grande	9	15
	lo^{A} ntqa^{B}	Oaxaca City	19	22

2.2 Indicating Strategies (dem. Alone, dem. + Manual Point, dem. + chin point) by Target

Scale	Target, Quiahije Chatino	Target, Translated	D	DM	DC
Activity	ntenq^{F} tiyuq^{G}	Plain of the Spring	22	18	5
	tu^{C} kchi^{C}	Mountain Pass	21	9	1
	ke^{A} tiyeq^{B}	Sour Rock	22	8	3
	ke^{A} ku^{E} suq^{C}	Turkey Breast Rock	27	9	3
	lo^{A} si^{K} kyqya^{C} kcheq^{B}	Petition Monument	25	10	1
	kyqya^{C} kcheq^{B}	Peak of Thorn Mountain	36	15	0
District	kchin^{A}	San Juan Quiahije	9	11	3
	ntenq^{F}	Cieneguilla	36	15	0
	tqwa^{A} tyku^{E}	San José Ixtapan	18	21	0
	skwi^{E}	San Miguel Panixtlahuaca	29	19	4
	sqwe^{F}	Santa Catarina Juquila	21	26	1
	ke^{G} xin^{E}	Santiago Yaitepec	12	21	4
State	se^{A} na^{A} nya^{K}	Santiago Minas	15	20	0
	tsi^{C}	San Marcos Zacatepec	15	13	0
	ntenq^{F} tyku^{E} jlyu^{B}	Rio Grande	9	11	4
	lo^{A} ntqa^{B}	Oaxaca City	16	22	3

2.3 Demonstrative Forms (nde^{C}/re^{C} = Proximal, kwa^{F} = Neutral) by Participant

Participant	nde^{C}/re^{C}	kwa^{F}
R04	65	25
R05	47	46
R06	24	27
R07	35	38
R08	59	43
R09	64	26
R10	43	14
R11	28	17

2.4 Indicating Strategies (dem. Alone, dem. + Manual Point, dem. + Chin Point) by Participant

Participant	D	DM	DC
R04	35	52	3
R05	38	52	3
R06	23	18	10
R07	36	35	2
R08	51	49	2
R09	58	30	2
R10	47	6	4
R11	22	14	9

3. REGRESSION MODELS

3.1 Analysis 1: Outcome: Choice of Demonstrative, Predictors: Scale and Distance

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	1.598	0.262	6.109	0.000
distance.rescaled	−2.833	0.520	−5.452	0.000
scale.factordistrict	−0.809	0.371	−2.182	0.029
scale.factorstate	−1.576	0.322	−4.901	0.000
distance.rescaled:scale.factordistrict	2.222	0.706	3.150	0.002
distance.rescaled:scale.factorstate	2.577	0.659	3.909	0.000

3.2 Analysis 2: Outcome: Chin Point, Predictors: Scale and Distance

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	−3.115	0.517	−6.032	0.000
distance.rescaled	0.437	0.885	0.494	0.622
scale.factordistrict	0.982	0.659	1.491	0.136
scale.factorstate	−0.252	0.713	−0.353	0.724
distance.rescaled:scale.factordistrict	−1.831	1.316	−1.392	0.164
distance.rescaled:scale.factorstate	0.505	1.236	0.408	0.683

3.3 Analysis 3: Outcome: Manual Point, Predictors: Scale and Distance

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	−1.555	0.328	−4.747	0.000
distance.rescaled	1.606	0.487	3.296	0.001
scale.factordistrict	1.568	0.363	4.317	0.000
scale.factorstate	1.452	0.321	4.526	0.000
distance.rescaled:scale.factordistrict	−1.810	0.683	−2.653	0.007
distance.rescaled:scale.factorstate	−1.449	0.642	−2.258	0.024

3.4 Analysis 4: Outcome: Elbow Height, Predictors: Scale and Distance

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	−1.481	0.589	−2.513	0.012
distance.rescaled	1.985	0.907	2.187	0.029
scale.factordistrict	1.476	0.616	2.398	0.017
scale.factorstate	2.712	0.626	4.333	0.000
distance.rescaled:scale.factordistrict	−1.772	1.151	−1.539	0.124
distance.rescaled:scale.factorstate	−2.404	1.132	−2.125	0.034

3.5 Analysis 5: Outcome: Head Point, Predictors: Scale and Demonstrative

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	−3.371	0.529	−6.376	0.000
dem.factorkwaF	0.912	0.582	1.566	0.117
scale.factordistrict	0.093	0.586	0.159	0.873
scale.factorstate	0.660	0.669	0.987	0.324
dem.factorkwaF:scale.factordistrict	0.014	0.803	0.018	0.985
dem.factorkwaF:scale.factorstate	−1.412	0.987	−1.431	0.152

3.6 Analysis 6: Outcome: Manual Point, Predictors: Scale and Demonstrative

	Estimate	Std. Error	*z* value	Pr(>|z|)
(Intercept)	−0.980	0.331	−2.957	0.003
dem.factorkwaF	−0.430	0.330	−1.305	0.192
scale.factordistrict	1.240	0.257	4.836	0.000
scale.factorstate	1.666	0.346	4.815	0.000
dem.factorkwaF:scale.factordistrict	−0.488	0.433	−1.128	0.259
dem.factorkwaF:scale.factorstate	−0.955	0.515	−1.867	0.063

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 January 2021
doi: 10.3389/fpsyg.2020.579992

The Grammatical Incorporation of Demonstratives in an Emerging Tactile Language

Terra Edwards[1] and Diane Brentari[2]*

[1] Department of Sociology & Anthropology, Saint Louis University, St. Louis, MO, United States, [2] Department of Linguistics, The University of Chicago, Chicago, IL, United States

In this article, we analyze the grammatical incorporation of demonstratives in a tactile language, emerging in communities of DeafBlind signers in the US who communicate via reciprocal, tactile channels—a practice known as "protactile." In the first part of the paper, we report on a synchronic analysis of recent data, identifying four types of "taps," which have taken on different functions in protacitle language and communication. In the second part of the paper, we report on a diachronic analysis of data collected over the past 8 years. This analysis reveals the emergence of a new kind of "proprioctactic" tap, which has been co-opted by the emerging phonological system of protactile language. We link the emergence of this unit to both demonstrative taps, and backchanneling taps, both of which emerged earlier. We show how these forms are all undergirded by an attention-modulation function, more or less backgrounded, and operating across different semiotic systems. In doing so, we contribute not only to what is known about demonstratives in tactile languages, but also to what is known about the role of demonstratives in the emergence of new languages.

Keywords: protactile, language emergence, deixis, demonstratives, intersubjectivity, tactile signed language, DeafBlind, tactile phonology

OPEN ACCESS

Edited by:
Kenny Coventry, University of East Anglia, United Kingdom

Reviewed by:
Bencie Woll, University College London, United Kingdom
Rose Stamp, Bar Ilan University, Israel

****Correspondence:***
Terra Edwards
terra.edwards@slu.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *03 July 2020*
Accepted: *29 October 2020*
Published: *13 January 2021*

Citation:
Edwards T and Brentari D (2021) The Grammatical Incorporation of Demonstratives in an Emerging Tactile Language. Front. Psychol. 11:579992. doi: 10.3389/fpsyg.2020.579992

INTRODUCTION

In this article, we analyze the grammatical incorporation of demonstratives in a tactile language, currently emerging in communities of DeafBlind signers in the US who communicate via reciprocal, tactile channels—a practice known as "protactile." We argue that this process is undergirded by reconfiguration of intersubjective relations, including habitual modes of attention to others and the environment. It is well known that deictic systems—and demonstratives in particular—are a powerful means of facilitating intersubjective coordination (e.g., Agha, 1994; Benveniste, 1971; Rommetveit, 1976; Duranti, 2010; Hanks, 2013; Sidnell, 2014; Evans et al., 2018a,b). However, in order to be effective, participants must assume reciprocal, perceptual access to each other and the environment. The systems we analyze in this article are emerging in DeafBlind communities where reciprocal modes of access have been re-organized around tactile channels (Edwards, 2015). In this article, we identify linguistic resources that have emerged since then for modulating attention within those newly re-contoured environments. In doing so, we contribute not only to what is known about demonstratives, but also to what is known about their role in the emergence of new languages.

BACKGROUND

Protactile language[1] has emerged in groups of DeafBlind people who, for the most part, were born sighted, acquired American Sign Language (ASL) as children, and became blind slowly over several subsequent decades. As that process unfolded, visual communication in general, and ASL in particular, became increasingly untenable. Prior to the protactile movement, this was addressed via increased dependence on sighted interpreters. Since the inception of the protactile movement, there has been a politically and culturally framed demotion of visual communication and ASL, and an explicit push toward experimentation and innovation aimed at maximizing the potential of the tactile channel for purposes of communication (Edwards, 2014; McMillen, 2015; Granda and Nuccio, 2018; Clark, unpublished). As a result, new grammatical systems are beginning to emerge, which are optimized, as never before, to the tactile modality (Edwards and Brentari, 2020). As grammatical systems that interact most extensively with sensory-motor and interactional interfaces, phonology and deixis are at the center of this transformation.

In our prior research, we have shown that in roughly 10 years, a new phonological system has become conventional in protactile, DeafBlind communities, and that conventionalization of protactile phonology involves assigning specific grammatical roles to the four hands (and arms) of Signer 1 ("conveyer") and Signer 2 ("receiver") in "proprioceptive constructions" (PCs), which are comparable to "classifier constructions" in visual signed languages (Edwards and Brentari, 2020).[2] In producing a PC, Signer 1 and Signer 2 work together to define the global space of articulation (similar to a "place of articulation"), within which, and to which, attention can be directed. We argue in what follows that the grammatical system governing the unfolding articulation of the PC incorporates and constrains protactile demonstratives. Protactile demonstratives are expressed using a combination of movement and contact that can be described as "tapping." However, this is only one of several types of tapping that occur. In what follows, we trace the divergence of taps as they take on distinct functions in protactile language and communication.

Demonstrative Categories

Diessel (1999) categorizes demonstratives according to their morphosyntactic properties from crosslinguistic and diachronic perspectives, and argues that demonstratives occur in four syntactic contexts (p. 1): (i) they are used as independent pronouns in argument position of verbs and adpositions (pronominal); (ii) they may co-occur with a noun in a noun phrase (adnominal); (iii) they may function as verb modifiers, and (adverbial); (iv) they may occur in copular and non-verbal clauses (identificational). Insofar as i–iv are distinguished formally, Deisssel assigns each to a corresponding grammatical category: (i) demonstrative pronoun; (i) demonstrative determiner; (ii) demonstrative adverb; and (iv) demonstrative identifier. He argues that the grammatical pathway demonstratives will take are determined by the syntactic context in which they occur (p. 2). In this study, we employ a modified set of Deissel's demonstrative categories for several reasons. First, this study does not include elicitations designed to establish a noun-verb distinction (i.e., Abner et al., 2019). Therefore, we have replaced adnominal and adverbial demonstratives with a single category: "demonstrative modifier," which can be applied either exophorically or endophorically, i.e., to refer to referents in the immediate environment, or to refer to linguistic aspects of the unfolding discourse. Second, we are tracking the diachronic development of a single form: "tap." In the data we have analyzed here, tap does not appear in pronominal or identificational contexts. This reduces Deissel's four categories to one: demonstrative modifier. The third reason we depart from Deissel's categories is that our frame, by necessity, must handle an emerging (rather than well-established) linguistic system, and includes non-linguistic interactional signals, which, we argue, preceded, and contributed to the emergence of demonstratives with linguistic properties.

Protactile Taps

In this article, we present evidence for four distinct types of taps: A tactile backchanneling tap, which is not part of the deictic system, but has an attention-modulating function, and may have functioned as a precursor to demonstrative and propriotactic taps; two kinds of demonstrative taps—one used for endophoric demonstrative reference and the other for exophoric demonstrative reference. In addition, we have identified a type of tap that is used to organize sequences of linguistic units by coordinating the four articulators of Signer 1 and Signer 2 (as discussed below). These forms, which we call "proprioactic" taps, are taps that have been co-opted by the phonological system, thereby entering the grammar of protactile language.

While only two of the four forms we analyze are demonstratives, we are interested in the intersubjective, attention-modulation function that underlies all four forms in more or less backgrounded ways. The order in which these forms emerge suggests a trajectory along which patterns in attention modulation, as part of a broader process of intersubjective coordination, are incorporated into, and integrated with, grammatical systems as those systems emerge. Tracking the way taps take on new functions in increasingly grammatical systems offers some insight into how this process can unfold, and helps us to understand the crucial role that demonstratives (and deixis, more generally) might play in that process.

The emergence and differentiation of protactile taps is part of a broader divergence between protactile language and ASL—the visual language on which it was originally scaffolded. Therefore, some background is needed on the relationship between the two languages.

[1] We use the term "protactile language" since this is the term that is currently in widespread use within the DeafBlind community that uses it.

[2] Classifier constructions in sign languages are polymorphemic predicates, in which the movement is a verbal root (i.e., TO-GO or BE-LOCATED) and the handshape is an affix that represents a class of objects or size and shape of objects (Supalla, 1982; Zwisterlood, 2012).

The Relationship of Protactile Language to ASL

Some DeafBlind people live as minorities within larger Deaf, sighted communities, while others are active members of a signing or non-signing DeafBlind community. Still others interact solely with hearing sighted people, and have no contact with Deaf or DeafBlind communities. Therefore, language and communication vary widely from community to community and across individuals. The dominant language in some DeafBlind communities in the United States is English, perceived via adaptive technologies such as amplification systems. In others, the dominant language is ASL. In order to perceive ASL through touch, the receiver places their hand(s) on top of the hand(s) of the signer to track the production of signs. Just as spoken languages require adaptive measures to be perceived by DeafBlind signers, adaptations and innovations are necessary for the perception of visual languages by DeafBlind signers as well. All of the participants in this study were fluent in ASL—whether perceived visually or tactually—prior to becoming DeafBlind, and hence can access ASL linguistic representations in some form.

However, for DeafBlind signers, ASL has the great disadvantage of being difficult to perceive, and therefore to *use*. According to DeafBlind, protactile leaders and educators aj Granda and Nuccio (2018), this difficulty is grounded in one, fundamental problem: the use of "air space," or the space on and around the body of the signer. Protactile language is produced instead in "contact space," or the space on the addressee's body. This shift unlocks proprioception as a rich and accessible dimension of the tactile channel. In air space, locations are perceived relative to each other against a visual backdrop that is inaccessible for DeafBlind signers (e.g., "to the right of the mouth" vs. "to the right of the eye"). In contrast, locations in contact space can be clearly perceived against the proprioceptive backdrop of the listener's own body. **Figure 1** demonstrates this shift in the sign for SAME **Figure 1**; the citation form of the ASL sign is in **Figure 1** (left). In **Figure 1** (right) we see that both signers co-produce this sign. Signer 1 (right) produces an ASL "Y" handshape: 🤙 as in the ASL sign SAME; however, the sign is produced by making contact with Signer 2's hand (left). The ASL handshape is not articulated in air space; instead, it has been transferred to contact space. All of the demonstratives analyzed in this paper occur in contact space.

In what sense, then, is this an emerging language of its own and not simply a variety of ASL? Edwards and Brentari (2020) have shown that the move to contact space triggered the emergence of new atomic units out of which signs are built, as well as new well-formedness constraints, which determine how protactile signs can and cannot be articulated. These constraints differ, fundamentally, from ASL. For example, instead of two hands, as in ASL, protactile language has four hands which can be activated in the creation of utterances. We label the four hands used in many protactile signs neutrally as "Articulators": A1 (signer 1 dominant hand); A2 (signer 2 dominant hand); A3 (signer 1 non-dominant hand) and A 4 (signer 2 non-dominant hand). Each has its own set of specific linguistic functions, as Edwards and Brentari (2020) have described, and as we summarize in the section on proprioceptive constructions below.

In contact space, it is important that the addressee can feel signs as they are produced on their own body, that they can distinguish signs from one another, and that iconic and indexical grounds are maintained, linking signs, wherever relevant, to resonant and accessible tactile experiences, that can be shared by all speakers of the language. As reported by protactile signers themselves, the aim is not to preserve ASL to the greatest extent possible, but to embrace the potential of the proprioceptive/tactile modality for representing and evoking shared experiences. Granda and Nuccio (2018, p. 13) explain:

> As Deaf children, we were drawn to visual imagery in ASL stories— transported into the vivid details of the worlds created for us. As DeafBlind adults, we still carry those values within us, but ASL doesn't evoke those same feelings for us anymore. When you are perceiving a visual language through touch, the precision, beauty, and emotion are stripped away; the imagery is lost. [. . .] If you try to access an ASL story through an interpreter [. . .], you just feel a hand moving around in air space [. . .].

FIGURE 1 | Handshape transferred to contact space via PC devices and conventions [ASL sign SAME (left) is re-produced from Hochgesang et al. (2018)].

> In air space we are told what is happening for other people, but nothing happens for us.

This orientation suggests that protactile signers are prioritizing intuitive and effective communication over and against the preservation of ASL structures. Elements taken from ASL, such as "handshapes" are admitted into protactile language insofar as they meet these criteria. For example, there are two classifier handshapes for representing a "person" in ASL—the "1" handshape:  and an upside-down "V" handshape:  . The "1" handshape does not articulate to contact space easily because the bottom of the wrist is difficult to position and move on the body of the addressee in a precise and perceptible way. In contrast, in the "V" classifier , the two extended fingers representing the legs and the tips of the fingers make contact with the body. This handshape is perceptible. In addition, it can be modified for manner and direction of movement such that iconic and indexical grounds can be established and maintained in the unfolding of the communicative event. Given this, the upside-down V handshape is selected and the "1" handshape is discarded. In other words, only one of the two handshapes lends itself to the application of emerging, protactile constraints. This suggests that instead of working within the ASL grammar, and "adapting" or "compensating" as needed, protactile signers are operating within the new, tactile system, retrieving elements from ASL only insofar as they can conform to emerging patterns and rules. They are treating ASL as a kind of archival lexicon, or in the words of one participant, a "junk yard." Furthermore, archived elements of ASL are only one source of material for building new protactile forms.

As we demonstrate in this article, another source of new protactile forms is interaction, and more specifically, cues that have emerged to facilitate intersubjective coordination, such as backchanneling and turn-taking (Edwards, 2014, p. 144–158)[3]. This paper is concerned with one type of form, which we refer to simply as "taps." Anyone observing protactile communication for the first time will be struck by the sheer quantity of taps present in the stream of discourse, and will have some difficulty in interpreting those taps in ways their interlocutors seem to expect. In what follows, we argue that the complex multifunctionality taps have taken on can be traced back to a simple backchanneling cue used to signal continued attention or agreement. Backchanneling signals like these have been described in other DeafBlind communities (e.g., "YES-tapping" in Mesch, 2001; and see Willoughby et al., 2018, p. 9–11). We argue that these signals have been co-opted by protactile language to serve several different attention-modulating functions, including demonstrative reference.

Pointing in Language Emergence

Typological and historical studies of language emergence are informative, but there are no cases of emergent spoken languages recent enough to be studied *al vivo*. Much of what we know about language emergence, then, comes from studying sign languages. In this growing body of work, the semiotic diversification and grammatical incorporation of pointing has become a focus (Meir, 2003; Coppola and Senghas, 2010; Pfau, 2011; Hopper and Traugott, 2003 [1993]; De Vos, 2014; Kocab et al., 2015; Mesh, 2017). As others have noted, grammaticalization has traditionally been studied in spoken languages and started with lexical forms, tracing how those forms take on new, grammatical functions as part of larger processes of language change (Hopper and Traugott, 2003 [1993]). However, there is growing interest in co-speech gesture and other forms of "visible action" (Kendon, 2004) as input to grammaticalization and related processes of language emergence in both spoken and visual signed language communication. This research has raised questions about how that input is treated by the linguistic system as containing structure that is accessible to the agents of language creation and language change. Deictic systems figure prominently in these debates.

For example, Kathryn Mesh, in a study examining the gestural origins of signs in San Juan Quiajije Chatino Sign Language, argues that "indicating" gestures are not, as McNeil (1992) has claimed, holistic "gesticulations," but rather, forms with internal structure. For targets near the gesturer, elbow height is low and it increases as the distance of the target from the gesturer increases (Mesh, 2017, p. 65). This supports earlier findings (cited in Mesh, 2017, p. 47–48) that changes in the height of an indicating gesture correspond to the distance of the target among both hearing gesturers (Kendon, 1980; Levinson, 2003; Haviland, 2009; Ola Orie, 2009; Le Guen, 2011) and Deaf gesturers (van der Kooij, 2002; De Vos, 2014). Mesh shows that the internal structure of these indicating gestures is perceptible visually, without any access to the accompanying speech, and therefore constitutes rich input for creators of a new signed language (2017, p. 37–1122). Dachovksy (2018) argues facial expressions that are mutually accessible to signers of in the young Al-Sayyid Bedouin Sign Language serve as input to the creation of relative clause construction. In a similar vein, we are concerned with structured and meaningful resources available to the DeafBlind creators of protactile language. We focus here on two domains that are likely sources for such resources: ASL, which all of the people in this study acquired as children, and non-linguistic communication conventions that have emerged as part of broader patterns in protactile interaction.

While ASL is not perceptible enough to facilitate unimpeded communication among DeafBlind people (Reed et al., 1995), there are forms of knowledge that come with being a (former) speaker of ASL that are useful in creating a new language. For example, the intuition that "space" can be seized on for purposes of expressing grammatical relations. The concept of "air space," as theorized by protactile signers, constitutes structured, input, which is then re-structured by the creators of protactile language, according to emerging principles, to yield "contact space."

Another example: Former speakers of ASL are likely to have the intuition that new signs can be created via iconically motivated selection of some aspect of a referent

[3]We do not use the term "gesture" here to describe backchanneling. As Kendon (2004) notes, much of what would be considered gesture from one perspective might be considered language from another. There are too many conflicting uses of the term. We therefore identify backchanneling taps in terms of their primary communicative function—to modulate intersubjective attention in interaction.

to metonymically represent the whole (see Boyes-Braem, 1981, p. 42; Mandel, 1981, p. 204–211; Brennan, 1990, p. 11–36; Taub, 2001, p. 43–60). This pattern, too, can be transferred, at some level of abstraction, from visual to tactile domains.

ASL also provides a lexicon to the degree that lexical items are still cognitively accessible in individual speakers. These and other aspects of ASL are not raw materials, nor are they readily accessible as elements within a larger, structured system. Rather, they are *wrought products*—pieces of a language, now ill-suited to the world inhabited by its speakers. As protactile creators sift through the debris, they select elements that have affordances for communication in their new environment, this time, organized around tactile, rather than visual access. As that process unfolds, new patterns emerge that are different from the ones that had previously broken apart. These new patterns begin to govern what can and cannot be incorporated in the emerging system.

In addition to ASL (now sold for parts), protactile communicators draw on tactile communication conventions that emerged prior to, and operate beyond the bounds of, protactile language (Edwards, 2014). In this article, we focus on one of the many backchanneling cues that emerged and became conventional as part of that process: a repetitive tap, which is used to signal agreement or continued attention. We argue that these backchanneling cues have been drawn on in building a new deictic system that operates entirely via tactile channels.

Claims about grammaticalized pointing in signed languages tend to start with pointing gestures. In addition, space is often treated as the primary contextual variable driving semantic distinctions in the emergent system. Here, following the semiotic diversification of "taps" in protactile communication, we arrive at grammaticalized pointing not by way of pointing gestures or space, but by way of intersubjective attention modulators. While ego-centric spatial distinctions such as proximity to speaker, can, and often are, encoded in deictic systems, the heart of deixis is not space, but *access*. As Hanks (2009, p. 12) puts it: "The question is not, 'Where is the referent?' but 'How do we identify the referent in relation to us?'" The diachronic trajectory we trace in this article—from backchanneling cues to demonstrative modifiers, and from there, to more grammatical/functional units, reflects the idea that demonstrative reference is not grounded in spatial relations, but rather, in the multidimensional, intersubjective worlds within which pointing makes sense. It is not surprising, from this perspective, that interactional signals associated most closely with attention-modulation are precisely the kind of thing a protactile signer would intuitively seize on in building a new deictic system. This connection draws us away from thinking about language emergence as deriving from the grammaticalization of "space" (Coppola and Senghas, 2010) and instead shifts attention to the grammaticalization of intersubjectivity (Evans et al., 2018a,b, p. 113). From there, the richness of the semiotic input does not derive merely from the internal, structural features of a gesture or set of gestures available in the environment, but from any aspect of the environment that speaker and addressee can converge on as meaningful.

The Grammaticalization of Intersubjectivity

In its most basic formulation, pointing is a mechanism for intersubjective coordination. The general tasks involved can be distributed over interlocking semiotic systems—such as grammatical systems, co-linguistic gesture, and interactional signals. Cross-linguistic research on spoken language has shown that deictic systems are particularly powerful in this regard, since they not only direct attention to objects of reference, but they do so according to diverse and conventional cues regarding what, where, and how to attend (Hanks, 1990; Diessel, 1999; Bühler, 2001 [1934]; Evans et al., 2018a). For example, Yucatec Maya offers speakers the option of signaling—by way of three distinct and conventional enclitics—that the referent is tactually, visually, or audibly accessible (Hanks, 2009, p. 14). Jahai, a language spoken in Malaysia, offers an "elevation" set, which includes distinctions such as "superjacent vs. subjacent," i.e., located above the speech situation, as in "overhead, uphill, or upstream" vs. located below the speech situation, as in "underneath, downhill, or downstream" (Burenhult, 2003 cited in Evans et al. (2018a), p. 129, also see Forker, 2020). Each of these categories primes receptive attention in the addressee in its own, special way. Deictics, then, are a key resource, which can be drawn on by speakers and addressees to build up intersubjective access to, and knowledge about a shared world.

The protactile movement led to a radical re-configuration of human-human and human-environment relations (Edwards, 2018). As part of this, protactile signers learned to attend to one another and their environment in ways that were expectable to others, given no presumed access to visual channels (Edwards, 2015). In this article, we show how these new, reciprocal modes of attention are being enshrined in grammar. This case highlights the fact that deictics are not only tools for modulating attention; they also act as repositories for routine modes of access, including the channels along which attention can dependably be directed (Hanks, 2009, p. 22). Insofar as elevation or differences in visual, tactile, and auditory sources of information can function as organizing dimensions of life, they can become useful landmarks, and insofar as those landmarks are routinely referred to, they can be imprinted on the language as a system of choices for how to expediently orient the "searching perceptual activity" of one's interlocutors (Bühler, 2001 [1934], p. 121). In other words, deictic systems anticipate the intersubjective worlds that shape them. The present study offers a glimpse of how that anticipatory capacity begins to develop in the earliest stages of language emergence, and the role demonstratives play. To this end, we begin with a brief summary of recent findings regarding the structure of proprioceptive constructions, which is necessary for understanding how taps are entering the grammar of protactile language.

PROPRIOCEPTIVE CONSTRUCTIONS

Expressions for events of motion and location have been an area of protactile grammar where a great deal of innovation has taken place, and in this section we describe some of the innovations that provide a backdrop for our analysis of demonstratives

(further details can be found in Edwards and Brentari, 2020). In contrast to visual signed languages, where sign production involves the two articulators of the signer, protactile language has four potential articulators: the hands and arms of Signer 1 (the "speaker") and the hands and arms of Signer 2 (the "listener"). The incorporation of the listener's body as part of the articulatory apparatus yields a new kind of articulatory space, unattested in the world's languages. Granda and Nuccio (2018) call this space "contact space," and they distinguish this sharply from "air space," used in visual sign languages such as ASL. In air space, locations are perceived relative to each other against a visual backdrop, such as the nose and eyes of the signer. In contact space, locations on the body of Signer 2 are activated and perceived against the backdrop of their own body.

For example, in **Figure 2**, Signer 1 (right) is describing a lollipop to Signer 2 (left). The cylindrical stick of the lollipop is represented by the arm of Signer 2 ("A2"), as is the spherical candy portion. Their spatial relationship to one another is clear, since those relations are perceived by Signer 2, via proprioception, in the movements and positionings of their own body. Incorporating Signer 2's body into the articulatory system unlocks great potential in the proprioceptive channel. However, it also generates a problem for the language: how can the articulators of Signer 1 and Signer 2 be coordinated in an efficient and effective manner?

One of the earliest stages in the emergence of protactile phonology resolves this problem by establishing conventional ways of signaling how and when Signer 1 wants Signer 2 to contribute to co-articulation of signs. Edwards and Brentari (2020) show that the conventionalization of such mechanisms involves assigning specific linguistic tasks to four articulators ("A1"–"A4" in **Figure 2**) in much the same way that the two hands in visual signed languages ("H1" and H2") are assigned consistent and distinct tasks (Battison, 1978). A detailed account is available in Edwards and Brentari (2020). Here we provide a summary of those findings, which is required for understanding the synchronic and diachronic analysis of protactile demonstratives presented in this article.

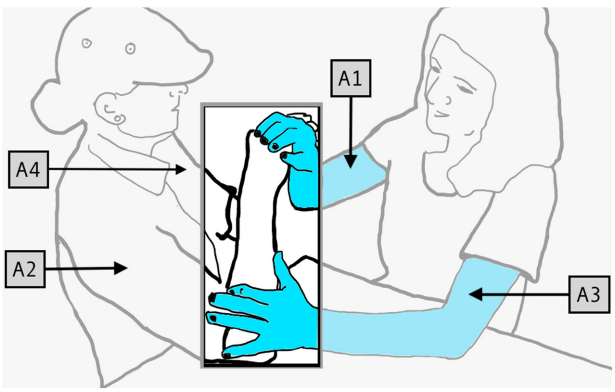


FIGURE 2 | Four articulators used to produce PCs.

Each PC includes at least one unit from each of the following categories, labeled according to their role in the larger construction: initiate (I), proprioceptive object (PO), prompt to continue (PTC), and movement-contact type (MC). These units, which are defined in **Table 1**, combine in the order given, to form a unified construction via rapid interchange between Signer 1 and Signer 2.

Initiate

In the temporal unfolding of the PC, the first to occur is "initiate". As its name suggests, its function is to *initiate* a four-handed construction. There are seven attested types of initiate in the data we analyze here, one of which, is: "INITIATE-PROMPT-TAP". In **Figure 3A**, Signer 1 (left) produces this form by tapping on the back of Signer 2's (right) non-dominant hand. Signer 1 taps on Signer 2's non-dominant hand ("A4") with her non-dominant hand ("A3"). This alerts signer 2 that their active participation is required, and further instructions will follow. Attested initiates include: TOUCH, GRASP, MOVE, HOLD, TRACE, PO, PROMPT-TAP, AND PROMPT-PO.

Proprioceptive Object

Once the PC has been initiated, a meaningful and phonologically constrained space, on which, or within which, further information can be conveyed must be established. We call that space, which is actively produced by Signer 2, the "proprioceptive object," or "PO." In **Figure 3B**, Signer 1 produces a second initiate, telling Signer 2 to select the PLANE PO. In **Figure 3C**, Signer 2 produces the PLANE PO using A2. The PO is significant for understanding demonstrative modifiers, because it generates the discourse-internal referents, to which endorphoric demonstrative modifiers refer. For example, once a PLANE PO, like the one in **Figure 3C** has been produced, Signer 1 can establish relations on the plane, and then tap on locations within it to foreground those locations. Attested POs include: PLANE (with sub-types: PENETRABLE, BENT, INCLINE); CYLINDER (with sub-type: TWISTED); SPHERE (with sub-type: PENETRABLE); INDIVIDUATED OBJECTS; and CONTAINER.

Prompt to Continue

The third task in producing a PC is to maintain the active, contact signing space generated by the PO. It tells Signer 2, "Leave this hand here. There is more to come; or in the case of PUSH, relax this hand, we are done with it." Therefore, we call this category of forms, Prompt-to-Continue (PTC). In **Figure 3D**, after Signer 2 has produced the requested PO (using A2), Signer 1 grips the PO

TABLE 1 | Articulators and signing space in the protactile system.

Articulatory components of Protactile Constructions		
1	Articulator 1	Dominant hand—Signer 1
2	Articulator 2	Dominant hand—Signer 2
3	Articulator 3	Non-dominant hand—Signer 1
4	Articulator 4	Non-dominant hand—Signer 2
5	Contact Space	Locations on or near Signer 2's body—"signing space" for protactile language.

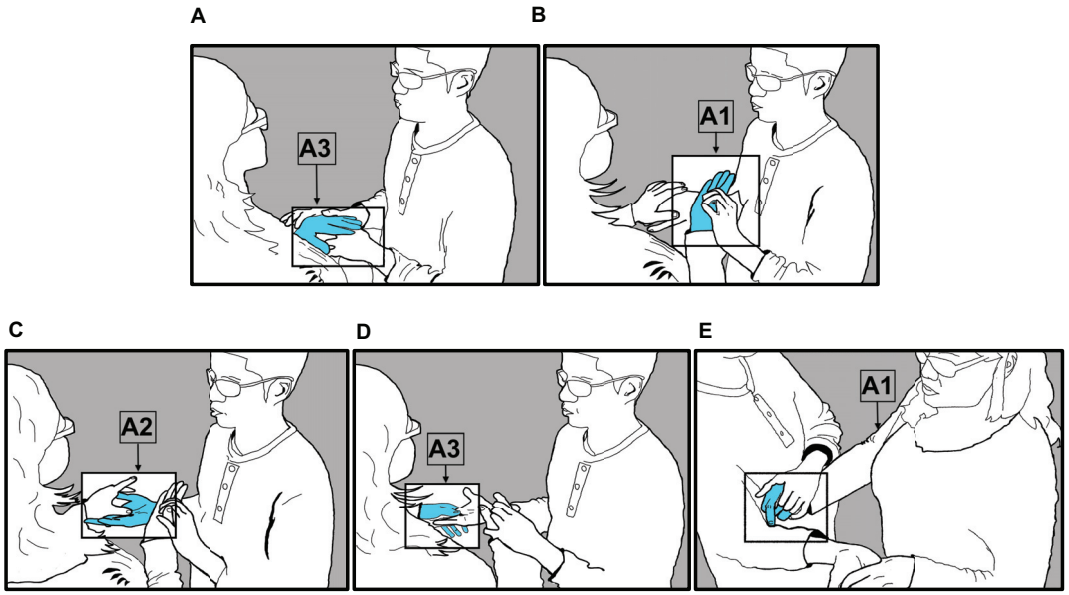


FIGURE 3 | (A–D) Signer 1 (left), Signer 2 (right); **(E)** Signer 1 (right), Signer 2 (left). Proprioceptive Construction (PC) Units: **(A)** a general INITIATE form for the utterance; **(B)** a specific INITIATE form for the PC; **(C)** establishing the proprioceptive object (PO); **(D)** holding the PO in place for further specification (PTC), and **(E)** articulating a movement-contact unit (MCs).

(using A3) and holds onto it for the remainder of the PC. This gripping action is an example of a PTC unit. PTC is significant for understanding demonstrative modifiers, because it maintains the discourse-internal referents, to which endophoric demonstraive modifiers refer. It also maintains an active signing space, within which, demonstrative modifiers can be articulated. Attested PTCs include: GRIP, PENETRATE, PRESS, HOLD, and PUSH.

Movement Contact Type

The fourth task in producing a PC is to draw attention to, and characterize, certain aspects of the PO, or a language-external referent, by producing tactile and proprioceptive cues that contain information about size, shape, location, or movement of an entity. These cues are called "Movement Contact Types," or "MCs". For example, in **Figure 3E**, Signer 1 (left) uses A1 (her right hand) to *trace* a line from the palm of A2 to the inside of the elbow. **Figure 3E** shows the end of a SLIDE describing a long, rectangular object. Attested MCs include: TRACE (with sub-type PO); GRIP (with sub-types TWIST, WIGGLE SLIDE, PULL, TRILL); SLIDE (with sub-type TRILL); PENETRATION; TAP (with sub-type TRILL); press (with sub-types WIGGLE and PO); SCRATCH; MOVE; and PUSH.

It is within this PC structure that taps can be differentiated formally and functionally. In the next section, after describing our methodology, we present evidence for establishing such distinctions, synchronically.

STUDY DESIGN AND PROCEDURES

In this article, we report on two studies. Study 1 is a synchronic analysis of the most recent data, collected in 2018. This analysis shows how taps function within, and are constrained by, protactile phonology. The results of this study will help orient the reader to the landscape of the current system. Study 2 zeros in on taps, tracing the different functions they take on over time, and how those functions change. This longitudinal study examines data collected at four moments in the emergence of the protactile language: In 2010, in the early stages of emergence; in 2015, 2016, and 2018. In the sections that follow, we include in-depth information about data collection, participants, procedures, stimuli, and transcription for each data set.

STUDY 1 METHODS

Procedures

Recruitment took place in two phases. First, an email was sent to relevant community leaders explaining the project and requesting participation. That email was shared by them to a group of potential community members. A local DeafBlind educator then selected a subset of those who responded, based on her evaluation of high protactile proficiency. During data collection events, prior to filming, we gave consent forms to participants in their preferred format (e.g., Braille or large print). We also offered the services of qualified interpreters who could translate the consent form into protactile language. The first author, who is fluent in protactile language, offered to discuss the consent forms with each of the participants and answered questions/offered clarification, where requested. The consent forms included questions requesting permission to include images of these communication events in published research and other research and education contexts, such as

conferences and classrooms. Once consent had been obtained, we commenced with data collection.

The 2018 data were generated in a description task, designed to elicit PCs. Data collection took place a privately owned home. Dyads of protactile signers were asked to stand next to a "cocktail" table—or a small, round table, which was tall enough to comfortably reach stimuli, while standing. Tactile landmarks were placed on the ground to signal locations where the cameras could pick up linguistic productions. The interactions were always between two protactile signers, both of whom were participants in the study. The stimuli were placed on the table in pseudo-random order and Signer 1 was instructed to "describe what they feel". Signer 2 was told that Signer 1 would be describing something they felt. After the description, Signer 2 was offered an opportunity to feel the stimulus. After a certain number of stimuli, Signer 1 and Signer 2 changed roles. However, sometimes, after feeling the stimulus, Signer 2 chose to repeat a description with added changes or feedback. One of the co-authors and one member of the research staff were present throughout the task to operate the video cameras and place stimuli on the table, but were only in tactile contact with the participants while placing stimuli. The cameras were attached to the ceiling, using hooks and wire, and pointed down toward the participants, in order to capture contact and motion between them.

Participants

The six participants in this study (3 males, 3 females, ages 32–53) were all DeafBlind individuals, who had participated in a protactile network for at least 3 years. The average number of years participating in a protactile network across the group was 6 years, and the range was 3–11 years. All but one of the participants were exposed to a visual sign language prior to age 5, via visual perception (those who became blind in adolescence or adulthood). One participant (who was born blind) had access to a visual sign language via tactile reception since birth. At the time these data were collected (2018), all six participants had been working full time in an environment where protactile language was in wide-spread use. Three of the participants also lived with protactile signers, and all of the participants attended informal protactile social events in the evenings and on weekends. In total, they reported an average of 49 h per person, per week, of protactile interaction and language-use. When asked what proportion of that time was spent with DeafBlind protactile signers, and what proportion was spent with sighted protactile signers (either hearing or Deaf), five participants said that most of their time was spent with DeafBlind protactile signers; One participant responded with irritation to the question, saying, "It doesn't matter—the point is, they all know Protactile." All but one of the participants (who reported growing up ambidextrous) reported that they grew up right-handed, but said that since being immersed in protactile environments, they now consider themselves more ambidextrous than they used to be.

Stimuli

Proprioceptive constructions were elicited by presenting a series of tactile stimuli to the participants (**Table 3**). These objects were chosen because they were the same, or had corresponding characteristics as stimuli that were used in prior elicitation sessions, such as shape, size, or the presence of movable parts. Tactile stimuli like these offer opportunities for participants to convey information about motion and location events in protactile language, using real objects that can be explored tactually. The first two were presented as singular objects. The rest were presented in singular and plural conditions, as well as "multiple" conditions, which included a set of three, where two were the same, and one was different, along some dimension.

In the case of the toy car stimulus, differences included size, shape, material, and whether or not the car was self-propelled (i.e., when you press it down, into a surface, and pull back, does it spring forward and travel out of reach? Or does it stay in place?). The lollipop stimulus involved differences in size and type of wrapper. In the case of the pen stimulus, the difference was whether the pen had a cap or was a ball point pen, where the ball point pop out when you press on the end of the pen with your thumb. In addition, some participants described the relative locations of each object on the table.

Transcription

Descriptions of the stimuli were videotaped, labeled, and annotated using ELAN (Crasborn and Sloetjes, 2008). Coding one tier at a time, we identified the units produced by each articulator. In order to identify units of analysis within PCs, we assigned Signer 1 and Signer 2 independent tiers. Recall that Signer 1 is the principal conveyer of information. Signer 2 contributes to the articulation of the message, but in terms of information, is the principal receiver. In visual signed languages, the dominant hand (H1) and the non-dominant hand (H2) are assigned complementary roles; H1 is more active than H2 (Battison, 1978). In protactile language, four anatomical structures are available for producing each sign, which we assign to roles based on the degree to which they are active in linguistic productions as described in **Table 1** : **A1** (dominant hand of Singer 1) > **A2** (dominant hand of Signer 2) > **A3** (non-dominant hand of Signer 1) > **A4** (non-dominant hand of Signer 2). Edwards and Brentari (2020) show that one of the earliest stages in the conventionalization of protactile phonology is the consistent alignment of particular articulators with particular linguistic functions, as described in **Table 2**: A1 is primarily responsible for producing MCs; A2 is primarily responsible for POs; A3 is primarily for producing PTCs; and A4 is rarely involved in linguistic production. As the present analysis highlights, A4 is least active in linguistic productions because its primary role is to track the movements of A1; Indeed, in these data contact between A4 and A1 is rarely broken. Its secondary role is to produce backchanneling cues. While A4 is tracking A1, simultaneous backchanneling cues can be produced by tapping on A1 with a few fingers, while maintaining a light grip with the remaining fingers. A2 can also produce backchanneling cues while performing other, linguistic tasks—though, as is discussed below, this is less common.

TABLE 2 | Functional units of proprioceptive constructions and their associated articulators.

		Functional Units of Proprioceptive Constructions	Articulator
1	Initiate (I)	A request for active involvement of S2 in co-producing a PC	A1, A3
2	Proprioceptive Object (PO)	Active articulatory space- type selected in response to type of Initiate produced.	A2
3	Prompt to Continue (PTC)	Keeps selected articulatory space active for further information to be added.	A3
4	Movement Contact Type (MC)	Tactile and proprioceptive combinations of movement and contact that contain information about size, shape, location, or movement of an entity.	A1

Analyses

Analyses were aimed at identifying articulatory, distributional, and combinatory characteristics, which distinguish each of the taps from one another. To this end, we annotated each functional PC unit, including, but not limited to, taps on the tier corresponding to the articulator that produced it. In addition, we annotated taps produced outside of a PC unit, including taps on objects in the immediate environment, and taps functioning as backchanneling cues. We identified four types of taps, which we coded: I-PROMPT-TAP, MC-TAP, EX-TAP, and BC-TAP (**Table 4**).

In the sections that follow we outline our results and provide figures which contain the quantitative patterns found in the data; however, due to the small number of participants in these students, these results should be considered qualitative in nature.

STUDY 1 RESULTS AND DISCUSSION

Phonological Characteristics

One of the earliest stages in the conventionalization of protactile phonology is the consistent alignment of particular articulators with particular linguistic functions (Edwards and Brentari, 2020). This is, then, a potential resource for protactile signers for distinguishing the functional category to which they belong. For example, In the case of taps, I-PROMPT-TAP and MC-TAP belong to the INITIATE and MC categories, respectively. In the 2018 data, I-PROMPT-TAP and MC-TAP are associated with different articulators. I-PROMPT-TAP is produced with A3 most of the time (79% of 77 tokens), while MC-TAP is produced with A1 most of the time (89% of 392 tokens, **Figure 4**).

Given this clear pattern in articulatory-functional alignment, the articulator used to produce the tap is one dimension along which demonstrative taps and propriotactic taps can be distinguished from one another. In addition, when a demonstrative modifier is used exophorically, its phonological characteristics change: it is not produced in contact space, i.e., on the body of Signer 2, with three articulators A1, A2, and A3. Instead, it is produced on an object in the immediate environment with a single articulator—either A1 or A3. For example, in **Figure 5**, Signer 1 (left) produces an exophoric demonstrative tap (A3) at the edge of the napkin, to indicate which edge he will fold next. The addressee (right) perceives the tap from a "listening position" (A4) but does not actively participate in articulation.

TABLE 3 | Stimuli for study one (2018).

Stimuli	Conditions
Large doll with braids and movable arms	Singular
Electronic braille display	Singular
Toy car ± self-propelled	Singular, plural, multiple
Lollipop	Singular, plural, multiple
Pen (± cap; ± ballpoint)	Singular, plural, multiple

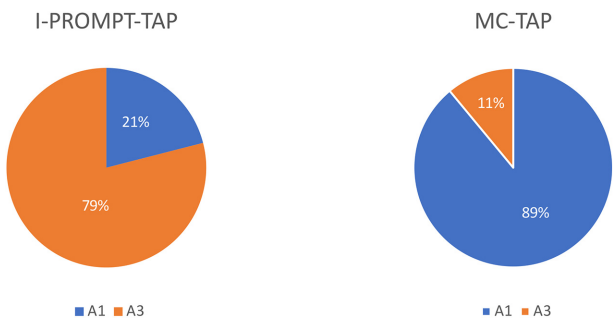


FIGURE 4 | Articulatory-functional alignment in 2018 data.

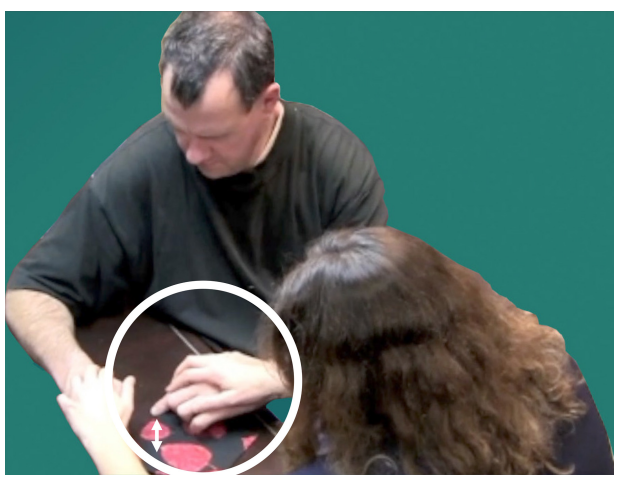


FIGURE 5 | Signer 1 (left), Signer 2 (right). Signer 1 produces an exophoric demonstrative tap (A3) at the edge of a napkin.

TABLE 4 | Tap functional units and coding labels.

Coding Labels for TAP Functional Units		
Proprioactic	Instruction by S1 to S2 to activate A2 for purposes of articulation, and/or that a "prompt-PO" is coming next	I-PROMPT-TAP
Demonstrative modifier	Draws attention to some aspect of a referent, or singles out one referent among others, in contact space	MC-TAP
Exophoric demonstrative modifier	Draws attention to some aspect of a referent, or singles out one referent among others, in the immediate environment	EX-TAP
Backchanneling signal	Signals S2's continued attention and/or agreement	BC-TAP

Finally, backchanneling taps can be distinguished from exophoric and endophoric demonstrative modifier taps and proprioactic taps, since they are produced by Signer 2 by either A2 or A4 (the two articulators, which belong to the "listener," or Signer 2).

Combinatory Patterns

Recall that proprioactic taps are a type of INITIATE, which function as an instruction from Signer 1 to Signer 2 to activate A2 for purposes of articulation and/or signal that another, more specific instruction will soon follow. We observe in these data that proprioactic (I-Prompt) taps cannot combine with other units to add information to that request. Signer 1 cannot, for example, produce the tapping motion with two fingers to signal that they are requesting two articulators instead of one. In contrast, demonstrative modifier taps can and do participate in various combinations. In addition, while proprioactic taps can only be used endophorically, demonstrative modifiers can be used alone (i.e., outside of a PC context) to refer, exophorically, to an object in the immediate environment (**Figure 5**, i.e., "this"), or they can combine with other meaningful elements in contact space to express information about number and/or location (e.g., "these two here" or "this here"); identify one item in a sequence of items (e.g., the second one of these three); information about shape (e.g. this cylindrical-thing"); or information about size (e.g., "this large one"). For example, in **Figure 6A**, Signer 2 (left) produces a PO representing three individuated objects (A2), after being prompted to do so by Signer 1. In **Figure 6B**, Signer 1 taps on the second of the three (A1), to indicate that the referent is second in the series of three.

Distribution

Proprioactic (I-Prompt) taps occur at the beginning of a PC in the "initiate" slot, whereas the demonstrative modifier taps occur in the third position in a PC—the "MC" slot (**Figure 3**).

Types of Taps in Protactile Language and Communication

Based on these differences in the 2018 data set, we propose the following types of taps in protactile language and communication:

(i) **backchanneling:** interactional signal for continued attention or agreement, is produced by Signer 2 using A2 or A4; occurs in various positions within, before, and after the completion of a PC; and cannot combine with other units.

(ii) **exophoric demonstrative modifier (Ex-tap)**: a TAP used to draw attention to, and add information about, a referent in the immediate environment; is produced most often with A1 and in all other cases, with A3; occurs outside of the PC context; and can combine with handshapes to indicate information about size, shape, and location.

(iii) **endophoric demonstrative modifier (MC-tap)**: a TAP used to draw attention to, and add information about, a referent in the unfolding discourse; is produced most often with A1 and in all other cases, with A3; occurs in the third position of the PC; and can combine with other PC units to convey information about size, shape, location, and movement of referent.

(iv) **proprioactic (I-Prompt)**: a TAP used to draw attention to, and request an action from, an articulator belonging to Signer 2; is produced most often with A3 and in all other cases, with A1; occurs in the first position of the PC; and cannot combine with other PC units.

While types i-iii are commonly found cross-linguistically in both signed and spoken languages, type iv—proprioactic taps—have, to our knowledge, never before been reported[4].

[4]As a general rule, it is assumed that turns at talk occur sequentially—that is, participants endeavor to speak "one-at-a-time" (Sacks et al., 1974, p. 696–735). Momentary deviations from this rule are permissible, for example, in the anticipatory completion of the other's turn (Lerner, 1989, 1996), overlapping turns, as in assessments (Goodwin and Goodwin, 1987), and in collective turns, for example, when a group collectively greets a newcomer (Lerner, 1993; Schegloff, 2000). These activities display a greater range of interactional functions than backchanneling, and also are not seen by participants as problematic in the way that an "interruption" would be. Rather, they are treated as permissible, momentary deviations. In addition to these forms of overlap, speakers may grant "conditional access to the turn" (Schegloff, 2000, p. 5, and see also Lerner, 1996, p. 239). This occurs when a second speaker is invited to complete the turn of the first speaker, "conditional on the other's use of that opportunity to further the initial speaker's undertaking" (Schegloff, 2000, p. 5). For example, the "word search," where the second speaker will be invited to produce the name of someone that the initial speaker is having trouble retrieving. This could be considered a self-initiated other-repair, since the speaker flags a source of trouble in the interaction and the addressee executes the correction (thanks to Simeon Floyd for pointing this out, personal communication). The phenomenon we are describing among protactile signers is most similar to conditional access to the turn, where the "invitation" to participate has become conventionalized in the initiate category of the proprioceptive construction. There are three conventional ways to "invite" the second speaker to co-produce signs— INITIATE-TOUCH, INITIATE-GRASP, and INITIATE-PROMPT. In addition, the production format (Goffman, 1981) is the same in cases of conditional access to the turn and PCs—while the animator role is distributed across two participants, the author role remains assigned to a single

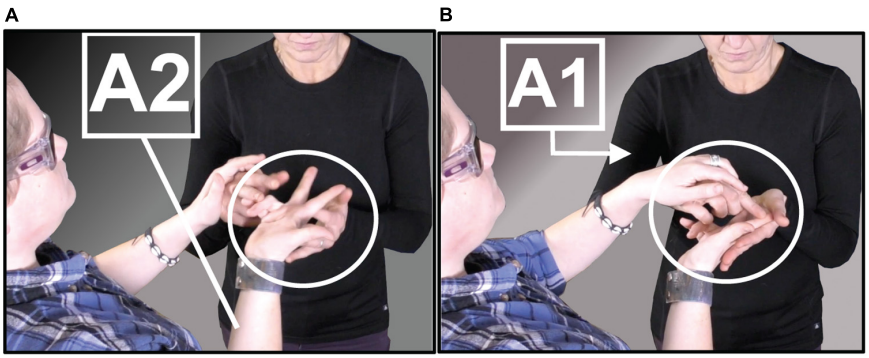


FIGURE 6 | Signer 1 (right), Signer 2 (right). In **(A)**, Signer 2 produces PO-INDIVIDUATED-OBJECTS-3 (A2); In **(B)**, Signer 1 produces MC-TAP (A1) on the second of three individuated objects, to indicate position of referent in sequence.

These forms are a type of Initiate, which function as a kind of "reception signal" (Bühler, 2001 [1934], p. 122). They tell Signer 2: " Be receptive here in this region of signing space—more information is coming." For the phonological system, the presence of such a unit helps to optimize language to the tactile modality by shifting the ground of perception to the body of the addressee, where articulatory distinctions are made accessible through tactile channels, alone. In contrast, in visual sign languages, phonologically distinctive locations are perceived relative to each other against a visual backdrop that is inaccessible for DeafBlind signers (e.g., "to the right of the mouth" vs. "to the right of the eye"). Proprioceptive taps function as a conventionalized mechanism for activating the *listener's* body, rather than the *signer's* body, as the ground of perception **Table 5**. This new kind of unit, then appears to be specific to the tactile modality. In the remainder of this article, we examine the relationship of propriotactic taps to similar forms that emerged earlier in the roughly 10-year history of protactile language.

STUDY 2 METHODS

Study 2 is a diachronic analysis of the emergence and conventionalization of the categories of taps identified in Study 1. Data include those from 2018, as well as those that were collected three times prior to 2018: in Seattle in 2010, early in the emergence of protactile communication practices (Edwards, 2014); in Seattle in 2015, as protactile practices were becoming conventionalized; and in 2016, when conventionalized practices were being transmitted from Seattle signers to a new group of DeafBlind students in Washington DC. Below, we provide information about the study design and procedures for each data collection event.

participant. However, there is one significant difference between conditional access and the co-construction of PCs: This collaborative construction of the utterance is not an exception to a general rule, it *is* the general rule—this is how protactile utterances are produced.

Design and Procedures

The data used in this diachronic study were not collected in an identical manner, as is often the case when there are long intervals between data collection sessions, with changing linguistic and social circumstances. We describe each of the data sets we use in this study in the following sections.

2010 Data Collection

2010 recruitment procedures took place as part of a year-long period of sustained ethnographic fieldwork conducted by the first author. First, several meetings were held with relevant community leaders, in order to identify a context that would be appropriate for linguistic and interactional research. In those meetings, the community gave permission to the first author to videorecord a series of protactile workshops, where 11 DeafBlind signers and 2 instructors/organizers met twice weekly for 2.5 h, for a total of 10 weeks, in order to experiment with protactile communication in a range of activities. The workshops were held in a private room within a DeafBlind organization in Seattle.

Participants selected for the workshops by the DeafBlind instructors, were invited to discuss the research at length in individual meetings, prior to the start of the workshops, with Edwards. After those meetings, they made an informed decision of whether or not to participate. If they chose to participate, they were given a consent form in their preferred format (e.g., braille or large print). Edwards also offered to interpret the form into the preferred language of the participant. The consent forms included questions requesting permission to include images of interactions in the workshops in published research and other research and education contexts, such as conferences and classrooms. Once consent was obtained, Edwards and a team of videographers proceeded to videorecord approximately 120 h of interactional data generated during the 2010 protactile workshops, which were subsequently labeled, organized, and stored.

For the purposes of this study, we reviewed these data, looking for contexts that were maximally similar to the elicitation contexts created for the 2018 study. This included activities where objects, such as a tea strainer, a movable toy snake, or a phone charger, were being described by one DeafBlind participant to another; when objects were referred

TABLE 5 | Place of articulation for each type of tap.

Type of tap	Sender	Place of articulation
endophoric demonstratives	Signer 1	contact space often on a PO (Signer 2): A1
propriotactic taps	Signer 1	contact space often on a PO (Signer 2): A3
exophoric demonstratives	Signer 1	on an object in the environment
backchannel	Signer 2	contact space on Signer 1

to as part of demonstrations/instructional activities, where one DeafBlind participant explains how to do something, such as use a crochet hook; and direction-giving activities—all of which were organized by the DeafBlind instructors. While these contexts were not elicitation contexts, and the objects introduced in the workshops were not framed as "stimuli," we think these contexts offer an opportunity for comparison with the more targeted elicitations we conducted later.

2015 Data Collection

The 2015 recruitment procedures and consent process were identical to those described above for Study 1, and the consent forms included questions requesting permission to include images of interactions in the workshops in published research and other research and education contexts, such as conferences and classrooms. The data were generated by a description task, designed to elicit PCs using three stimuli: a lollipop, a jack (the kind children use to play the game "jacks"), and a complex wooden toy involving movable arms, magnets, and magnetized pieces. The first two stimuli were presented in a singular context (1 object) in a plural context (several of the same object in a row), and a "multiple" context (2 objects that are the same and one that differs in size, shape, or movability).

Data collection took place at a dining room table in a private home by both co-authors. Dyads of protactile signers were seated at the corner of the table. The interactions were always between two protactile signers, both of whom were participants in the study. They changed roles after a given item was completed, and discussed and gave feedback to one another about the clarity of a description, as it unfolded. We placed a cloth napkin with thick edges on the tabletop to provide a tactile boundary within which the stimuli would be placed. The stimuli were placed on the napkin in pseudo-random order and Signer 1 was instructed to "describe what they feel." Signer 2 was told that Signer 1 would be describing something they felt. After the description, Signer 2, who was not exposed to the stimulus prior, picked up the object and explored it tactually. The co-authors were present throughout the task to operate the video camera, but were only in tactile contact with the participants when placing stimuli. The camera was on a tripod on the table, positioned above the participants pointing down, in order to capture contact and motion between them. In all cases, the dyads discussed aspects of the object and adjusted their descriptions—sometimes at great length. In addition, the stimuli had many different pieces and parts, each of which was described by the participants. Therefore, we collected a large number of tokens in response to a limited number of stimuli.

2016 Data Collection

The 2016 data we analyze were collected in two events. First, the two organizers of the 2010 Seattle workshops hosted a second set of protactile workshops in Washington, DC for a group of DeafBlind signers who were actively involved in local protactile networks. We analyze a series of interactions between 5 of the 7 participants, who took part in a direction-giving exercise. During this exercise, each participant was asked to give directions to locations within the building, to nearby buildings, and to other locations in the district. Workshops took place in an auxiliary classroom space at a local university. The second source of data from 2016 was generated in a description task, led by the first author, as well as a direction-giving task, similar to those that occurred in the workshops, also led by the first author. These data collection events took place in a lab on a university campus. The stimuli that were used in the description task included a soft block made out of fabric, a lollipop, and a jack. Two of the workshop participants were included in these sessions, along with one additional protactile signer. This yielded a total of 8 participants in the 2016 data set as a whole. 2016 recruitment procedures and consent were the same as those described above for the 2010 protactile workshops. However, instead of being invited to videorecord all of the sessions, the first author was invited to videorecord a subset of sessions, as determined by the group.

Participants

There were a total of 15 participants in Study 2 (6 males and 9 females, ages 32–53 in 2018).

Thirteen participants were born sighted or partially sighted and acquired ASL as children via visual reception (all but one by the age of 7). 2 participants were born blind, and acquired ASL via tactile reception (both prior to the age of 7). 12 of the 15 participants were, at the time of data collection, immersed in protactile environments—at work, where protactile language was in wide-spread use, and/or in the evenings and on weekends, when they attended community events, or interacted with their protactile roommates. Three of the participants interacted with protactile signers somewhat often, according to their own reports, but with less frequency and consistency than the others, as they did not work in environments where protactile language was widespread. In **Table 6**, an X is placed under data collection event(s) for which each participant was present.

Participants 1 and 2 (**Table 6**) were leaders in the community, and took on the role of instructor or facilitator in the data collection events they were present for. They had been in close contact as colleagues since 2007 and during that period, developed a framework for thinking about tactile ways of doing everyday tasks, including communication. They hosted the 2010 workshops together in an effort to broaden their efforts across the community. From the outset, then, they had more experience with "protactile principles" (Granda and Nuccio, 2018) than any of the other participants. Participants 2–7 were exposed to protactile principles for the first time during the 2010 protactile workshops and were involved in the early innovations described below. Participants 8–15 who resided in either Seattle or Washington DC (or both at different times during the study)

TABLE 6 | Longitudinal participation.

	2010	2015	2016a	2016b	2018
Participant 1	X	X	X		
Participant 2	X	X	X		X
Participant 3	X	X			X
Participant 4	X				
Participant 5	X				
Participant 6	X				
Participant 7	X				
Participant 8		X			X
Participant 9		X			
Participant 10		X			
Participant 11		X	X	X	X
Participant 12			X	X	X
Participant 13				X	
Participant 14			X		
Participant 15					X

were exposed to protactile principles at least 1 year prior to their participation in their first data collection event, all via contact with protactile signers from Seattle.

Transcription and Analyses

In a first pass, we located moments in each data set, when participants were asking questions such as "which one?" "where?" or moments where it seemed that the signer was trying to contrast one thing compared with another—e.g., if there were two chairs, and the signer was trying to draw attention to the one they wanted their interlocutor to sit in. We also looked at descriptions of objects with multiple parts, such as a lollipop (including the candy, the stick, and the wrapper), a series of such objects laid out on a table, and a series of such objects, where one differed in size, shape, or some other characteristic from the others. We identified moments in these descriptions when the signer foreground aspects of the object against a background (often paired with some characterization, i.e., "this [MC-TAP] + spherical thing [PO-SPHERE]"), or one object in a series against the others (i.e., "this one [MC-TAP] is the larger one"). We observed that MC-TAP was the most common form used in such contexts. We identified several additional taps, as well, which had related, but not identical functions: backchanneling taps (which we labeled "BC-TAP"), exophoric demonstrative modifier taps (which we labeled "EX-TAP"), and propriotactic demonstrative taps (which we labeled "I-PROMPT-TAP").

We imported all video data (described above) into ELAN (Crasborn and Sloetjes, 2008). We created one tier for each of the four articulators (A1, A2, A3, and A4), and annotated each functional PC unit, including, but not limited to taps, on the tier corresponding to the articulator that produced it. In order to determine whether a form was in use, and how frequent its use was, we counted numbers of tokens per minute of active, transcribed, signing time. We also recorded the proportion of signing time spent in the "Signer 1" vs. "Signer 2" roles. Diachronic analyses, showing changes in the rate of occurrence of each category of taps is presented below.

STUDY 2 RESULTS AND DISCUSSION

The results of Study 2 show that over time, the total number of taps produced by protactile signers are (a) becoming differentiated into a greater number of structural and semiotic types; and (b) being distributed differently across those categories in ways that optimize semiotic load (see **Figure 7**).

In the 2010 data, 483 taps were produced and in the 2018 data, 557 taps were produced. In 2010, backchanneling taps ("BC-taps") were the most common, making up 52% of the total. Endophoric taps ("MC-taps") were next, at 34%, followed by exophoric taps ("Ex-taps"), which accounted for 14%. Propriotactic taps ("I-Prompt-taps") were not present in 2010 at all. From 2010 to 2018, we see the emergence and steady increase of propriotactic taps to reach 14% of the total by 2018. This coincides with a decrease in exophoric taps from 14% to 0. This suggests that a device for requesting the active participation of the addressee in sign production (i.e., the propriotactic tap), reduces dependence on exophoric reference.

In a parallel pattern, we see the proportion of endophoric (MC) taps increase from 34 to 71%, while backchanneling decreases from 53 to 15%. It may seem intuitive that an increase in affirmative backchanneling suggests an increase in understanding. However, early protactile communication was tenuous, and comprehension could not be taken for granted. Without conventionalized mechanisms for unlocking contact space, and without established modes of intersubjective access to the environment, consistent reassurance was necessary. As propriotactic taps and endophoric demonstratives emerged and became conventionalized, such frequent confirmation became far less necessary.

Of these four types of taps, we can see in **Figure 7** that there are actually two pairs of taps that interact in terms of frequency, and, we would like to suggest, also in function. Backchannel taps and exophoric taps seem to lay the ground work for the two new types of taps, made possible by the proprioceptive construction that is argued for in Edwards and Brentari (2020). Both backchanneling taps and MC-taps index elements already present in the discourse, though of different types. Backchannel cues are responses to what was just said by the other person in the dyad, while MC-taps refer anaphorically or cataphorically to an element within the proprioceptive construction produced by the signer. This possibility derives from the conventionalized structure of the propriotactic construction: MC-tap (like all MCs) is interpreted in terms of its relevance to the preceding PO in the proprioceptive construction. In a parallel fashion, both exophoric taps and propriotactic taps introduce *new* information into the discourse, but, again, of different types. Exophoric taps introduce new entities from the surrounding environment. Propriotactic taps introduce new entities by way of new proprioceptive objects (POs).

In order to analyze the emergence of MC-taps as compared to backchannel taps, and I-prompt taps as compared to exophoric taps, within-subject comparisons were carried out across two time points. For this analysis we compared individuals who were active in the dyad at least 20% of the time, and whose data were

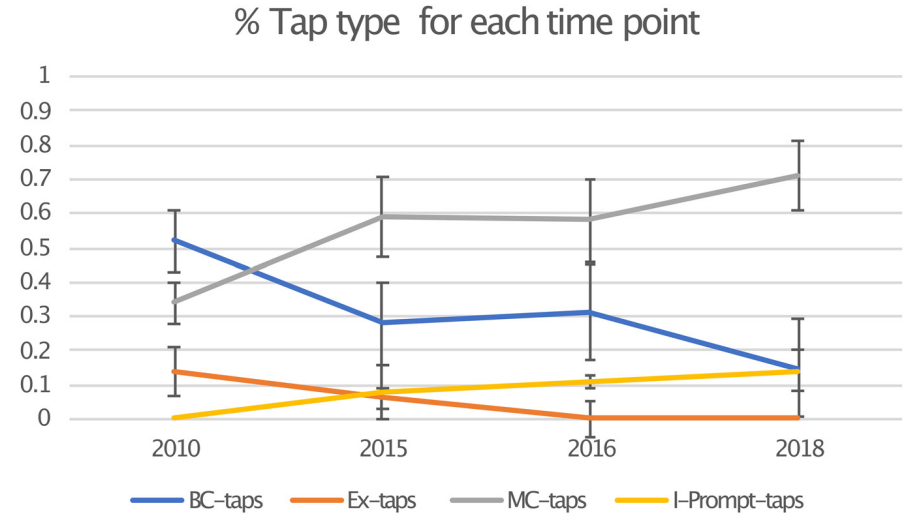


FIGURE 7 | Semiotic re-distribution across categories (with standard error bars).

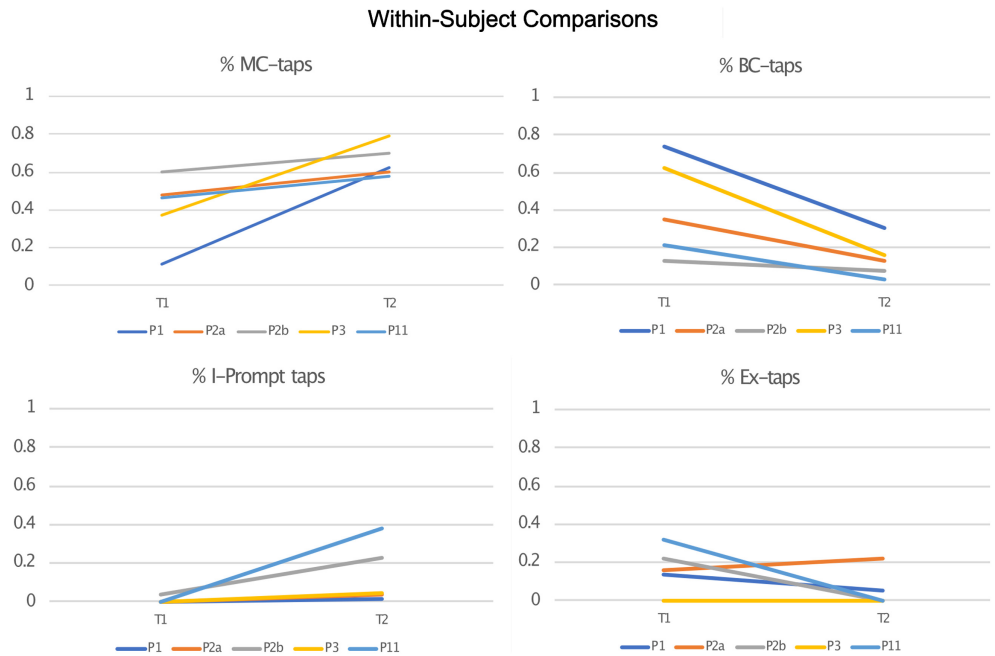


FIGURE 8 | Within-subject comparisons of proportion of taps in each target group.

sampled at time points that were at least 2 years apart. Five of the 15 subjects[5] met these criteria and are presented in **Figure 8**. In all within-subject comparisons, MC-taps and I-prompt taps increased over time, all Backchannel taps decreased over time, and in four out of five comparisons exophoric taps (one increased from 16 to 22%).

In analyzing these data, we were struck by the degree to which protactile signers struggled to communicate with each other in 2010. This often led to "checking" or "proving" that descriptions and instructions were accurate, by guiding one's interlocutor to the aspects of the environment under discussion. By 2018, many of the mechanisms that were mere experiments in 2010 had become conventionalized, and therefore, there was a level of confidence in production, reception, and comprehension, which seemed to obviate strategies that involved excessive dependence on exophoric

[5]Participant 2 had two sessions that met the criteria for inclusion which are labeled Participant 2a and Participant 2b.

reference. We expect exophoric reference to remain available (despite its absence in the most recent data presented here). However, it is highly desirable for communicators to be able to give directions to a location in the immediate environment without walking the person to that location; It is also desirable to be able to describe or depict an object without having that object handy. It seems that these abilities were made possible by a process of semiotic redistribution across systems—I-Prompt taps are primarily organized by, and in service of, the linguistic system, broadly construed, while exophoric demonstratives are organized by, and in service of, an emerging deictic system. Over time, these systems have come to work in tandem, distributing attention-modulation tasks in ways the optimize the linguistic system to the intersubjective environment of protactile signers.

CONCLUSION

In this study, we have followed the diversification and distribution of taps in protactile language. We have shown that backchanneling taps, which maintain continuity of attention across utterances, gain a new, related function in MC-taps, which use the developing linguistic system to maintain continuity of attention across related elements *within* a single utterance. Exophoric taps and propriotactic taps both introduce new entities into the discourse. Exophoric taps do so by directing attention to an object of reference in the immediate environment. This function then expands to include propriotactic taps, which introduce new entities into the discourse via proprioceptive constructions. The PC helps to optimize language to the tactile modality by incorporating the body of the addressee into the articulatory system, thereby making the proprioceptive sense available for purposes of perception. In doing so, it also offers a structure, within which, anaphoric and cataphoric reference can be reliably achieved.

Prior anthropological research has revealed the significant work protactile people have done to re-route reciprocal modes of attention through tactile channels, generating a new and re-contoured environment within which communication unfolds (Edwards, 2014). In this article, we have shown how protactile demonstratives are scaffolded on conventional backchanneling signals that emerged as part of, and were instrumental in, that process, and how from there, forms with more specialized grammatical functions began to emerge. In line with prior research, this supports the idea that deixis plays an important role in language emergence (e.g., Coppola and Senghas, 2010; De Vos, 2014; Kocab et al., 2015; Mesh, 2017). However, the path protactile language has taken, suggests a different basis for the connection. Coppola and Senghas (2010, p. 17) observe that "[g]rammaticalization processes need original forms on which to operate," and following Heine et al. (1991) and Bybee (2003), they propose that "the sources for grammar are drawn from the most universal concrete and basic aspects of human experience, particularly the spatial environment and parts of the body." The evidence presented here turns attention instead toward intersubjectivity as a potentially universal basis for grammaticalization. If languages are built under intersubjective pressures, we would expect that as a new language emerges, its grammar would develop sensitivities not only to space, but to whatever elements and relations are routinely and reciprocally accessible to its speakers.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available in order to protect the confidentiality of the participants to the greatest degree possible. Requests to access the datasets should be directed to TE, terra.edwards@slu.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB, Saint Louis University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

TE is Assistant Professor of Anthropology at Saint Louis University, and has been conducting linguistic and anthropological research in DeafBlind communities in the United States for more than 14 years. DB is Professor of Linguistics at the University of Chicago and has written extensively on the phonology of signed languages.

FUNDING

The National Science Foundation (BCS-1651100) supported the synchronic study reported in this article, different stages of the diachronic study were supported by the National Science Foundation (BCS-1651100), the Wenner-Gren Foundation for Anthropological Research Dissertation Fieldwork Grant (8110), and Post Ph.D. Research Grant (9146).

ACKNOWLEDGMENTS

Thank you to all of the DeafBlind people who participated in this research. Thanks to Jelica Nuccio, Halene Anderson, Vince Nuccio, Abby Clements, and Lilia McGee-Harris for their research assistance. We also thank those who contributed to earlier phases of the project, including Yashaira Romilus, Oscar Serna, aj granda, and Eddie Martinez. For comments on earlier drafts of this article, we thank Simeon Floyd, Constantine Nakassis, Chris Prener, Kenny Coventry, Bencie Woll, and Rose Stamp.

REFERENCES

Abner, N., Flaherty, M., Stangl, K., Coppola, M., Brentari, D., and Goldin-Meadow, S. (2019). The noun-verb distinction in established and emergent sign systems. *Language* 95, 230–267. doi: 10.1353/lan.2019.0030

Agha, A. (1994). Honorification. *Annu. Rev. Anthropol.* 23, 277–302. doi: 10.1146/annurev.an.23.100194.001425

Battison, R. (1978). *Lexical Borrowing in American Sign Language*. Silver Spring, MD: Linstock Press, 240.

Benveniste, E. (1971). *Problems in General Linguistics*. Coral Gables, FL: University of Miami Press.

Boyes-Braem, P. K. (1981). *Features of the Handshape in American Sign Language*. Doctoral Dissertation, University of California, Berkeley, CA.

Brennan, M. (1990). *Word Formation in BSL*. Doctoral Dissertation, University of Stockholm, Stockholm.

Bühler, M. (2001 [1934]). *Theory of Language: The Representational Function of Language*. Philadelphia, PA: John Benjamins.

Burenhult, J. (2003). Attention, accessibility, and the addressee: the case of the Jahai demonstrative ton. *Pragmatics* 13, 363–379. doi: 10.1075/prag.13.3.01 bur

Bybee, J. (2003). "Cognitive processes in Grammaticalization," in *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, ed. M. Tomasello (Mahwah, NJ: L. Erlbaum), 145–167.

Coppola, M., and Senghas, A. (2010). "Deixis in an emerging sign Language," in *Sign Languages*, ed. D. Brentari (Cambridge: Cambridge University Press), 543–569. doi: 10.1017/cbo9780511712203.025

Crasborn, O., and Sloetjes, H. (2008). "Enhanced ELAN functionality for sign language corpora," in *Sixth International Conference on Language Resources and Evaluation*.

Dachovksy, S. (2018). *Grammaticalization of Intonation in Israeli Sign Language: From Information Structure to Relative Clause Relations*. Ph. D. Thesis, University of Haifa, Haifa.

De Vos, C. D. (2014). The kata kolok pointing system: morphemization and syntactic integration. *Top. Cogn. Sci.* 7, 150–168. doi: 10.1111/tops.12124

Diessel, H. (1999). The morphosyntax of demonstratives in synchrony and diachrony. *Linguist. Typol.* 3, 1–49. doi: 10.1515/lity.1999.3.1.1

Duranti, A. (2010). Husserl, intersubjectivity and anthropology. *Anthropol. Theory* 10, 1–20.

Edwards, T. (2014). *Language Emergence in the Seattle DeafBlind Community*. Ph. D. Thesis, The University of California, Berkeley, CA.

Edwards, T. (2015). Bridging the gap between DeafBlind minds: interactional and social foundations of intention attribution in the Seattle DeafBlind community. *Front. Psychol.* 6:1497. doi: 10.3389/fpsyg.2015.01497

Edwards, T. (2018). Re-channeling language: the mutual restructuring of language and infrastructure at Gallaudet University. *J. Linguist. Anthropol.* 28, 273–292. doi: 10.1111/jola.12199

Edwards, T., and Brentari, D. (2020). Feeling phonology: the emergence of tactile phonological patterns in protactile communities in the United States. *Language* 96, 1–22.

Evans, N., Bergqvist, H., and Lila, S. R. (2018a). The grammar of engagement: Part I, fundamentals. *Lang. Cogn.* 10, 110–140. doi: 10.10.1017/langcog.2017.21

Evans, N., Henrik, B., and Lila, S. R. (2018b). The grammar of engagement: Part II, typology. *Lang. Cogn.* 10, 141–170. doi: 10.1017/langcog.2017.22

Forker, D. (2020). Elevation as a grammatical and semantic category of demonstratives. *Front. Psychol.* 11:1712. doi: 10.3389/fpsyg.2020.01712

Goffman, E. (1981). *Footing: Forms of Talk*. Oxford: Basil Blackwell.

Goodwin, C., and Goodwin, M. H. (1987). Concurrent operations on talk: notes on the interactive organization of assessments. *IPRA Pap. Pragm.* 1, 1–54. doi: 10.1075/iprapip.1.1.01goo

Granda, A., and Nuccio, J. (2018). *Protactile Principles*. Monmouth, OR: Tactile Communications.

Guen, O. L. (2011). Modes of pointing to existing spaces and the use of frames of reference. *Gesture* 11, 271–307. doi: 10.1075/gest.11.3.02leg

Hanks, W. F. (1990). *Referential Practice: Language and Lived Space Among the Maya*. Chicago: University of Chicago Press, 616.

Hanks, W. F. (2009). Fieldwork on deixis. *J. Pragm.* 41, 10–24. doi: 10.1016/j.pragma.2008.09.003

Hanks, W. F. (2013). Counterparts: co-presence and ritual intersubjectivity. *Lang. Commun.* 33, 263–277. doi: 10.1016/j.langcom.2013.07.001

Haviland, J. (2009). "Pointing, gesture spaces, and mental maps," in *Language and Gesture*, ed. D. McNeil (Cambridge: Cambridge University Press), 13–46. doi: 10.1017/cbo9780511620850.003

Heine, B., Claudy, U., and Hünnemeyer, F. (1991). *From Cognition to Grammar: Evidence from African languages*. Amsterdam: Johns Benjamins.

Hochgesang, J., Crasborn, O., and Lilo-Martin, D. (2018). *ASL Signbank*. New Haven, CT: Haskins Lab.

Hopper, P. J., and Traugott, E. C. (2003 [1993]). *Grammaticalization*. Cambridge: Cambridge University Press.

Kendon, A. (1980). A description of a deaf-mute sign language from the Enga Province of Papua New Guineaa with some comparative discussions: Parts I-II. *Semiotica* 32, 1–34. doi: 10.1515/semi.1980.31.1-2.1

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kocab, A., Pyers, J., and Senghas, A. (2015). Referential shift in nicaraguan sign language: a transition from lexical to spatial devices. *Front. Psychol.* 5:1540. doi: 10.3389/fpsyg.2014.01540

Kooij, E. V. D. (2002). *Phonological Categoriess in Sign Language of the Netherlands*. Utrecht: LOT Publications.

Lerner, G. (1989). Notes on overlap management in conversation: the case of delayed completion. *West. J. Speech Commun.* 53, 167–177. doi: 10.1080/10570318909374298

Lerner, G. (1993). Collectivities in action: establishing the relevance of conjoined participation in conversation. *Text* 13, 213–245.

Lerner, G. (1996). "On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker," in *Interaction and Grammar*, eds E. A. S. E. Ochs and S. A. Thompson (Cambridge: Cambridge University Press), 238–276. doi: 10.1017/cbo9780511620874.005

Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.

Mandel, M. (1981). *Phonotactics and Morphophonology in American Sign Language*. Doctoral Dissertation, University of California, Berkeley, CA.

McMillen, S. K. (2015). *Is Protactile Habitable at Gallaudet University: What does it take?*. Washington, DC: Gallaudet University.

McNeil, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

Meir, I. (2003). Grammaticalization and modality: the emergence of a case-marked pronoun in Israeli sign language. *J. Linguist.* 39, 109–140. doi: 10.1017/s0022226702001664

Mesch, J. (2001). *Tactile Sign Language: Turn Taking and Questions in Signed Conversations of Deaf-blind People*. Hamburg: Signum.

Mesh, K. (2017). *Points of Comparison: What Indicating Gestures tell Us About the Origins of Signs in San Juan Quiahije Chatino Sign Language*. Ph. D. Thesis, The University of Texas, Austin, TX.

Orie, O. O. (2009). Pointing the Yoruba way. *Gesture* 9, 237–261. doi: 10.1075/gest.9.2.04ori

Pfau, R. (2011). "A point well taken: on the typology and diachrony of pointing," in *Deaf Around the World*, eds D. J. Napoli and G. Mathur (Oxford: Oxford University Press).

Reed, C. M., Delhorne, L. A., Durlach, N. I., and Fischer, S. D. (1995). A study of the tactual reception of Sign Language. *J. Speech Hear. Res.* 38, 477–489. doi: 10.1044/jshr.3802.477

Rommetveit, R. (1976). "On the architecture of intersubjectivity," in *Social Psychology in Transition*, eds L. H. Strickland, K. J. Gergen, and F. J. Aboud (New York, NY: Plenum Press).

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/s0047404500001019

Sidnell, J. (2014). "The architecture of intersubjectivity revisited," in *The Cambridge Handbook of Linguistic Anthropology*, eds N. Enfield, P. Kockelman, and J. Sidnell (Cambridge: Cambridge University Press).

Supalla, T. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. Doctoral Dissertation, University of California, San Diego, CL.

Taub, S. (2001). *Language from the Body: Metaphor and Iconicity in American Sign Language*. Cambridge: Cambridge University Press.

Willoughby, L., Iwasaki, S., Bartlett, M., and Manns, H. (2018). "Tactile sign languages," in *Handbook of Pragmatics*, eds J.-O. Östman and J. Verschueren (North York, ON: Benjamins), 239–258. doi: 10.1075/hop.21.tac1

Zwisterlood, I. (2012). "Classifiers," in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 158–185.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fcomm.2021.567774

How Pointing is Integrated into Language: Evidence From Speakers and Signers

Kensy Cooperrider[1]*, *Jordan Fenlon*[2], *Jonathan Keane*[2], *Diane Brentari*[2] and *Susan Goldin-Meadow*[1]

[1]Department of Psychology, University of Chicago, Chicago, IL, United States, [2]Department of Linguistics, University of Chicago, Chicago, IL, United States

When people speak or sign, they not only describe using words but also depict and indicate. How are these different methods of communication integrated? Here, we focus on pointing and, in particular, on commonalities and differences in how pointing is integrated into language by speakers and signers. One aspect of this integration is *semantic*—how pointing is integrated with the meaning conveyed by the surrounding language. Another aspect is *structural*—how pointing as a manual signal is integrated with other signals, vocal in speech, or manual in sign. We investigated both of these aspects of integration in a novel pointing elicitation task. Participants viewed brief live-action scenarios and then responded to questions about the locations and objects involved. The questions were designed to elicit utterances in which pointing would serve different semantic functions, sometimes bearing the full load of reference ('load-bearing points') and other times sharing this load with lexical resources ('load-sharing points'). The elicited utterances also provided an opportunity to investigate issues of structural integration. We found that, in both speakers and signers, pointing was produced with greater arm extension when it was load bearing, reflecting a common principle of semantic integration. However, the duration of the points patterned differently in the two groups. Speakers' points tended to span across words (or even bridge over adjacent utterances), whereas signers' points tended to slot in between lexical signs. Speakers and signers thus integrate pointing into language according to common principles, but in a way that reflects the differing structural constraints of their language. These results shed light on how language users integrate gradient, less conventionalized elements with those elements that have been the traditional focus of linguistic inquiry.

Keywords: pointing, gesture, sign language, deixis, demonstratives, language

OPEN ACCESS

Edited by:
Olga Capirci,
Italian National Research Council, Italy

Reviewed by:
Brendan Costello,
Basque Center on Cognition, Brain and Language, Spain
Olivier Le Guen,
Center for Research and Higher Studies in Social Anthropology, Mexico
N. J. Enfield,
The University of Sydney, Australia

***Correspondence:**
Kensy Cooperrider
kensycoop@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *30 May 2020*
Accepted: *29 January 2021*
Published: *03 May 2021*

Citation:
Cooperrider K, Fenlon J, Keane J, Brentari D and Goldin-Meadow S (2021) How Pointing is Integrated into Language: Evidence From Speakers and Signers. Front. Commun. 6:567774. doi: 10.3389/fcomm.2021.567774

INTRODUCTION

When people communicate—whether by speaking or signing—they interweave at least three different methods (Clark, 1996; Enfield, 2009; Clark, 2016; Ferrara and Hodge, 2018). First, they use the categorical, highly conventionalized symbols—words—used within their language community. That is, they *describe*. These communicative resources are the traditional focus of linguistic inquiry. But people also use resources that are gradient and less conventionalized: they iconically represent images, actions, and sounds—they *depict*—and they draw attention to locations,

objects, and people—they *indicate*. This three-method framework—which builds on the semiotic theories of Charles Peirce (Peirce, 1940)—presents a powerful lens through which to understand the complexity and heterogeneity of human communication. But it also prompts a key question: How are these disparate methods of communication integrated? How do resources like depictions, points, and words come together into a coherent and fluent stream of meaning? And how do the mechanisms of integration differ in signed vs. spoken communication?

To shed light on these issues, we focus here on the case of pointing. Pointing—in which one person directs another's attention to a target location—is a ubiquitous form of indicating (Clark, 2003; Kendon, 2004; Cooperrider et al., 2018). It has been described as "a basic building block" (Kita, 2003a) of human communication: it is universal across cultures (e.g., Eibl-Eibesfeldt, 1989); is among children's first communicative acts (e.g., Bates, 1979; Liszkowski, 2006; Capirci and Volterra, 2008; Liszkowski et al., 2012); and is pervasive and multifunctional in both spoken (Clark, 2003) and signed (e.g., McBurney, 2009; Johnston, 2013) interaction. For these and other reasons, pointing has been widely examined across the cognitive and linguistic sciences (e.g., contributions in Kita, 2003b), and has been a focus of recent efforts to compare gesture and sign (e.g., Cormier et al., 2013; Johnston, 2013; Meier and Lillo-Martin, 2013; Goldin-Meadow and Brentari, 2017; Fenlon et al., 2018; Fenlon et al., 2019). Yet little work to date has closely examined how pointing is integrated with words (for some exceptions, see: Bangerter, 2004; Cartmill et al., 2014; Cooperrider, 2016; Floyd, 2016). The question of how pointing is integrated with the more highly conventionalized components of language is multilayered, and we focus on two important aspects of it.

First, how is pointing integrated with the intended meaning of the broader utterance in which it is embedded? We term this *semantic integration*, and several observers have noted that a point's form reflects this type of integration. Kendon (2004) noted that speakers point differently depending on how the target of the point "is presented in the speaker's discourse" (p. 201). Enfield et al. (2007) examined the size of a pointing gesture in relation to its function. They found that speakers of Lao were more likely to produce big points—that is, points with a greater degree of arm extension—when those points convey "a primary, foregrounded part of the message" (p. 1723). (They also suggested that, beyond involving greater arm extension, such points tended to be longer in duration, but they did not analyze this systematically [p. 1728]). The researchers described these contexts as "location-focus" utterances. The central example, they observed, occurs when someone answers a question about where something is; speakers will often respond to a "where" or "which" question by producing a point along with a demonstrative such as "here" or "there." Thus, they find that the form of a pointing gesture is integrated with utterance-level meaning.

Building on the proposals in Enfield et al. (2007), we suggest that a point's form might also be integrated with meaning at a finer level: the lexical semantics of the phrase in which the point is embedded. Within location-focus utterances, pointing can carry more or less of the burden of specifying location. For example, the question "Where did you park?" could be answered simply with a wordless point, or with a point along with a demonstrative, such as "there." In such cases, the point bears the full load of specifying location; we call these instances "load-bearing points." The same question could also be answered, however, with a point plus a longer description of location, such as "Over on the left, in the far back." In this example, the point shares the load of specifying location with lexical material; we call such cases "load-sharing points." To date, beyond the work of Enfield et al. (2007) on Lao co-speech gesture, we are not aware of any studies that have closely examined either of these types of semantic integration, utterance-level (location-focus or not) or lexical-level (load-bearing or load-sharing). A major goal of the present study is to determine whether pointing signs (points that signers produce) reflect semantic integration in the same way as pointing gestures (points that speakers produce).

Second, how is pointing as a manual signal structurally integrated with other signals involved in the utterance? We term this *structural integration*. Our focus is manual pointing, although non-manual forms are prominent in some communities (see, e.g., Cooperrider et al., 2018). This aspect of integration has the potential to differ for speakers and signers because spoken and signed communication are put together differently. Spoken communication involves different articulatory channels operating in parallel, with speech produced in tandem with gestures of the hands and movements of the face and body. Importantly, in spoken communication, much of the describing—that is, the use of highly conventionalized symbols—is done with the mouth, and much of the depicting and indicating is done with the hands. Signed communication similarly involves different articulatory channels working in parallel: manual signs are produced in tandem with movements of the face and body. But, in contrast to spoken communication, the bulk of the describing, depicting, and indicating all occur in the hands (though enriched with critical information in the face and body).

These differences have clear consequences for the structural integration of pointing: manual pointing gestures must integrate with spoken words, which are produced in a different articulatory channel, whereas manual pointing signs must integrate with signs within the same articulatory channel. Fenlon et al. (2019) described this situation as a "same-channel constraint" and suggested that the constraint may account for certain differences between points that accompany sign language and points that accompany spoken language. For instance, in their comparison of pointing signs and pointing gestures in two conversational corpora, Fenlon et al. (2019) found that pointing signs were much shorter in duration overall than pointing gestures, which may have been related to the pressure to slot those points in to a stream of manual signals. Speakers, of course, do not face the same-channel constraint; their pointing gestures are therefore free to take more time, spanning across much of an utterance or even bridging over adjacent utterances. Fenlon et al.'s (2019) account of these tendencies, however, needs to be further investigated. After all, signers do sometimes point with their non-dominant hand, thus allowing those points to span

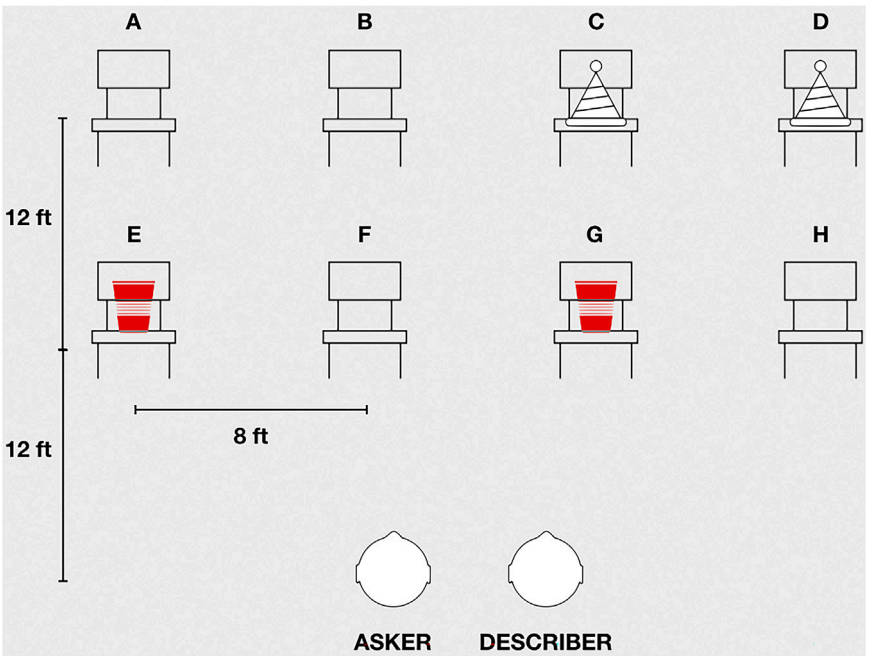


FIGURE 1 | A schematic depiction of the task layout (not to scale). Participants sat before an array of chairs (labeled **(A–H)** in a large performance space. After watching an actor carry out actions involving two pairs of objects (e.g., hats, cups), the describer answered a series of questions posed by the asker.

across signs produced with their dominant hand (Johnston, 2013). Moreover, speakers occasionally slot their gestures in to breaks in speaking, a phenomenon occasionally called "component gestures" (Clark, 1996). Thus, a second goal of the present work is to examine these aspects of structural integration more closely for both pointing gestures and pointing signs.

In the present study, we sought to investigate both semantic and structural aspects of integration in speakers and signers. To this end, we developed a novel pointing elicitation task. English speakers and ASL signers watched brief live-action vignettes in which an actor interacted with objects at different locations in an eight-location grid. Participants were then asked questions about the locations and objects in the vignettes. Some questions were designed to elicit utterances in which location would be focal (e.g., an answer specifying where a particular object was located), whereas other questions were designed to elicit utterances in which location would be mentioned but not focal (e.g., what action was performed at a given location). The elicited utterances also provide a data set within which to examine how points were integrated with other signals, whether manual or vocal. We thus extend Enfield et al. (2007) by examining the semantic integration of pointing in another spoken language (English) and, for the first time, in a sign language (ASL). We also extend Fenlon et al. (2019) by focusing on points to objects and location rather than persons; by using a controlled task rather than conversational data; and by looking at the consequences of the "same-channel" constraint in further detail. The overarching goal is to shed light on how speakers and signers integrate disparate forms of communication into coherent streams of discourse.

METHODS

Participants

24 adults (mean age = 30.3, 8 women) from the Chicago area participated in exchange for payment. One group, referred to as 'speakers' (n = 12; 5 women), consisted of hearing participants who were native or near-native speakers of English. None of the speakers reported knowing ASL. The other group, referred to as 'signers', consisted of deaf participants (n = 12, 3 women) who were native or near-native signers of American Sign Language (ASL). The study was conducted in pairs of participants from the same group.

Materials and Procedure

The study was conducted in a large performance space. The participants sat in two chairs near the stage, facing out toward where the audience would typically be seated. In the middle of the room were two rows of four chairs, one row approximately 12 feet from participants' seats and the other approximately 12 feet beyond that. Two rows at different distances offered the possibility of analyzing whether target distance affects pointing form. Within each row, the chairs were spaced approximately eight feet apart (**Figure 1**).

For each trial, one participant served as the *describer* and the other as *asker*. Before the beginning of a trial, the asker left the room; the describer stayed seated and watched as an experimenter acted out a scripted scenario. On the chairs were four objects of two types (e.g., two identical cups and two identical party hats). Each scenario would involve the experimenter carrying out a series of four actions, which involved the four objects at different

chairs. For example, in one scenario (see **Figure 1**), the four objects were located on four different chairs: two identical red cups on chairs E and G, and two identical white party hats on chairs C and D. At the start of the scenario—signaled by the experimenter raising her hand—the experimenter moved to chair E (far left chair, front row) and retrieved one of the cups (action 1); she then moved to chair B (left middle, back row) and took a sip from the cup (action 2); she then walked to chair H (far right, front row) and placed the cup on the chair (action 3); finally, she walked to chair D (far right, back row) and put on the party hat (action 4). Finally, the experimenter signaled the end of the scenario by raising her hand.

At the end of each scenario, the asker re-entered the room and was given a single piece of paper with five questions to ask the describer. The first four questions asked for "which"/"where" information. These questions were designed to elicit utterances focused on a single, specific location (location-focus utterances). For example, the question "Which cup was used?" might elicit a location-focus utterance such as "That one." The final question, identical for all scenarios, was: "Could you please explain everything that happened, in the order that it happened?" This prompt was designed to elicit utterances that may have mentioned locations, but were focused on explaining the actions and the sequence in which they took place ("explanatory utterances"). For example, in response to this question a speaker might begin a longer explanation with the utterance, "So she picked up the cup from that chair."

Each pair completed two practice scenarios—one per participant—followed by eight primary scenarios. The participants switched roles after every scenario, with each participant completing four scenarios as describer and four as asker. The chairs remained in the same position across all scenarios; the four objects varied, as did the relevant locations and the actions performed at each location. All locations were used across the scenarios but not equally. Full information about the scenarios, as well as a video of the experimenter performing them, is available in the online Supplementary Material: https://osf.io/wckx5/). Finally, pointing was not mentioned in the instructions and all sessions were video-recorded.

Analysis

The videos were segmented into scenarios and analyzed using ELAN annotation software (Lausberg and Sloetjes, 2009). Only the describer's behaviors are discussed here. The primary coder was a hearing signer with experience analyzing bodily communication; this coder performed all the pointing and language analyses, with another experienced coder performing reliability as described below.

Pointing Gesture Analysis

Following prior work (e.g., Fenlon et al., 2019), points were defined as movements toward a region of space that were intended to direct attention to that region. Points produced by the speakers (pointing gestures) and points produced by the signers (pointing signs) that were directed to any of the eight target locations—i.e., the chairs—were identified for further analysis (see 'Pointing analysis' section of the coding manual in the online **Supplementary Material**). We only considered those points that exclusively conveyed location, sometimes called "pure points" (Kendon, 2004). Pointing gestures that simultaneously conveyed location and other information (e.g., gestures directed to a location but also conveying the action of picking up an object or moving points showing the trajectory the experimenter took between locations) were excluded, as were points judged to be non-communicative (e.g., produced as part of a private rehearsal of the scenario).

Points were coded for two features: extension and duration. These features were selected, following prior studies, because they seem to vary according to a point's communicative importance and function (Enfield et al., 2007; Peeters et al., 2015). For extension, we distinguished points with full extension—an arm that was straight across the elbow—from points with partial extension—an arm that had a bend at the elbow. (We initially tried to distinguish a third category of "minimal extension" points, which had a bend at the elbow and no raising of the upper arm, but were unable to achieve high reliability for this distinction.) For duration, we coded the onset of the point—defined as the first frame of motion from rest or from a prior gesture or sign—and the offset—defined as the last frame of the hold phase (or frame of fullest extension, if no hold), before returning to rest or beginning another gesture/sign. Duration was then determined based on the time between pointing onset and offset.

Reliability was assessed by having a second coder (another hearing signer with experience analyzing bodily communication) analyze one randomly selected scenario from each participant (of the four completed; i.e., 25% of the data). Agreement on the presence of pointing was 80%. Discrepancies resulted from: one coder identifying a point that the other did not; one coder identifying a single point where the other identified two or more points; and one coder considering a point to be iconic or non-communicative, whereas the other considered it a "pure point." For those points identified by both coders as pure points (N = 238; other points were not coded for extension), they agreed on whether extension was partial or full 93% of the time (Cohen's $K = 0.87$). Further, the durations attributed to points by the two coders were highly correlated with each other ($r = 0.94$).

Language Analysis

For each point, we identified the utterance in which it occurred. Utterances were segmented using a combination of grammatical criteria (i.e., clause boundaries) and visible/audible criteria (e.g., pauses) (Du Bois et al., 1993; Sandler et al., 2005; Fenlon et al., 2007; Brentari et al., 2011) (for full details, see 'Language analysis' section of the coding manual). These utterances thus closely aligned with the notion of an "intonational phrase" (Nespor and Vogel, 1986), but we use "utterance" for simplicity. After each utterance containing a point was demarcated, its contents were transcribed; transcriptions were crosschecked by the second coder. As part of this transcription process, the primary coder marked, within the string of words or signs, the hold onset of each point to the nearest word boundary (with an open bracket) and the hold offset of each point to the nearest word boundary (with a

TABLE 1 | Examples of utterance types.

Utterance type	Examples from speakers	Examples from signers
Location-focus		
Load-bearing	[that] chair	HAT [___]
	right [there]	[___]
	[___] that one	BAG [___]
Load-sharing	the [back] left chair	BACK ROW [___] THIRD [___]
	that [far corner] brush	LEFT [___]
	[front] far right	BACK [___] ALL-THE-WAY [___]
Explanatory		
	then she moved back one [chair]	DRINK [___] FINISH
	she picked up the cup from [that] chair	WALK [FRONT SECOND]
	and she put the ball in [the] chair	PUT-DOWN-CUP [___]

Note: *Points not produced along with words or signs are marked as [___]. For points that were produced along with words or signs, we indicate the onset and offset of the hold with brackets around the overlapping lexical material.*

closed bracket). This step allowed for subsequent analysis of different aspects of structural integration.

These transcribed utterances were then coded for their semantic function at two levels of granularity (**Table 1**). At one level, we coded whether the utterances were "location-focus" or "explanatory." "Location-focus" utterances served to specify *which* or *where* information, and were most often sentence fragments with no verb (e.g., "In the back row"). "Explanatory" utterances were defined as those that included a verb describing the actions of the experimenter (e.g., "Then walked to that chair"). At a second, more fine-grained level, we further divided the location-focus utterances based on the semantic role of the point vis-à-vis the other lexical material in the same utterance[1]. If the point specified location entirely on its own, it was considered a "load-bearing point"; if the utterance in which the point occurred contained any other spoken words or lexical signs that helped specify location (e.g., *back*, *front*, *left*, *right*), the point was considered a "load-sharing point." We also coded whether speakers' points were co-produced with a demonstrative—*this*, *that*, *here*, or *there* (e.g., Diessel, 2006). (The same analysis was not carried out for signers because the lexical item THAT was used rarely—by only three of the signers—and primarily served an anaphoric function.)

Points were also coded for whether, and how, they were structurally integrated with the other lexical material. We first noted whether the utterances that included a point also included lexical material. Points within utterances that did contain other material were further analyzed as follows. Points were considered to slot in if they were not produced concurrently with other words (e.g., "[___] that one") or signs (e.g., "PUT BOWL [___]"). Points were considered to span across if they were produced concurrently with words (e.g., "that [far corner] brush") or signs (if produced with one hand while the other hand produced additional signs; e.g., "WALK [FRONT SECOND]") (as indicated by square brackets/underlining containing words or glosses); and they were considered to bridge over if they were held across adjacent utterances (speaker example: "So, walked over to this [chair /picked up the ball]"; signer example: "WALK [___] FRONT THIRD [/ PUT DOWN HAT]"). On the basis of the transcriptions, utterances were further coded for the number of words or signs, in addition to the point, that they contained.

RESULTS

General

Pointing—though not mentioned in the instructions or modeled in any way—was a prominent part of how both speakers and signers carried out the task (see examples, **Figure 2**). Speakers produced a total of 513 points (mean per participant = 42.8) (again, including only pure points to target locations and excluding points with iconic aspects, as described earlier). At least one point was produced in 93% of responses to the "which"/"where" questions and 98% of responses to the explanation prompts. Signers produced a total of 508 points (mean per participant = 42.2), and at least one point occurred in 90% of responses to the "which"/"where" questions and 90% of responses to the explanation prompts. Despite the fact that pointing is usually considered optional in spoken communication but

[1]A similar, more fine-grained analysis of explanatory utterances was not possible because this class was more heterogeneous than the location-focus utterances. Some of the speakers' points in explanatory utterances, for instance, did not occur with location references in speech.

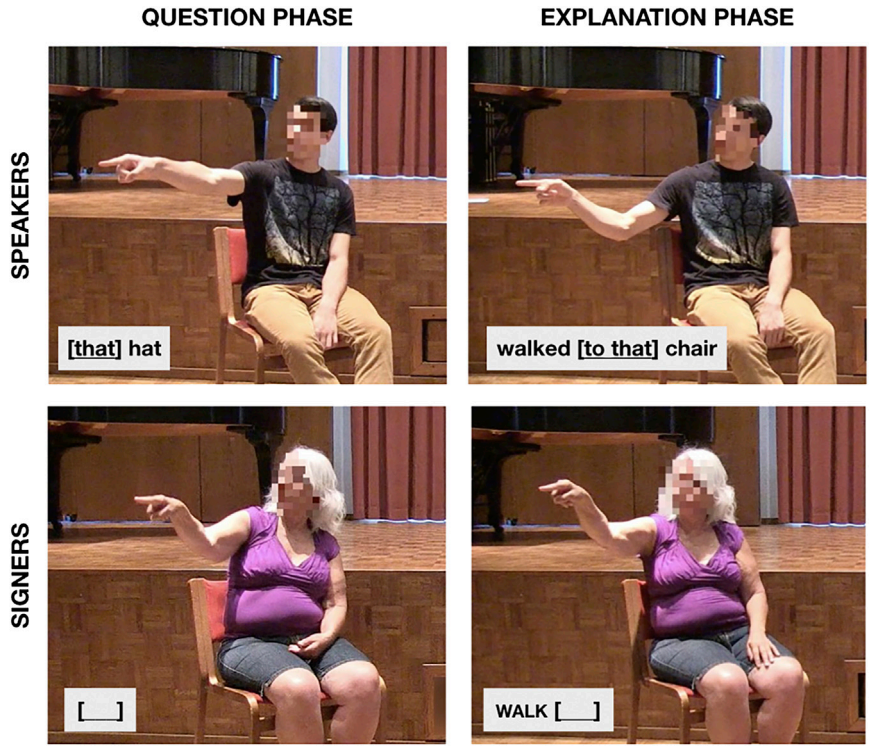


FIGURE 2 | Examples of points produced by a speaker **(top row)** and signer **(bottom row)**. Points in the left column were produced in response to a "which"/"where" question; points in the right column were produced in response to a prompt to explain everything that happened in the scenario.

integral to the language in signed communication, the amount of pointing was thus comparable in the two groups.

We next confirmed that the two different question types—"which"/"where" or explanation prompt—successfully elicited points embedded in different types of utterances. Of the points produced in response to the "which"/"where" questions, most were part of location-focus utterances (speakers = 90%; signers = 86%). Conversely, of the points produced in response to the explanation prompts, most were part of explanatory utterances (gesturers = 82%; signers = 73%). These percentages confirm that the elicitation procedure worked as expected. In what follows, rather than use question type (i.e., "which"/"where" or explanation prompt) in our analyses, we use semantic function—load-bearing location, load-sharing location, and explanatory—as the key predicting variable.

All three semantic functions were common in both groups (Note that semantic integration analyses excluded the 68 points—64 in speakers, 4 in signers—that bridged over utterances, as these were difficult to assign to a single semantic function.) Speakers produced 108 load-bearing location points (24% of all speakers' points), 146 load-sharing points (33%), and 195 explanatory points (43%). Signers produced 130 load-bearing location points (26% of signers' points), 194 load-sharing location points (38%), and 184 explanatory points (36%). 17 participants produced points associated with all three semantic functions; the other seven participants (four gesturers and three signers) produced points with only two of the semantic functions. For a breakdown of the number of observations of each type contributed by each participant, see **Supplementary Table S1**.

Results Bearing on Semantic Integration

We next analyzed the points for whether their form reflected their semantic integration with the surrounding language, focusing on extension and duration. For all analyses, we built hierarchical regression models in R (R Core Team, 2016) using the lme4 package (Bates et al., 2015). We used the maximal models that were both justified by theoretical concerns (e.g., including group and semantic function, and their interaction) and were able to converge. Unless otherwise specified, values in all figures and the text are estimates from these models rather than raw values. For ease of interpretability, we present these estimates in their natural interpretation space (i.e., probabilities or untransformed milliseconds) along with 95% confidence intervals (CIs), rather than coefficient estimates and standard errors (available in the **Supplementary Material**, along with t or z values).

To judge the direction and magnitude of effects, we use CIs, rather than p values (e.g., Gelman and Tuerlinckx, 2000; Cumming and Finch, 2005). An advantage of CIs is that they provide an intuitive visual format for judging the precision of point estimates and the size of effects, and thus discourage the "dichotomous thinking" (Cumming, 2014, p. 8) associated with p values. We judge whether two model estimates are different in a gradient manner. If the CIs of the estimates do not overlap at all, this indicates a robust difference; if one estimate falls outside of

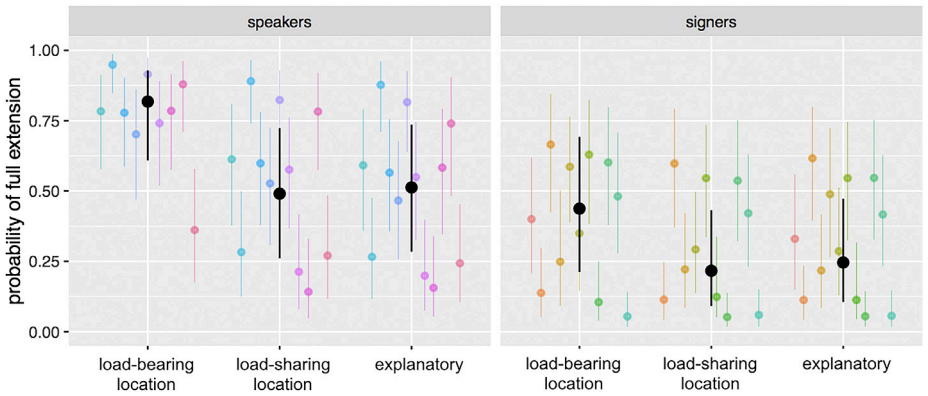


FIGURE 3 | Results from the model predicting the probability of using full arm extension, as a function of participant group and semantic function. Speakers were much more likely to use full extension overall, but in both groups load-bearing location points were more likely to involve full arm extension than other points. Black dots represent group means, and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

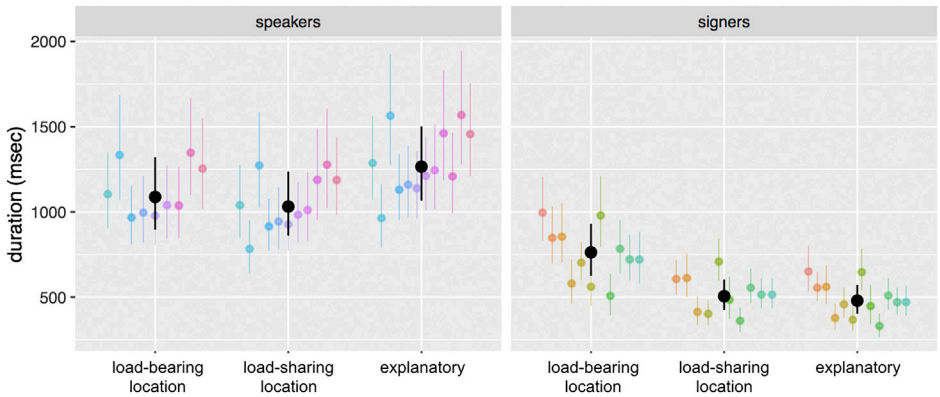


FIGURE 4 | Results from the model predicting the duration of a point as a function of participant group and semantic function. Speakers' points were longer in duration overall than signers' points, and were longer when part of explanatory utterances. In contrast, signers' load-bearing location points were longer than either of the other two types. Black dots represent group means, and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

the CIs of the other (but the bounds of the CIs overlap), this suggests weaker evidence for a difference; if one estimate falls squarely inside the CIs of the other, this suggests a lack of strong evidence for a difference. We use these as guidelines rather than rigid cut-offs; they are not meant to substitute for traditional significance testing. For discussion of similar guidelines and their relation to significance testing, see Cumming and Finch (2005). Finally, in all figures, we also estimate confidence intervals for each participant using bootstrap sampling.

Extension. To analyze extension, we first built a hierarchical logistic regression model with group (speakers or signers) and semantic function (load-bearing location, load-sharing location, explanatory) as predictors, and with random effects for participant and target location (intercept only). In this and other models, these random effects were chosen because they were expected to exhibit variability that was orthogonal to our measures of interest—i.e., individuals may vary in pointing style and different locations may prompt points with different degrees of arm extension. Overall, speakers' points were much more likely to involve full extension than signers' points, with speakers using full extension 57% of the time and signers only 36% of the time (raw percentages). However, extension in both groups was shaped by how the point fit semantically with the surrounding speech or signing, with participants in both groups being more likely to use full extension for load-bearing location points (speakers: 82% [95% CI: 61–93%]; signers: 44% [95% CI: 21–69%]), compared to either load-sharing location points (speakers: 49% [95% CI: 26–72%]; signers: 22% [95% CI: 9–43.2%) or explanatory points (speakers: 51% [95% CI: 26–74]%; signers: 25% [95% CI: 11–47%]) (**Figure 3**). This pattern of data—though somewhat more robust in the speakers than in the signers—suggests that whether the utterance as a whole is location-focus or explanatory is not the important factor for predicting extension in either group. Rather, what matters is how

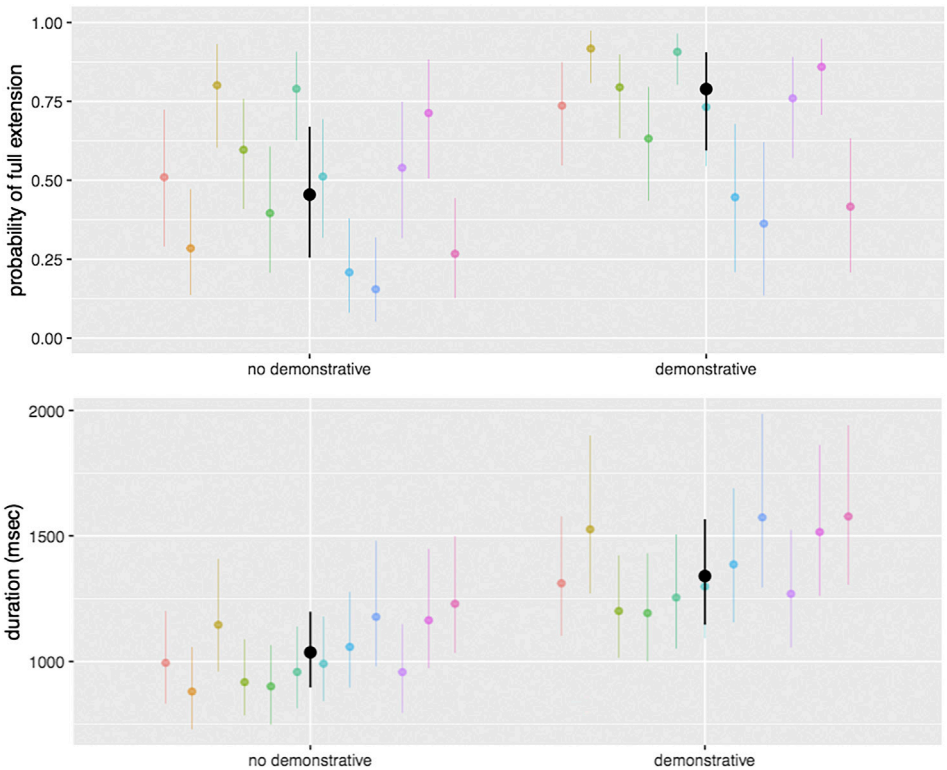


FIGURE 5 | Results from a model of speakers only, predicting the probability of full arm extension **(top)** and duration **(bottom)** as a function of whether or not a point is associated with a spoken demonstrative (*this, that, here, there*). Points that were associated with a demonstrative were more likely to involve full arm extension, and were longer in duration, than points that were not associated with a demonstrative. Black dots represent overall means and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

the point relates to the lexical semantics within that utterance. We also built another version of the model with number of words in the utterance added as a predictor, but this addition did not substantially change the observed effects; on its own, number of words does not have an effect on extension. Lastly, we built a version of the model with location added as a predictor. Some target locations (i.e., B, C, and H, all far from the speaker) were associated with full-extension points more than others, but the effects of semantic function remain.

Duration. Next, we built an analogous model to analyze duration. Overall, speakers' points were much longer in duration (in milliseconds) than signers' points (speakers: M = 1327; signers: M = 724) (raw group means). Moreover, duration patterned differently across the different semantic functions in the two groups (**Figure 4**). In speakers, points embedded in explanatory utterances (explanatory = 1266 msec [95% CI: 1066–1502 msec]) were longer than those embedded in either type of location-focus utterance (load-bearing = 1088 msec [95% CI: 896–1321 msec]; load-sharing= 1032 msec [95% CI: 861–1237 msec]). In signers, the pattern was reversed; mirroring the results for extension, signers' points were longer when embedded in load-bearing location utterances (load-bearing location = 763 msec [95% CI: 625–930 msec]) than in either of the other two types of utterances (load-sharing location = 506 msec [95% CI: 424–603 msec]; explanatory = 479 msec [95% CI: 402–572 msec]). As in our extension analyses, we built a model in which we added number of words in the utterance as a predictor. Adding this factor does not substantially change the observed effects. Moreover, there was no strong relationship between duration and number of words on its own, though in speakers there is a weak trend toward longer point durations as the number of words in the utterance increases; in signers there is no evidence of such a trend.

Demonstratives. Finally, we also conducted a separate analysis for speakers, asking whether points that were associated with spoken demonstratives (N = 175, or 39% of all points) differed from points that were not. (Again, we did not attempt to analyze this in signers; although ASL signers do make occasional use of a lexical item glossed as THAT, only three signers in our study used this sign and the predominant function appeared to be anaphoric rather than to support a co-occurring pointing sign.) We built analogous models predicting extension or duration, but adding in demonstrative presence as a predictor (and removing function and participant group, as we were only looking at speakers). We found that pointing gestures that were associated with demonstratives were more likely to be fully extended (demonstrative present = 79% [95% CI: 59–91%]) than pointing gestures that were not associated with demonstratives (demonstrative absent = 45% [95% CI: 26–67%]). Points that were associated with demonstratives were also longer in duration

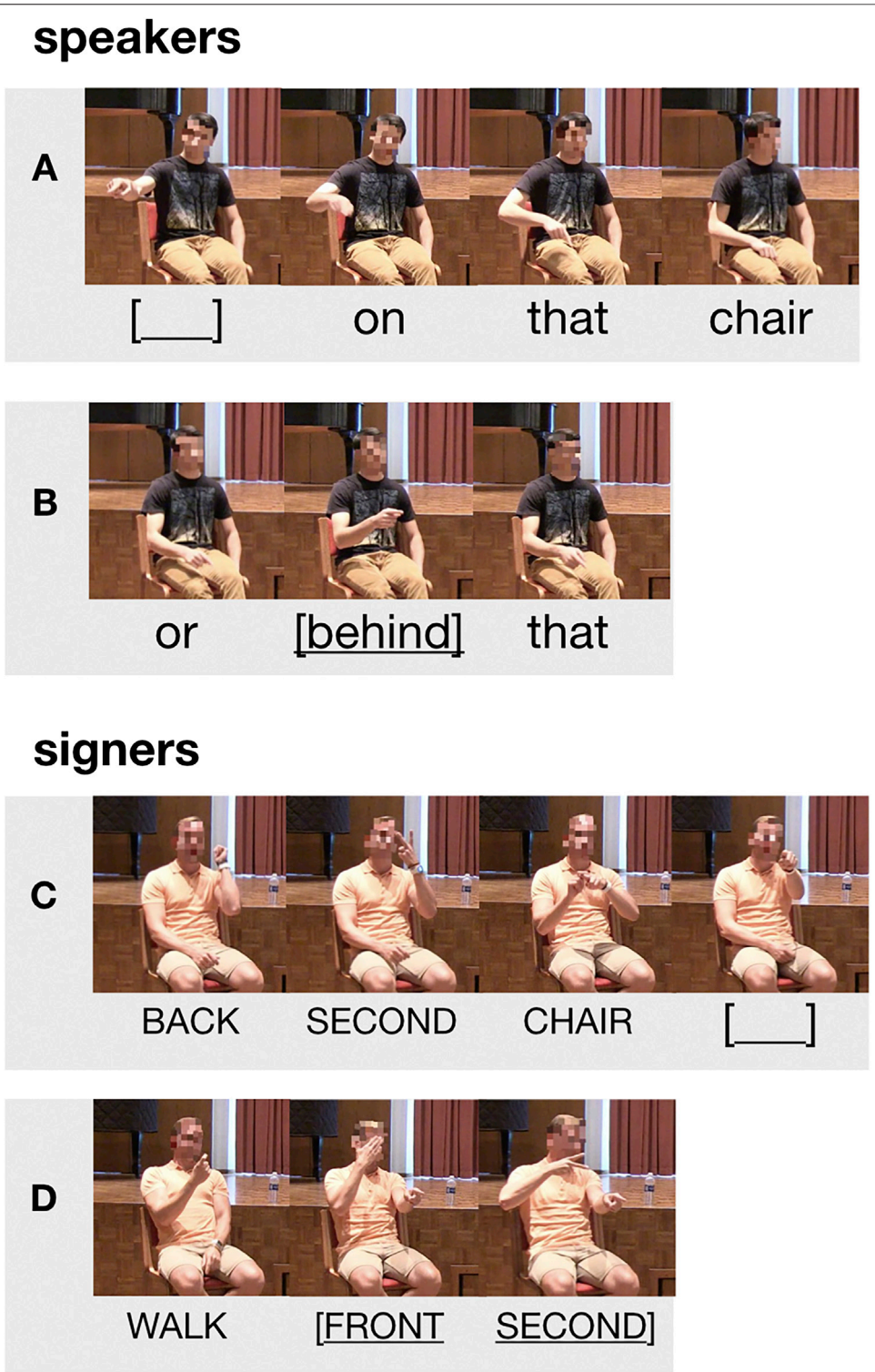


FIGURE 6 | Examples of different structural integration possibilities. Each example consists of an utterance containing a point and three words; each frame shows the participant during production of the word (or standalone point) transcribed below. In some cases, the points slotted in to the string of words, a pattern observed rarely in speakers **(A)** but very commonly in signers **(C)**. In other cases, the points spanned across words in the utterance, a pattern observed very commonly in speakers **(B)** but rarely in signers **(D)**.

TABLE 2 | Structural integration.

Type of point	Speakers	Signers
	count (raw mean duration in msec)	
Stand-alone points	**20** (717)	**110** (963)
Points contained within an utterance	**429** (1356)	**394** (657)
Spanning across words	399 (1381)	26 (1125)
Slotting in to words	30 (1020)	368 (624)
Points bridging over adjacent utterances	**64** (2801)	**4** (2736)
Total	**513**	**508**

(demonstrative present = 1340 msec [95% CI: 1146–1566 msec]) than points that were not associated with demonstratives (demonstrative absent = 1036 msec [95% CI: 896–1198 msec]) (**Figure 5**). We built additional models adding in semantic function as a predictor, which attenuates the effect of demonstrative presence on extension but not on duration.

Results Bearing on Structural Integration

We turn now to issues of structural integration. We first analyzed whether points were stand-alone in that they were the only signal within an utterance; whether they were contained within an utterance that included other lexical material; or whether they bridged over adjacent utterances. In both speakers and signers, the majority of points were produced within an utterance that also contained other lexical material (speakers = 84%; signers = 78%). Signers produced a greater percentage of stand-alone points than speakers (speakers = 4%; signers = 22%), and speakers produced a greater percentage of bridging over points than signers (speakers = 12%; signers = 1%) (**Table 2**).

Zooming in on those points produced as part of an utterance that contained other lexical material, we next considered how points are integrated with that material. In particular, we were interested in whether points "slotted in" (that is, fit into gaps in the string of words or signs) or "spanned across" (that is, overlapped with other words or signs in the utterance) (**Figure 6**). Although speakers' points did occasionally slot in to the lexical string (7%), the overwhelming majority of speakers' points spanned across part (or all) of that string (93%). Signers showed the opposite pattern, with their points slotting in to other signs (93%) more often than spanning across other signs (7%).

In sum, all the structural integration categories we considered were deployed by both groups. Yet each group nonetheless exhibited a clear characteristic profile, likely because of the same-channel constraint that exerts pressure on signers but is absent in speakers. Moreover, informal inspection of the durations of points associated with these different categories reveals other patterns (**Table 2**). In both groups, points that span across lexical material are numerically longer than points that slot in; and, similarly, points that bridge over utterances are substantially longer than points that are contained within an utterance (though there are few examples of bridging over points in signers). Thus overall differences in pointing duration between speakers' and signers' points may be due largely to the contrasting structural integration profiles that each group exhibits.

DISCUSSION

Human communication—whether spoken or signed—involves the fluent integration of pointing into language. Here, we used a novel elicitation task to better understand how this integration is achieved, at both semantic and structural levels. To investigate semantic integration, we used different prompts expected to elicit utterances in which pointing served different semantic roles vis-à-vis linguistic elements. At the broadest level, these points sometimes supported utterances in which location was focal and other times supported utterances that were more explanatory; at the finer lexical level, these points sometimes bore the full load of communicating location and sometimes shared that load with words. We found that both speakers and signers were more likely to produce points with full arm extension when the point bore the full load. (The duration of the points proved more nuanced, a finding we discuss in more detail below.) To investigate structural integration, we used the resulting utterances from this elicitation task to better understand how pointing coordinates with surrounding words/signs and, in particular, how it slots in to, or spans across, other spoken or manual signals. We found that, while speakers and signers both integrate pointing with other signals in diverse ways, they also show strong characteristic profiles, with pointing gestures usually spanning across words—and sometimes bridging across utterances—and pointing signs usually slotting in between signs—and rarely spanning across them or bridging across utterances. These differences, we argue, are best understood in light of the differing modality constraints faced by speakers and signers. Put together, our results show how speakers and signers integrate pointing into language in response to common pressures, while at the same time navigating constraints that are particular to spoken or signed communication.

Our findings about semantic integration conceptually replicate and build on earlier observations. We extended Enfield et al.'s (2007) core finding to another spoken language and, for the first time, to a signed language. However, it is important to note key differences between the present findings and these prior ones. One is that we further subdivided location-focus utterances into two finer categories—those in which the point uniquely specified location (i.e., 'load-bearing points') and those in which the point was supplemented by other lexical material that helped specify location (i.e., 'load-sharing points'). In our data, the significant difference proved to be between these two finer subtypes rather than between the coarser-grained distinction between location-focus and not, as Enfield at al. reported. A possible reason for this discrepancy is that the difference between location-focus and other points was more subtle in our task than it was in Enfield et al. study. Our points all concerned a small set of possible locations, all relatively near the interlocutors. By contrast, their data—which was from naturalistic interviews—likely yielded more heterogeneous

pointing behaviors, produced for a wider variety of pragmatic purposes and toward referents at different distances. If we had elicited a broader range of utterances, we might well have seen a basic difference between location-focus points and other points. Our findings about extension also fit with earlier observations that gestures with heightened communicative status exhibit a greater degree of effort (Peeters et al., 2015; Cooperrider, 2017). For instance, one study found that participants were more likely to point with their arms fully extended when their interlocutors could see them, compared to when they could not (Bangerter and Chevalley, 2007).

Although we found that the extension of points reflected semantic integration in a way common to speakers and signers, we did not find the same pattern for duration. Signers' points were longer in duration when they were load-bearing (much as their load-bearing points were more likely to involve full arm extension). But speakers' points did not fit this pattern—there was no difference in duration between load-bearing and load-sharing points. (This is another way that our data appear to depart from Enfield et al. (2007). They observed—but did not formally analyze—a pattern in which big points tended to be held for longer.) Although at first puzzling, this pattern becomes intelligible in light of our findings about structural integration. Signers' points are highly constrained by the need to produce other signs in the manual channel. Speakers' points are not so constrained, particularly because speakers do not often produce more than one gesture per utterance (McNeill, 1992). As a result, during longer utterances—such as the explanatory utterances we observed here—speakers' points may "stretch out" for longer periods of time. Further studies will be needed to identify the precise factors that determine how long pointing gestures are maintained. One possibility is that duration in speakers' points reflects how the point relates to the speech it accompanies—i.e., how it "takes scope" over some relevant portion of the utterance. The upshot is that, in signers, duration serves as a reliable cue to the importance of the point in the utterance but, in speakers, duration is more complexly determined.

We also analyzed, in the speakers, whether points that were associated with demonstratives differed in form from points that were not. (This analysis could not be carried out in signers because ASL does make regular use of lexical signs that serve to highlight co-occurring points in the same way.) We found that points associated with demonstratives were more likely to involve full arm extension, and were longer in duration, than points not associated with demonstratives. One way to understand these findings is that speakers face a semantic integration problem that signers do not. As discussed, signers' points occur in the same articulatory channel as the majority of the referential content—the hands. Speakers' points, by contrast, occur in a different channel from the majority of the referential content; speakers thus face the task of stitching manual content in with the spoken content when it is critical to do so. Spoken demonstratives support this stitching by signaling that there is critical content in the secondary, manual channel. In turn, fully extending the point and holding it for a longer period of time further supports this stitching by enhancing the gesture's salience. The fact that points associated with demonstratives are longer in duration than points not associated with demonstratives further underscores the fact that, in speakers, the duration of a pointing gesture is complexly determined by its coordination with speech. Sometimes it may reflect constraints of structural integration—or the *absence* of a same-channel constraint, in this case—and other times it may reflect constraints of semantic integration—the need to make a point salient in order to help stitch it in with speech.

A final set of findings concerned structural integration. In a recent study, we described a same-channel constraint that exerts pressure on pointing signs, but not on pointing gestures (Fenlon et al., 2019). In this earlier work, we did not analyze the mechanics of this constraint in detail, but we identified several hallmarks of pointing signs that might reflect it. Here, we sought to confirm these broader patterns while also delving more deeply into how structural integration actually plays out in speakers and signers. As in our earlier work (Fenlon et al., 2019), we found here that speakers' points were overall much longer in duration than signers' points—in both studies about twice as long, despite a number of differences between the data sets. We also found that signers were much less likely to produce full-extension points than speakers, perhaps reflecting a general economy of effort that the same-channel constraint encourages.

Both of the broader patterns in speakers and signers just mentioned make sense when we look more closely at how these points articulate with other signals. Speakers' points showed a strong tendency to span across neighboring lexical material (spoken words), and occasionally even bridge over adjacent utterances. Although signers' points are free to span across lexical content in this way—and sometimes did—their strong tendency was to slot in between neighboring lexical material (manual signs). The pressure to slot in encourages shorter, less effortful points in signers; the absence of this constraint makes room for longer lasting, more effortful points in gesturers. A question that arises is whether these broad tendencies are intrinsic to these types of communication—spoken vs. signed—or are part of the conventional practices one comes to master. One way to investigate this would be to see whether young speakers and signers show these same strong tendencies from the start, or whether they are gradually acquired like other discourse conventions.

CONCLUSION

Human communication is increasingly understood as composite in nature—as an activity that integrates different types of communicative elements (e.g., Clark, 1996; Enfield, 2009; Vigliocco et al., 2014). These elements include not only the highly conventionalized symbols that have traditionally been the focus of linguistic inquiry, but also more tailored, gradient elements—from depicting signs and ideophones, to size-specifying constructions and pointing gestures. An emerging question within this framework concerns how such elements are integrated into seamless, fluent discourse (e.g., Davidson, 2015; Clark, 2016; Dingemanse and Akita, 2017; Lu and Goldin-Meadow, 2018). Here, we used pointing—a ubiquitous and multi-functional act—as a paradigm case. We have attempted to shed light on how pointing is integrated into language—and, moreover, how this

integration differs across spoken and signed communication. Of course, there are a number of aspects of integration that we did not touch on, even for the case of pointing. For example, work remains to be done on how pointing might be coordinated with specific grammatical structures or sequential environments. Work also remains to be done on whether aspects of the integration of pointing into language, such as those described here, generalize—for instance, to other types of pointing (e.g., points to invisible entities; Flack et al., 2018) or to communities with different articulator preferences (e.g., those who rely heavily on non-manual pointing; Cooperrider et al., 2018) and practices (e.g., points to the sun's arc to refer to time of day; Floyd, 2016). Beyond pointing, a host of questions await about the integration of other types of signals into language. Such questions will become more and more central as linguistic inquiry broadens its focus to account for the heterogeneous nature of communication—and how it coheres.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://osf.io/wckx5/ 200601.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Social and Behavioral Sciences Institutional Review Board, University of Chicago. The participants provided their written informed consent to participate in this study.

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors conceived of the research questions and study design. JF supervised the data collection, with input from all authors. KC and JF supervised the gesture and language coding, with input from all authors. JK conducted all statistical analyses, with input from KC. KC drafted the manuscript, with critical revisions provided by all authors. All authors approved the final version of the manuscript.

FUNDING

This article was supported by a grant from the Neubauer Collegium at the University of Chicago, awarded to Diane Brentari and Susan Goldin-Meadow.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this research from the Neubauer Family Collegium for Culture and Society and the Center for Gesture Sign and Language at the University of Chicago. We also thank Averi Ayala and Zena Levan for research assistance and three reviewers for their valuable feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.567774/full#supplementary-material.

REFERENCES

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi:10.1111/j.0956-7976.2004.00694.x

Bangerter, A., and Chevalley, E. (2007). "Pointing and describing in referential communication: when are pointing gestures used to communicate?," in *CTIT proceedings of the workshop on multimodal output generation (MOG)*. Editors I. Van Der Sluis, M. Theune, E. Reiter, and E. Krahmer (Scotland: Aberdeen).

Bates, D., Mächler, M., Ben, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67 (1), 1–48. doi:10.18637/jss.v067.i01

Bates, E. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.

Brentari, D., González, C., Seidl, A., and Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Lang. Speech* 54, 49–72. doi:10.1177/0023830910388011

Capirci, O., and Volterra, V. (2008). Gesture and speech: the emergence and development of a strong and changing partnership. *Gesture* 8 (1), 22–44. doi:10.1075/gest.8.1.04cap

Cartmill, E., Hunsicker, D., and Goldin-Meadow, S. (2014). Pointing and naming are not redundant: children use gesture to modify nouns before they modify nouns in speech. *Dev. Psychol.* 50 (6), 1660–1666. doi:10.1037/a0036003

Clark, H. H. (1996). *Using language*. (Cambridge: Cambridge University Press).

Clark, H. H.(2003). "Pointing and placing," in *Pointing: where language, culture, and cognition meet*. Editor S. Kita (Mahwah, NJ: Lawrence Erlbaum), 243–268.

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123 (3), 324–347. doi:10.1037/rev0000026

Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53 (8), 632–656. doi:10.1080/0163853X.2015.1094280

Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture* 16 (3), 176–202. doi:10.1075/gest.16.2.02coo

Cooperrider, K., Slotta, J., and Núñez, K. (2018). The preference for pointing with the hand is not universal. *Cogn. Sci.* 42 (4), 1375–1390. doi:10.1111/cogs.12585

Cormier, K., Schembri, A., and Woll, B. (2013). Pronouns and pointing in sign languages. *Lingua* 137, 230–247. doi:10.1016/j.lingua.2013.09.010

Cumming, G. (2014). The new statistics: why and how. *Psychol. Sci.* 25 (1), 7–29. doi:10.1177/0956797613504966

Cumming, G., and Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *Am. Psychol.* 60 (2), 170–180. doi:10.1037/0003-066X.60.2.170

Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics Philos.* 38 (6), 477–520. doi:10.1007/s10988-015-9180-1

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguistics* 17 (4), 463–489. doi:10.1515/COG.2006.015

Dingemanse, M., and Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese. *J. Linguistics* 53 (3), 501–532. doi:10.1017/S002222671600030X

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., and Paolino, D. (1993). Outline of discourse transcription. In Edwards & Lampert (Eds.), *Talking data: transcription and coding in discourse research* (pp. 45-48). Hillsdale, NJ: Erlbaum.

Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York: Aldine de Gruyter. doi:10.1017/S0140525X00060416

Enfield, N. J., Kita, S., and de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *J. Pragmatics* 39 (10), 1722–1741. doi:10.1016/j.pragma.2007.03.001

Enfield, N. J. (2009). *The anatomy of meaning: speech, gesture, and composite utterances*. Cambridge: Cambridge University Press. doi:10.1111/j.1467-9655.2010.01661_28.x

Fenlon, J., Cooperrider, K., Keane, J., Brentari, D., and Goldin-Meadow, S. (2019). Comparing sign language and gesture: insights from pointing. *Glossa*. 4 (1), 1–26. doi:10.5334/gjgl.499

Fenlon, J., Denmark, T., and Campbell, R. (2007). Seeing sentence boundaries. *Sign. Lang. Linguistics* 10 (2), 177–200. doi:10.1075/sll.10.2.06fen

Fenlon, J., Schembri, A., and Cormier, K. (2018). Modification of indicating verbs in British Sign Language: a corpus-based study. *Lang.* 94 (1), 84–118. doi:10.1075/sll.10.2.06fen

Ferrara, L., and Hodge, G. (2018). Language as description, indication, and depiction. *Front. Psychol.* 9, 1–15. doi:10.3389/fpsyg.2018.00716

Flack, Z. M., Naylor, M., and Leavens, D. A. (2018). Pointing to visible and invisible targets. *J. Nonverbal Behav.* 42 (2), 221–236. doi:10.1007/s10919-017-0270-3

Floyd, S. (2016). Modally hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Lang.* 92 (1), 31–64. doi:10.1353/lan.2016.0013

Gelman, A., and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Comput. Stat.* 15. 373–390. doi:10.1007/s001800000040 10.1007/s001800000040

Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 1, 1–82. doi:10.1017/S0140525X15001247

Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian Sign Language): are the data sufficient to posit a grammatical class of "pronouns" in Auslan?. *Corpus Linguistics Linguistic Theor.* 9 (1), 109–159. doi:10.1515/cllt-2013-0012

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (2003a). "Pointing: a foundational building block of human communication," in *Pointing: where language, culture, and cognition meet*. Editor S. Kita (Mahwah, NJ: Lawrence Erlbaum), 1–8.

S. Kita (Editor) (2003b). *Pointing: where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

Lausberg, H., and Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behav. Res. Meth. Instrum. Comp.* 41 (3), 841–849. doi:10.3758/BRM.41.3.841

Liszkowski, U. (2006). "Infant pointing at 12 months: communicative goals, motives, and social-cognitive abilities," in *Roots of human sociality*. Editors N. J. Enfield and S. C. Levinson (Oxford, England: Berg Publishers), 153–178.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., and de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cogn. Sci.* 36, 698–713. doi:10.1111/j.1551-6709.2011.01228.x

Lu, J., and Goldin-Meadow, S. (2018). Creating images with the stroke of a hand: depiction of size and shape in sign language. *Front. Psychol.* 9 (1276), 1–15. doi:10.3389/fpsyg.2018.01276

McBurney, S. L. (2009). "Pronominal reference in signed and spoken language: are grammatical categories modality-dependent?," in *Modality and structure in signed and spoken languages*. Editors R. P. Meier, K. Cormier, and D. Quintos-Pozos (Cambridge: Cambridge University Press), 329–369.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.

Meier, R. P., and Lillo-Martin, D. (2003). The points of language. *Humana. Mente J. Philosoph. Stud.* 24, 151–176.

Nespor, M., and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications. doi:10.1007/978-94-009-2733-9_16

Peeters, D., Chu, M., Holler, J., Hagoort, P., and Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* 27 (12), 2352–2368. doi:10.1162/jocn_a_00865

Peirce, C. S. (1940). *Philosophical writings of Peirce*. Editor J. Buchler (New York: Dover).

R Core Team (2016). R: a language and environment for statistical computing. Vienna, Austria. Available at: https://www.R-project.org/.

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. USA* 102 (7), 2661–2665. doi:10.1073/pnas.0405448102

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 369, 20130292. doi:10.1098/rstb.2013.0292

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership