SIMPLE AND SIMPLIFIED LANGUAGES

EDITED BY: Andras Kornai, Sigal Uziel-Karl and Scott A. Hale
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication

frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88976-239-2
DOI 10.3389/978-2-88976-239-2

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

SIMPLE AND SIMPLIFIED LANGUAGES

Topic Editors:
Andras Kornai, Computer and Automation Research Institute (MTA), Hungary
Sigal Uziel-Karl, Achva Academic College, Israel
Scott A. Hale, University of Oxford, United Kingdom

Citation: Kornai, A., Uziel-Karl, S., Hale, S. A., eds. (2022). Simple and Simplified Languages. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-239-2

Table of Contents

frontiers | Frontiers in Communication

EDITORIAL
published: 05 May 2022
doi: 10.3389/fcomm.2022.910680

Editorial: Simple and Simplified Languages

*Sigal Uziel-Karl**

Department of Communication Disorders, Achva Academic College, Arugot, Israel

Keywords: pidgins and creoles, plain English, accessibility, augmentative and alternative communication, Automatic Text Simplification (ATS), language simplification, L2 learning and teaching

Editorial on the Research Topic

Simple and Simplified Languages

Language has always powerfully influenced people's lives (e.g., Fairclough, 2001). This influence is even more forceful in the current era, the Information Age, in which language-based products are abundantly available and extensively used, with information and communication constantly increasing their impact on our daily lives.

Managing this abundance of written or spoken information may pose a considerable challenge for specific populations. Language simplification is crucial for individuals with cognitive or sensory disabilities, language minorities, and economically or socially disadvantaged populations (e.g., migrant workers), for whom it may remove barriers to inclusive, equal, and independent participation in society (e.g., Uziel-Karl and Tenne-Rinde, 2018). Recognizing the social value of simplified language has led legislators and human rights organizations worldwide to promote laws and regulations on language simplification, e.g., the UN Convention on the Rights of Persons with Disabilities[1]. The benefit of language simplification for the general public has also been recognized in movements promoting "plain English" or "plain language." These campaigns have sought to make content accessible beyond specific disciplines, fighting increased complexity associated with highly technical language in legal, financial, or medical documentation to make it more understandable to laypersons[2].

The practical need for language simplification crosses time. Throughout history, the need for communication between speakers of different languages for trade or administrative purposes led to the development of pidgin languages. These would sometimes grow into creoles, becoming the first language of later generations[3]. Recent globalization trends and the prevalent use of the World Wide Web further highlight the necessity of language simplification for practical purposes like foreign language learning, language contact, and situations where technical vocabulary must be tightly controlled to promote cooperation[4].

Keeping up with the growing demand for simplified materials and adapting language and text simplification to diverse populations and settings requires efficient and fast methods of bulk simplification. This challenge creates fertile ground for research in the field. The papers in this Research Topic offer a broad perspective on current language simplification research encompassing numerous populations, typologically different languages, and various methodologies, addressing theoretical and practical questions.

OPEN ACCESS

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

***Correspondence:**
Sigal Uziel-Karl
sigal@alum.mit.edu

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

Received: *01 April 2022*
Accepted: *19 April 2022*
Published: *05 May 2022*

Citation:
Uziel-Karl S (2022) Editorial: Simple and Simplified Languages. Front. Commun. 7:910680. doi: 10.3389/fcomm.2022.910680

[1]The UN Convention on the Rights of Persons with Disabilities (CRPD). https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html (accessed March 18, 2022).
[2]http://www.plainenglish.co.uk/about-us.html; https://www.plainlanguage.gov/
[3]https://www.britannica.com/topic/language/Pidgins-and-creoles; and see Arends et al. (1994).
[4]http://www.asd-ste100.org

Yakpo examines creole prosodic systems from a typological perspective. The author argues that tone is not simplified or eliminated in creoles and contact languages. Instead, he proposes an areal continuum of tone systems roughly conterminous with tone in the east (Africa) and stress in the west (Americas). Kornai offers a way to determine the simplest "core" layer of vocabulary. He argues that a valuable notion of core vocabulary must synthesize both definitional simplicity (basic) and high occurrence (frequency) of a word. He recommends Kolmogorov complexity as the best formal means to integrate both aspects.

The next three papers discuss simplification in the context of non-verbal communication. Yum et al. examined whether Cantonese Chinese augmentative and alternative communication (AAC) users are sensitive to different types of communication partners during naturalistic AAC conversations. They describe differences in AAC users' strategies to communicate with peers vs. Speech and Language Pathologists. They suggest considering the base language and the communication partner in studies on graphic-based AAC. Savaldi-Harussi and Fostick studied the impact of graphic symbol modality on message construction. They compare verbal vs. graphic symbol production by Hebrew-speaking preschoolers. They demonstrate a significant difference in favor of verbal speech across different syntactic structures, concluding that graphic representation of complex linguistic structures requires explicit instruction. Astell et al. examined the efficiency of a non-verbal method of communication (Adaptive Interaction) in simplifying the interaction between caregivers and patients with dementia who can no longer speak. Their results suggest that non-verbal communication methods can streamline and improve caregiver-patient interaction.

The next four papers present various aspects of automated text simplification (Siddharthan, 2014). Dmitrieva et al. examined whether texts simplified for different learner groups are equally simple by investigating linguistic properties and specific simplification strategies used in Russian texts for three groups of primary school children (Native, Foreign, and Bilingual). They report that all text types are similarly accessible to young readers. However, different strategies are used for adapting or creating texts for each type of audience. Brunato et al. reviewed existing parallel corpora for Automatic Text Simplification (ATS) in different languages. They used Italian parallel corpora to compare different approaches to corpus building for ATS based on the methodology employed for their construction (manual vs. (semi)-automatic). They show that construction method affects original and simple corpora and report on differences between two variations of the manual corpora. Ebling et al. created a gold standard of sentence alignments based on four parallel corpora (standard/simplified German) compiled for evaluating automatic alignment methods on this gold standard. They note that one alignment method performs best on most data sources. They use two corpora as a basis for a sentence-based neural machine translation approach toward automatic simplification of German. They then extend the model to operate on multiple levels of simplified German. Harbusch and Steinmetz developed a computer-assisted writing tool for an extended version of Easy-to-Read German (LS) to enable LS readers to produce texts independently. They illustrate how to make dialogues of the automated tool intuitive and easy to use, reporting how well the software performs with different user groups.

Finally, Borghardt et al. examined how different online methods (eye-tracking, EEG, and fMRI) work in investigating the empirical validity of the Easy-Language guidelines by evaluating cognitive processing efficiency. They conclude that only examination of online methods combined with data triangulation in Easy Language research provides profound insights into the cognitive processing of simplified languages.

The papers presented here provide important insights into some major theoretical, technological, and practical questions in language simplification, and point to the challenges that still lie ahead.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

I wish to thank Andras Kornai and Scott Hale for the fruitful collaboration in editing the Research Topic. I would also like to thank *Editage* (www.editage.com) for English language editing.

REFERENCES

Arends, J., Smith, N., and Muysken, P. (1994). *Pidgins and Creoles*. Amsterdam: John Benjamins Publishing Company.

Fairclough, N. (2001). *Language and Power*. London: Longman.

Siddharthan, A. (2014). A survey of research on text simplification. *Int. J. Appl. Linguist*. 165, 259–298. doi: 10.1075/itl.165.2.06sid

Uziel-Karl, S., and Tenne-Rinde, M. (2018). "Making language accessible for people with cognitive disabilities: intellectual disability as a test case," in *Handbook of Communication Disorders*, eds A. Bar-On, and D. Ravid (Berlin: Muton De Gruyter), 845–862. doi: 10.1515/9781614514909-042

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Uziel-Karl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fpsyg.2021.686657

Sensitivity to Communication Partners During Naturalistic AAC Conversations in Cantonese Chinese

*Yen Na Yum[1], Soby Ka Wing So[2] and Rosanna Yuen-Yan Chan[3,4]**

[1] Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong, China, [2] SAHK, Hong Kong, China, [3] Centre for Perceptual and Interactive Intelligence, The Chinese University of Hong Kong, Hong Kong, China, [4] Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China

Previous studies have shown that graphic-based augmentative and alternative communication (AAC) output tend to be short and simple in structure with non-canonical word order, and that AAC users may show differences when communicating with peers compared to professionals such as speech therapists (STs). However, there was a lack of report for graphic-based AAC in the Chinese context, and the effect of communication partners had not been investigated systematically. In this study with 34 AAC users and 10 STs, we reported common and distinct features of free conversations in Cantonese graphic-based AAC, relative to AAC in other languages. We also found that AAC users were sensitive to different types of communication partners. In particular, when conversing with peers, AAC users produced long messages with equal proportion of questions and responses, which suggested active and bi-directional exchanges. In conversations with STs, AAC users showed high diversity in expressive vocabulary, indicating access to more semantic concepts. Results suggested that the base language and the communication partner are both influential factors that should be considered in studies of graphic-based AAC. The mobile AAC system facilitated free conversations in users with complex communication needs, affording an additional channel for social participation.

Keywords: augmentative and alternative communication, Cantonese Chinese, cerebral palsy, communication partner, complex communication needs, linguistic analysis, symbol

OPEN ACCESS

Edited by:
Sigal Uziel-Karl,
Achva Academic College, Israel

Reviewed by:
Eliseo Diez-Itza,
University of Oviedo, Spain
Ming-Chung Chen,
National Chiayi University, Taiwan

***Correspondence:**
Rosanna Yuen-Yan Chan
yychan@ie.cuhk.edu.hk

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

Received: *27 March 2021*
Accepted: *26 July 2021*
Published: *18 August 2021*

Citation:
Yum YN, So SKW and Chan RY-Y (2021) Sensitivity to Communication Partners During Naturalistic AAC Conversations in Cantonese Chinese. Front. Psychol. 12:686657. doi: 10.3389/fpsyg.2021.686657

INTRODUCTION

Augmentative and alternative communication (AAC) systems are commonly used by people with complex communication needs (CCNs) to supplement verbal communication, or in some cases, substitute for oral language. Systems developed for various communities have features (e.g., electronic vs. non-electronic, text-based vs. image-based) that facilitate usage in different contexts. McNaughton and Light (2013) raised important ways that mobile technologies impact the current use and continued development of ACC, including generating greater acceptance in society and wider dissemination of services. In addition, mobile technology allow for new patterns of communication. As AAC mobile applications support network connectivity and interactions between two remote devices, users can converse with others in real-time to achieve a two-way communication which was infeasible before. EasyDial^{TM} is a first-of-its-kind cloud AAC system

for people with CCN in Hong Kong that mimicked the mobile telephone functions, developed by the second and the third authors (Chan, 2014; Chan et al., 2016). In the pilot implementation of EasyDial^{TM}, various user tests showed that while some specific concepts were missing from the symbol list, the app interface was convenient and friendly. The majority of participants mastered symbol selection and message composition within 10 min and had conversations with minimal assistance from others (Chan et al., 2016). The EasyDial^{TM} system has been launched to the public sector in Hong Kong in November, 2019 by SAHK, which is a rehabilitation service provider serving more than 15,000 families on an annual basis. This paper reported a linguistic analysis on AAC conversational data collected from the cloud server and outlined potential applications of the results.

As an alternative to oral language, AAC systems may undergo a simplification process to facilitate language unit selection and output. However, the degree of change depends on the specific needs of the target group. On one end of the spectrum, an AAC system may contain the same lexicon as oral language, augmenting it with word typing and electronic speech synthesis. Users may produce identical messages as normal speech, although typically at a much slower rate and with altered syntax and pragmatics compared to non-AAC counterparts (Friginal et al., 2013, 2016). This type of AAC mainly caters for users who have normal verbal or intellectual abilities, but with difficulties in speech production. On the other hand, an AAC system may be drastically simplified from the base language to cater for users with limited verbal, intellectual, or motor abilities. An example is the picture exchange communication system (PECS), which is characterized by a much reduced vocabulary representing core concepts with words and simple pictures. This type of AAC is suited for children or individuals with developmental disabilities, e.g., autism spectrum disorder (Chen et al., 2015; An et al., 2017), intellectual disabilities (ID) (Deckers et al., 2017), or cerebral palsy (Chan et al., 2016; Soto and Clarke, 2017, 2018).

In aided AAC systems such as PECS, selection and display of core vocabularies is highly specific to the language and needs of the target users. Many corpora had been developed with the intention to identify the optimal set of symbols for AAC. Speech corpora came from individuals with normal speech abilities [e.g., English-speaking school-age children in Boenisch and Soto (2015); Mandarin speaking adults in Chen et al. (2009) and Liu and Sloane (2006)] or from the target population [e.g., Dutch-speaking children with Down Syndrome in Deckers et al. (2017)], and text corpora were also considered (e.g., Mühlenbock and Lundälv, 2011). An efficient AAC system should provide concepts that match what the users need to convey. For example, an AAC system used in school setting will need academic concept representation. McCarthy et al. (2017) showed that many age-appropriate concept words were not adequately represented or easily accessible in four commercially available AAC systems. The authors highlighted the need for educators to recognize and address the limitations of basic concept content in pre-packaged AAC software or applications. In informal settings, conversations may exhibit a greater social orientation, with the main focus on daily life and common topics of interest. Free conversational contexts provide another setting to examine vocabulary selection and usage in an ecologically valid manner.

Many studies have shown that AAC vocabulary and morpho-syntax does not parallel speech or written language (Bernardi and Tuzzi, 2011; Friginal et al., 2013, 2016). Indeed, it has been argued that using AAC requires separate skills apart from those supporting oral language (Light, 1997; Smith, 2015). Although AAC comprehension and expression exhibits a wide range of grammatical complexities, in general, utterances tend to be shorter than would be expected based on participant profiles (Binger and Light, 2008). In a review of 31 studies pertaining to morphology and syntax in graphic symbol-based AAC, Smith (2015) reported four main linguistic patterns: (1) dominance of single-symbol output, (2) persistence of simple clause structures, (3) changes in word order, and (4) errors in inflectional morphology. As Chinese has flexible word order and minimal inflectional morphology, it contrasts with the bulk of the literature in English and European languages.

Conversation-based AAC intervention can increase communication abilities in target users, such as spontaneous communication and use of requests (Soto and Clarke, 2017, 2018). Individuals with CCN using aided AAC have been described as passive responders, as they tended to ask few questions and follow the set topics (e.g., Light, 1988; Clarke and Kirton, 2003). In a study with 12 children with physical disabilities using AAC systems with their speaking peers in school (Clarke and Kirton, 2003), children with CCN were significantly more likely to respond than initiate interactions compared to their naturally speaking peers. Even so, the distribution of turn taking in these conversations was more equal than what had been identified in conversations between adults and children using AAC systems. Lund and Light (2007) found that during interactions with their caregivers, more experienced AAC users fulfilled most of their obligatory turns and more than half of their non-obligatory turns, with a majority of participants able to approach reciprocity in turn taking. The communicative functions most frequently used by AAC users were confirmations/denials and provision of information. In some cases, the passiveness may be because the physical limitations of the user led to less control over the use of the AAC system. Pinto and Gardner (2014) reported that a child was able to use eye–gaze strategies to indicate interests both within and outside the AAC system, and the communication partner is tasked to be sensitive to these signals. Overall, individuals with CCN demonstrated ability to use AAC to serve a variety of communicative functions, but there seemed to be differences in usage under different contexts and with different communication partners (e.g., peers vs. caregivers vs. professionals).

Considering the diversity and impact of language type, user characteristics, and usage contexts on AAC output, a cross-language study may be of particular value in examining the universality of previous observations. This study reported Cantonese Chinese graphics-based AAC free conversations to examine how individuals with CCN use AAC with different communication partners, i.e., peers or speech therapists (STs). The usage pattern of STs toward individuals with CCN was also investigated as a comparison group. On the symbol level,

we examined the core vocabulary used in free conversations, including frequency distribution and commonality of use in different groups. On the message-level, we compared message length, proportion of single symbols, type-token ratio, and communicative functions across groups. On the conversation level, we compared if the overall conversation length and the number of turns differed between user–user and user–ST conversations. We expected that individuals with CCN and STs would show different usage patterns that may reflect their verbal and motor abilities. The comparison between communication partners would show how social contexts influence language use in individuals with CCN. We also hypothesized that the data patterns would reveal some linguistic characteristics common to all AAC systems and some characteristics specific to Chinese.

MATERIALS AND METHODS

Participants

The participants were 34 individuals (14 females, mean age = 32.1 years, SD = 15.4, range = 10–55) with CCNs (hereafter, users) and 10 STs. Cerebral palsy (CP) was the main clinical diagnosis of the users (17 dyskinetic CP, 6 spastic CP, 6 mixed type CP, and 5 other diagnoses, e.g., epilepsy), and 85% of users had comorbid ID, of which 13 had mild ID, 15 had moderate ID, and 1 had moderate to severe ID. As the majority of users were adults, their language levels were assessed by STs using criterion-referenced assessment with the Reynell Developmental Language Scales, corresponding to receptive language age (mean = 4.1, SD = 1.2, range = 2–5). All of the participants were exposed to AAC before the study. They were able to use communication book or board with photos or picture presentation. A subset of them (n = 9) were also able to type *via* computer or mobile phone but at significantly reduced accuracy and efficiency. The participants were assisted by STs who were working in either rehabilitation or school settings with at least 4 years of experiences in clinical practice. All STs were experienced in training people with CCNs to use AAC.

The EasyDial^{TM} System

The current study was performed with the EasyDial^{TM} system. The system's purpose is to provide bidirectional mobile phone-like communication services in form of real-time exchange of AAC picture symbols over the networks. When using the system, users are shown an interface and symbol selection is accompanied by text-to-speech output of Cantonese Chinese in an adult female voice [see details in Chan et al. (2016)]. Currently, there are a total number of 665 communication symbols available in the system; while these symbols are classified into 17 categories (e.g., people, food, activities) according to the semantic nature. There are two important features in the system:

- The capability of performing semantic recommendation of AAC communication symbols using a recurrent deep learning algorithm designed by the project team; which greatly shortened the symbol selection time in users with severe motor and cognitive limitations; and
- The curation of a growing volume of anonymized AAC usage data that has a great potential to inform evidence-based speech therapy practices through big data analytics.

Besides, other usability adaptation features such as personalized application client interface and touch screen dexterity settings are also available.

Procedure

Data collection occurred in 2016 and 2019 during the development and prototyping stages of the system, respectively. Informed consent to participate in the study was obtained from all participants. Participants had unscripted free conversations using EasyDial^{TM} as AAC support, in either face-to-face conversations or AAC-based mobile phone calls. Free conversation here was defined as message production specifically for communication, without limit on conversation topic. Since EasyDial^{TM} supported 1-to-1 messaging, the messages produced were directed to a receiver and can be interpreted as intentional communication.

Although EasyDial^{TM} was also used by caregivers and acquaintances, only data from users and STs were included. About 40 messages contained repeated symbols that were caused by motor control of users and were excluded from analysis. The trimmed dataset included 1108 messages with 31.9% between users (UtoU, users n = 25), 33.9% from users to STs (UtoST, users n = 26), and 34.1% from STs to users (SToU, STs n = 10).

Data Analysis

On the symbol level, descriptive statistics of the dataset were reported in terms of the type and token frequencies of the selected symbols across groups. The frequency distribution of symbol use in the three groups was additionally examined by plotting the log of the frequency against the log of the rank. Commonality scores were calculated by counting the percentage of participants who used a particular symbol, e.g., commonality of "you" is 100% if all participants have used it at least once in their messages. This measure complements lexical frequency in that high commonality shows that a symbol is widely used, not only repeatedly used by a select few participants.

On the message level, message lengths across groups were compared using linear mixed-effects models (Baayen et al., 2008) using the Satterthwaite method for degrees of freedom. A by-participant random intercept was used in the model to account for clustering of paired participants in UtoU and UtoST conditions and to minimize the effect of unequal number of observations from participants. *Post hoc* tests were run with Bonferroni adjustments for multiple comparisons. Chi-square tests were used to determine group differences in count data, including number of single symbol messages, number of repairs, and distribution of communicative functions. *Post hoc* comparisons between groups were done with Bonferroni adjustments for multiple comparisons. Two independent raters judged the communicative functions of the messages (inter-rater agreement = 94.6%). The disagreements contained cases where the intended meaning was ambiguous, for example, "you; eat" may be a directive, a statement, or a question lacking

the question marker; "I; greet" may be a direct greeting or a comment on a past action; or a single message may be split into two: (1) "you; happy" + (2) "question marker." Disagreements were resolved after discussion, prioritizing contextual clues from the conversations.

On the conversation level, the number of messages and turns within a conversation and their ratio were examined. A conversation was either defined by greetings at the beginning or end, or a separation in time. The messages were included only if the sender and receiver each had at least one turn in the conversation. Eight messages did not meet this criterion and were excluded from this analysis (0.008%). The remaining data were log-transformed to approximate a normal distribution being before submitted to linear mixed models with by-pair random intercept to compare user–user and user–ST conversations.

The mixed-effects models were run using GAMLj (Gallucci, 2019) on Jamovi (2020) on the R Core Team (2019). For all inferential statistical tests, α level was set at 0.05.

RESULTS

Symbol Level

In the dataset, 2080 symbols were used (token counts), of which 290 were distinct (type counts). Overall, 182 symbols (type) appeared only once or twice in the corpus, so the core vocabulary used in free conversations is small. The 9 most frequently used symbols already accounted for over 50% of all symbols used, and 40 symbols represented 75% of total usage. The distribution of symbol use followed the Zipf's scaling law [$P(r) \sim r^{-\alpha}$] characteristic of natural languages, with R^2 above 95% in all three sub-groups (**Figure 1**). Due to the limited size of the dataset, we did not conduct further statistical tests for this measure. In terms of parts of speech, nouns and verbs were the most frequently used and accounted for 23.1 and 22.4% of the symbols. Pronouns and interjections occurred at the next highest frequency at 19.8 and 19.3%, respectively. Adjectives (6.5%) and adverbs (1.6%) occurred with low frequency, while particles (2.7%), conjunctions (0.1%), and prepositions (0%) were used minimally. Symbols representing phrases with two or more words (e.g., take off jacket, play on the computer) were used 4.4% overall (see **Table 1** for properties of symbols with the highest frequency, including the category and parts of speech).

Among the high frequency symbols, commonality was highest in the STtoU messages, suggesting uniform word choice across different STs, while commonality was lowest in UtoST messages, indicating idiosyncratic word usage. The breakdown of type and token frequencies across groups can be found in **Table 2**. While the three sub-groups had comparable values for number of messages, the type count was higher and token count was lower for the UtoST group relative to the other two groups.

Message Level

The random effect likelihood ratio test suggested clustering of data by participant ($p < 0.001$), but the proportion of variance explained by the random effect was not high, with intraclass correlation (ICC) = 0.071. Results showed a statistically significant group effect $F(2, 52.1) = 13.6$, $p < 0.001$. Pairwise

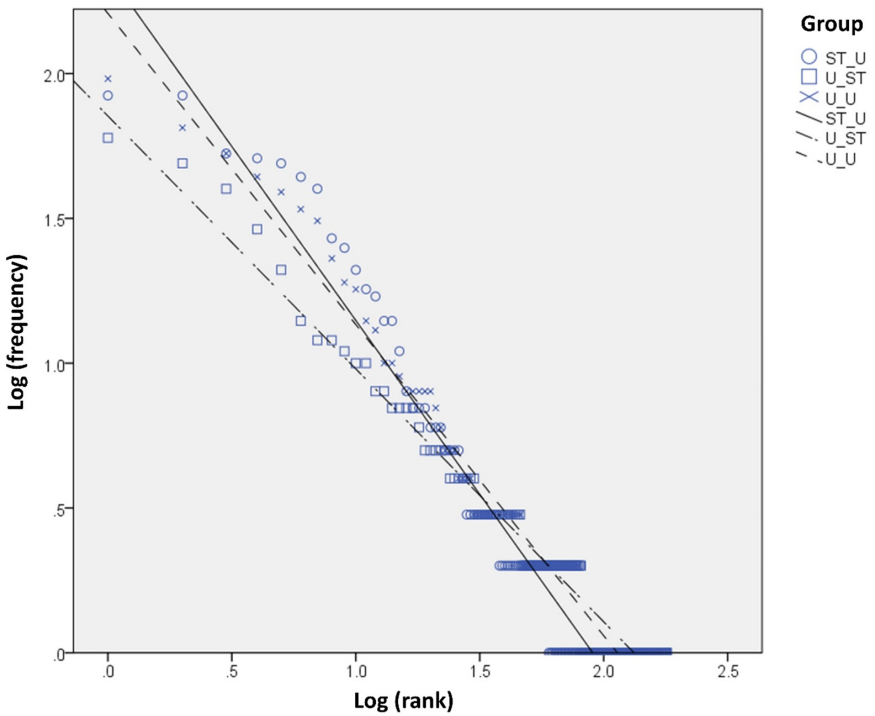

FIGURE 1 | Log–log plot by participant group.

TABLE 1 | Features of the most frequently used symbols in the dataset.

Rank	Symbol	Translation	Token Count	% of occurrence	Cumulative % of occurrence	Category	Part of speech	Commonality in UtoU	Commonality in UtoST	Commonality in STtoU
1	你好	Hello	189	9.1	9.1	Social expression	Interjection	**	***	****
2	我	I	172	8.3	17.4	People	Pronoun	***	**	***
3	你	You	149	7.2	24.5	People	Pronoun	***	*	***
4	食	Eat	139	6.7	31.2	Daily life	Verb	**	*	****
5	拜拜	Goodbye	131	6.3	37.5	Social expression	Interjection	**	***	****
6	鍾意	Like	92	4.4	41.9	Thought	Verb	*	*	***
7	甚麼	What	81	3.9	45.8	Function word	Pronoun	*		****
8	OK	OK	58	2.8	48.6	Social expression	Interjection	*	**	***
9	開心	Happy	56	2.7	51.3	Emotion	Adjective	**	*	**
10	嗎	Question	55	2.6	54.0	Function word	Particle			**
11	飲	Drink	50	2.4	56.4	Daily life	Verb	*	*	**
12	打招呼	Greet	30	1.4	57.8	Social expression	Verb	*		*
13	去	Go	28	1.3	59.2	Daily life	Verb			***
14	有	Have	28	1.3	60.5	Object property	Verb			*
15	麵包	Bread	20	1.0	61.5	Food	Noun	*		
16	多謝	Thank you	20	1.0	62.4	Social expression	Interjection		*	*
17	哪裡	Where	17	0.8	63.3	Function word	Adverb			***
18	做甚麼	Doing what	16	0.8	64.0	Function word	Phrase			
19	燒賣	Siu Mai	16	0.8	64.8	Food	Noun	*		
20	水	Water	15	0.7	65.5	Food	Noun	*		*

*****81–100%; ***61–80%; **41–60%; *21–40%.*

TABLE 2 | Message characteristics by participant group.

Participant group	No. of message	Type count	Token count	Message length		% single symbol	% repairs
				Mean	SD		
Users to users	354	144	741	2.09	1.24	**46.3**	10.7
Users to speech therapists	376	180	581	1.54	1.05	**66.5**	**17.6**
Speech therapists to users	378	122	757	2.01	1.22	50.5	10.1
Overall	1108	290	2079	1.88	1.19	54.6	12.8

Values in bold denote significant differences from expected values.

comparisons with Bonferroni adjustments indicated that the UtoST messages were significantly shorter than STtoU ($p = 0.002$) and UtoU ($p < 0.001$) messages, while the latter two groups did not differ from each other ($p = 1.00$; observed values in **Table 2**). A Chi-square test showed that there was group difference in the proportion of single symbol messages, χ^2 (2) = 33.7, $p < 0.001$. *Post hoc* tests revealed that UtoU messages had significantly less single symbol messages, UtoST messages had significantly more single symbol messages, while STtoU messages did not differ from the expected values. There was also group difference in the proportion of repairs, χ^2 (2) = 11.5, $p = 0.003$, where the UtoST messages had significantly more repairs than expected. Tabulated counts of four main communicative functions (i.e., convention, statement, question, and response), and an unspecified category are presented in **Table 3**. The Chi-square test indicated that communicative functions were different depending on group, χ^2 (8) = 211, $p < 0.001$. The most frequent function was conventionalized social language (e.g., greetings, thank you), which occurred significantly more often in STtoU messages and less often in UtoU messages. The next frequent function was statement, which included sharing new information or commenting, and STtoU messages had significantly less of this function than expected. Questions and responses occurred similar numbers of times overall, suggesting appropriate social responsiveness. As predicted, these functions showed a clear difference in ST and user interactions, where STs tended to ask questions rather than answer them, while users tended to answer questions but not raise them. Remarkably, users had a balanced question and answer proportion when talking amongst themselves. The unspecified category included null

TABLE 3 | Contingency table of communicative functions by participant group.

Groups		Communicative Functions					
		Convention	Statement	Question	Response	Unspecified	Total
Users to users	Observations	**89**	120	59	58	**28**	354
	% within row	**25.1**	33.9	16.7	16.4	**7.9**	100.0
Users to speech therapists	Observed	106	109	**18**	**125**	18	376
	% within row	28.2	29.0	**4.8**	**33.2**	4.8	100.0
Speech therapists to users	Observed	**135**	**48**	**143**	**46**	**6**	378
	% within row	**35.7**	**12.7**	**37.8**	**12.2**	**1.6**	100.0
Overall	Observed	330	277	220	229	52	1108
	% within row	29.8	25.0	19.9	20.7	4.7	100.0

Values in bold denote significant differences from expected values.

messages, consecutive repeated messages, and messages with unclear meanings. There are more unspecified UtoU messages and less unspecified STtoU messages than expected. The request function was notably absent in this dataset, likely due to the conversation context where the focus was casual social exchange rather than training for AAC utility.

Conversation Level

There were 35 conversations among 21 user–user pairs and 58 conversations among 29 user–ST pairs. The mean number of messages within a user–user conversation was 9.94 (SD = 12.1, range = 2–68), while that for user–ST conversation was 13.0 (SD = 7.68, range = 4–52). The random effect likelihood ratio test indicated significant clustering of data by different conversation pairs (p = 0.009), with ICC of 0.541 showing that pair-specific differences explained much of the variance in the data. Still, user–user conversations were statistically shorter than user–ST conversations, $F(1, 35.9) = 4.14$, p = 0.049. Likewise, the mean number of turns within a user–user conversation was 6.74 (SD = 5.93, range = 2–25), less than that for user–ST conversations, which was 9.84 turns (SD = 4.19, range = 2–26). The random effect likelihood ratio test for the pair random intercept was marginally significant (p = 0.095), with ICC of 0.471. The difference in number of turns between groups was also statistically significant, $F(1, 30.4) = 8.37$, p = 0.007. The ratio of number of turns to number of messages was 0.774 (SD = 0.201) and 0.791 (SD = 0.124) for user–user and user–ST conversations, respectively. The random effect of pairs was significant (p = 0.017), ICC = 0.309, however, there was no statistical difference between groups for this metric, $F(1, 38.6) = 1.49$, $p = 0.230$.

DISCUSSION

In the overall type count of the current dataset, less than half of the available symbols in EasyDial^{TM} were used. Despite clear differences in the linguistic features of graphic symbol-based AAC and the full Cantonese language, the frequency distribution of AAC data in free conversations in the three sub-groups all exhibited the Zipf's scaling law. The Zipf's law has been reported in many natural languages, including in specific populations such as child (e.g., Baixeries et al., 2013) or elderly language use (Abe and Otake-Matsuura, 2021). As seen from **Figure 1**, the exponent α is close to 1 for all three groups, consistent with these studies. Interestingly, Baixeries et al. (2013) observed that syntactic complexity measured by mean length of utterance is negatively related to the exponent values in child language development, but Abe and Otake-Matsuura (2021) found no relation between cognitive functions and the exponent value in free conversations in older adults. Although the present dataset was not of sufficient size to do more fine-grained analyses, future studies could explore such relationships with AAC user characteristics.

Cantonese graphic symbol-based AAC elicited distinctive morpho-syntactic usage patterns that aligned with previous reports in other AAC languages, specifically, single symbol use and simple clause structure (Smith, 2005; Binger and Light, 2008). The high proportion of single symbol utterances were partially due to inclusion of greetings common in phone calls (e.g., "hello," "goodbye"), and simple answers to questions. This is because even simple questions need several symbols while answers can be single symbol (e.g., "you like drawing?" vs. "yes"). Another reason is that symbol selection and output was usually effortful in AAC users, and there could be difficulty in motor control or command of the system for users in conversations, as evidenced by the number of repairs and messages with unspecified functions. So the prevalence of single symbol may be a strategic choice to maximize communication efficiency. Similarly, simple clause structure may be adopted to convey the central meaning of messages, with omission of function words or less important elements such as adjectives and adverbs. Overall, there were not many multi-symbol messages, and the symbol set included some common phrases to enhance communication efficiency, so coding of grammatical structure of individual messages was not done. Nevertheless, canonical SVO word order was observed for many messages with multiple symbols, and errors in word order were not particularly noted (unlike e.g., Binger and Light, 2008). However, since Chinese sentences may have a topic-comment structure and word order is not strict, even if symbols did not follow a typical word order, the message could still be interpreted – only about 5% messages had unspecified communicative function. Therefore, in terms of lexical choice and dominance of single symbols, we did not note much

cross-language differences in AAC usage. Meanwhile, because of the flexible word order and minimal morphological inflections that is characteristic of Cantonese Chinese, grammatical errors in word order or verb or number agreements did not occur, and message meanings were largely interpretable in context. In a small number of messages, errors appeared to be driven by picture processing. For example, a student used the symbol for "father" when addressing the ST, presumably because the picture depicted an older male. This suggested that picture representations in the AAC system is important for its appropriate use regardless of the verbal language.

A salient finding was the systematic differences in message construction in how users with CCN and professionals such as STs used AAC. Kent-Walsh et al. (2015) had reported in a meta-analysis that interventions by AAC partner instruction were highly effective across a range of participant types, intervention approaches, and outcome measure characteristics. In AAC interventions, modeling, expectant delay, and open-ended question asking were frequently targeted interaction skills. Although the current data were collected under free conversation instead of explicit intervention context, STs still employed a structured client-centered approach and asked many questions to stimulate responses from users with CCN. Results from the commonality scores and analysis of communicative functions supported this interpretation. In general, users appeared more reserved when talking to STs compared to peers, producing shorter messages with more repairs. This was likely because they were fulfilling their obligatory turns by responding to questions from STs. Users might also be sensitive of the usual social roles, i.e., that the STs were the "teachers" and they were the "students," and so they adopted a more receptive mode of communication. On the other hand, users produced a wider range of vocabularies when talking to STs compared to their peers, as indicated in the type count and the commonality scores. This suggested that even free conversations with STs may promote the diversity of expressive vocabulary in users. When users conversed with other users in free conversations, they initiated questions more often and produced more tokens than when they conversed with STs. They also used less conventions, indicating a more casual register. This pattern suggested that EasyDial^{TM} may facilitate active and bidirectional pattern of communication with peers (cf. Clarke and Kirton, 2003), in line with professionals' views that smart phone texting with picture symbols and speech can increase independence and participation in users with CCN (Buchholz et al., 2013). In terms of the overall quality of conversations, user–ST conversations were longer than user–user conversations with more messages and turns per conversation. However, reciprocity of conversation partners as indexed by turn-taking behavior did not differ between user–user and user–ST conversations.

In sum, this study reported linguistic analyses of graphic symbol-based AAC usage in a sample of users with CCN using Cantonese Chinese, with similarities and differences with AAC in other languages. We found that users with CCN had different usage patterns when conversing with peers and STs, suggesting sensitivity to communication partners or conversation topics, but both contexts could be valuable to their social communication. A limitation in this study is that we did not explore the effects of face-to-face vs. remote messaging, although this could yield differences in the choice of symbols and the contextual understanding. Further studies could address this question in light of the social distancing measures, which could restrict in-person communications. Although the usage data in the current report are limited in size compared to typical language corpora because of the nature of AAC, anonymized AAC usage data are continually accumulated as EasyDial^{TM} is used in the local community. Availability of these data in the future could allow for further cross-linguistic comparisons between AAC in Chinese and other languages. Our data can be used to deduce the communicative needs of the diverse and understudied population of people with CCN. Results will inform future enhancement of EasyDial^{TM} as well as other AAC systems, thereby improving service provision and ultimately equal access and social inclusion.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by SAHK. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YY contributed to the analysis of the results and to the writing of the manuscript. SS and RC designed the EasyDial^{TM} system and performed the user study. All authors contributed to the article and approved the submitted version.

FUNDING

Research work reported in this article was supported by Innovation and Technology Support Programme (ITSP) Public Sector Trial Scheme (PSTS) (Ref No: ITT/010/18GP) of the Hong Kong SAR Government.

ACKNOWLEDGMENTS

We thank all participants for their contribution to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.686657/full#supplementary-material

REFERENCES

Abe, M. S., and Otake-Matsuura, M. (2021). Scaling laws in natural conversations among elderly people. *PloS One* 16:e0246884. doi: 10.1371/journal.pone.0246884

An, S., Feng, X., Dai, Y., Bo, H., Wang, X., Li, M., et al. (2017). Development and evaluation of a speech-generating AAC mobile app for minimally verbal children with autism spectrum disorder in Mainland China. *Mol. Autism* 8:52.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Baixeries, J., Elvevåg, B., and Ferrer-i-Cancho, R. (2013). The evolution of the exponent of Zipf's law in language ontogeny. *PloS One* 8:e53227. doi: 10.1371/journal.pone.0053227

Bernardi, L., and Tuzzi, A. (2011). Analyzing written communication in AAC contexts: a statistical perspective. *Augment. Altern. Commun.* 27, 183–194. doi: 10.3109/07434618.2011.610353

Binger, C., and Light, J. (2008). The morphology and syntax of individuals who use AAC: research review and implications for effective practice. *Augment. Altern. Commun.* 24, 123–138. doi: 10.1080/07434610701830587

Boenisch, J., and Soto, G. (2015). The oral core vocabulary of typically developing English-speaking school-aged children: implications for AAC practice. *Augment. Altern. Commun.* 31, 77–84. doi: 10.3109/07434618.2014.1001521

Buchholz, M., Müller, I. M., and Ferm, U. (2013). Text messaging with pictures and speech synthesis for adolescents and adults with cognitive and communicative disabilities–professionals' views about user satisfaction and participation. *Technol. Disabil.* 25, 87–98. doi: 10.3233/tad-130370

Chan, R. Y. Y. (2014). "Cloud augmentative and alternative communication for people with complex communication needs," in *Proceedings of the 2014 IEEE Global Communications Conference*, (Austin, TX: Institute of Electrical and Electronics Engineers), 2727–2732.

Chan, R. Y. Y., Ding, J., Kong, L. W., Yan, G., Bai, X., Ma, X., et al. (2016). "Making telecommunications services accessible to people with severe communication disabilities," in *Proceedings of the 2016 IEEE Global Humanitarian Technology Conference (GHTC)*, (Piscataway, NY: Institute of Electrical and Electronics Engineers), 105–112.

Chen, M. C., Hill, K., and Yao, T. (2009). "Preliminary vocabulary frequency findings for mandarin Chinese AAC treatments," in *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*, (New York, NY: ACM digital library), 1–4.

Chen, S. H. K., Hill, K., Ridenour, T. R., Sun, K., Su, C., and Chen, M. C. (2015). Feasibility study of short-term, intensive augmentative and alternative communication intervention with mandarin Chinese speaking children with autism. *J. Speech Lang. Hear. Assoc. Taiwan* 34, 87–108.

Clarke, M., and Kirton, A. (2003). Patterns of interaction between children with physical disabilities using augmentative and alternative communication systems and their peers. *Child Lang. Teach. Ther.* 19, 135–151. doi: 10.1191/0265659003ct248oa

Deckers, S. R., Van Zaalen, Y., Van Balkom, H., and Verhoeven, L. (2017). Core vocabulary of young children with down syndrome. *Augment. Altern. Commun.* 33, 77–86. doi: 10.1080/07434618.2017.1293730

Friginal, E., Pearson, P., Di Ferrante, L., Pickering, L., and Bruce, C. (2013). Linguistic characteristics of AAC discourse in the workplace. *Discourse Stud.* 15, 279–298. doi: 10.1177/1461445613480586

Friginal, E., Pickering, L., and Bruce, C. (2016). "Narrative and informational dimensions of AAC discourse in the workplace," in *Talking at Work*, eds L. Pickering, E. Friginal, and S. Staples (London: Palgrave MacMillan), 27–54. doi: 10.1057/978-1-137-49616-4_2

Gallucci, M. (2019). *GAMLj: General Analyses for Linear Models. [Jamovi Module]*. Available online at: https://gamlj.github.io/ (accessed March 26, 2021).

Jamovi (2020). *Jamovi. (Version 1.2) [Computer Software]*. Available online at: https://www.jamovi.org (accessed March 26, 2021).

Kent-Walsh, J., Murza, K. A., Malani, M. D., and Binger, C. (2015). Effects of communication partner instruction on the communication of individuals using AAC: a meta-analysis. *Augment. Altern. Commun.* 31, 271–284. doi: 10.3109/07434618.2015.1052153

Light, J. (1988). Interaction involving individuals using augmentative and alternative communication systems: state of the art and future directions. *Augment. Altern. Commun.* 4, 66–82. doi: 10.1080/07434618812331274657

Light, J. (1997). "Let's go star fishing": reflections on the contexts of language learning for children who use aided AAC. *Augment. Altern. Commun.* 13, 158–171. doi: 10.1080/07434619712331277978

Liu, C., and Sloane, Z. (2006). Developing a core vocabulary for a mandarin Chinese AAC system using word frequency data. *Int. J. Comput. Process. Orient. Lang.* 19, 285–300. doi: 10.1142/s0219427906001530

Lund, S. K., and Light, J. (2007). Long-term outcomes for individuals who use augmentative and alternative communication: part II–communicative interaction. *Augment. Altern. Commun.* 23, 1–15. doi: 10.1080/07434610600720442

McCarthy, J. H., Schwarz, I., and Ashworth, M. (2017). The availability and accessibility of basic concept vocabulary in AAC software: a preliminary study. *Augment. Altern. Commun.* 33, 131–138. doi: 10.1080/07434618.2017.1332685

McNaughton, D., and Light, J. (2013). The iPad and mobile technology revolution: benefits and challenges for individuals who require augmentative and alternative communication. *Augment. Altern. Commun.* 29, 107–116. doi: 10.3109/07434618.2013.784930

Mühlenbock, K. H., and Lundälv, M. (2011). "Using lexical and corpus resources for augmenting the AAC-lexicon," in *Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies*, (Edinburgh: Association for Computational Linguistics), 120–127.

Pinto, M., and Gardner, H. (2014). Communicative interaction between a non-speaking child with cerebral palsy and her mother using an iPadTM. *Child Lang. Teach. Ther.* 30, 207–220. doi: 10.1177/0265659013518338

R Core Team (2019). *R: A Language and Environment for Statistical Computing. (Version 3.6) [Computer Software]*. Available online at: https://cran.r-project.org/ (accessed March 26, 2021).

Smith, M. (2005). The dual challenges of aided communication and adolescence. *Augment. Altern. Commun.* 21, 67–79. doi: 10.1080/10428190400006625

Smith, M. M. (2015). Language development of individuals who require aided communication: Reflections on state of the science and future research directions. *Augment. Altern. Commun.* 31, 215–233. doi: 10.3109/07434618.2015.1062553

Soto, G., and Clarke, M. T. (2017). Effects of a conversation-based intervention on the linguistic skills of children with motor speech disorders who use augmentative and alternative communication. *J. Speech Lang. Hear. Res.* 60, 1980–1998. doi: 10.1044/2016_jslhr-l-15-0246

Soto, G., and Clarke, M. T. (2018). Conversation-based intervention for adolescents using augmentative and alternative communication. *Augment. Altern. Commun.* 34, 180–193. doi: 10.1080/07434618.2018.1490926

Conflict of Interest: The authors declare that this study was partially supported by the Centre for Perceptual and Interactive Intelligence (CPII) Ltd. under the Innovation and Technology Fund. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yum, So and Chan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

REVIEW
published: 06 September 2021
doi: 10.3389/fcomm.2021.698044

Neuroscientific Research on the Processing of Easy Language

Liv Borghardt, Silvana Deilen, Julia Fuchs, Anne-Kathrin Gros, Silvia Hansen-Schirra, Arne Nagels, Laura Schiffl and Johanna Sommer*

Gutenberg Council for Young Researchers (GYR), Easy Language Research Group, Johannes Gutenberg University Mainz, Mainz, Germany

In an increasingly networked world, people who cannot participate in written communication are particularly at risk of social exclusion. Like other concepts of barrier-free communication, Easy Language ("Leichte Sprache") attempts to enhance perceptibility and comprehensibility of texts for people with reading impairments by means of a rule-based reduction of complexity on the text side. The underlying assumption of simplified languages is that the reduction of linguistic complexity correspondingly reduces cognitive processing costs. However, this interplay between linguistic complexity and cognitive processing costs still needs to be investigated by empirical research as up to date there are only a few studies investigating the perception and reception of Easy Language, mostly using offline methods (e.g., questionnaires or retrospective interviews). In contrast to offline methods, which are only capable of assessing comprehension products, online methods allow researchers to track what a participant is focusing their attention on at any given time and to thereby develop a detailed representation of the time-course of cognitive language processing. In our paper, we aim to point out how different online methods (eye-tracking, EEG and fMRI) can be used for investigating the empirical validity of the postulated rules for Easy Language by evaluating cognitive processing efficiency. Besides discussing the applicability of those neuroscientific online methods in Easy Language research, we discuss the importance of collecting personal and neuropsychological data to gain detailed profiles of the participants and therefore not only contribute to the explanation of variance but furthermore to determine the role of neuropsychological skills on reading proficiency. For each online method we elaborate basic principles, discuss some of the main findings in cognitive sciences and demonstrate the greatest advantages but also restrictions of the method and challenges related to the data collection process with impaired participants. Furthermore, we outline current challenges in Easy Language research and summarize remaining research gaps. On a final note, we emphasize that it is both the establishment of online methods and the data triangulation in Easy Language research that enable researchers to gain a profound insight into the cognitive processing of simplified languages.

Keywords: Easy Language, simplified languages, written language processing, cognitive processing costs, neurolinguistic methods, communicative impairments

OPEN ACCESS

Edited by:
Sigal Uziel-Karl,
Achva Academic College, Israel

Reviewed by:
Xavier Moonen,
University of Amsterdam, Netherlands
Anna Matamala,
Universitat Autònoma de Barcelona, Spain

****Correspondence:***
Julia Fuchs
fuchsj@uni-mainz.de

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

Received: *20 April 2021*
Accepted: *23 August 2021*
Published: *06 September 2021*

Citation:
Borghardt L, Deilen S, Fuchs J, Gros A-K, Hansen-Schirra S, Nagels A, Schiffl L and Sommer J (2021) Neuroscientific Research on the Processing of Easy Language. Front. Commun. 6:698044. doi: 10.3389/fcomm.2021.698044

INTRODUCTION

Easy Language is conceived as a variety of a language that is reduced systematically on different linguistic levels, for example on the lexical and syntactic level (cf. Maaß, 2015). On a continuum with different comprehensibility levels represented by specific language forms, Easy Language is characterized by a maximum comprehensibility level and constitutes one extreme pole of the continuum. The other extreme pole, the elaborate level, is represented by languages for special purposes. Plain Language and standard language are located between these two extreme poles, plain language having an intermediary comprehensibility level and standard language being the standard level (cf. Maaß, 2020: 51). Easy Language has initially been developed for individuals with intellectual and learning disabilities who have difficulties with understanding regular texts written in standard language. Today, however, the target groups are significantly larger, comprising also individuals with dementia, prelingual hearing impairment and aphasia as well as functional illiterates and people with German as a second language (Bredel and Maaß, 2016a: 140–172).

A major challenge Easy Language research is currently facing consists in the empirical investigation of the effectiveness of the different rules postulated by the guidelines for Easy Language, for example, with regard to German Easy Language, by the early practical guidelines (Inclusion Europe, 2009;[1] Netzwerk Leichte Sprache, 2009;[2] BITV 2.0, 2011[3]) and by the scientifically founded guidelines (Maaß, 2015; Bredel and Maaß, 2016a; Bredel and Maaß, 2016b; cf. Maaß, 2020: 69–87). It is thus necessary to investigate the comprehension of Easy Language. In linguistics and its neighbouring disciplines, the different methods used for the investigation of comprehension can be subdivided into two groups: online methods on the one hand and offline methods on the other. Online methods measure "the processes that come into play in comprehension itself" (Gillioz and Zufferey, 2020: 17); in this context, Christmann (2002) uses the term processes of comprehension. Examples of online methods are eye-tracking, electroencephalography (EEG) or functional magnetic resonance imaging (fMRI). On the other hand, offline methods "affect the final interpretations resulting from the comprehension process" (Gillioz and Zufferey, 2020: 17); for designating the object of offline methods, Christmann (2002) uses the term products of comprehension. Examples of offline methods are questionnaires or recall tasks.

The empirical investigation of the effectiveness of different rules on Easy Language has so far mainly been restricted to the use of offline methods (e.g., Lange, 2019). In contrast, only very few studies have addressed the comprehension of Easy Language using online methods (but: Pappert and Bock, 2019; Fuchs et al. submitted manuscript)[4]. Even though "[t]here are no good or bad measures in experimental linguistics" (Gillioz and Zufferey, 2020: 19), processes of comprehension can only be investigated via online methods — and it is in this field that the most significant current research gaps in Easy Language research are located. As an important basis for addressing these desiderata, the present article provides the relevant background information about online methods allowing to investigate processes of comprehension in the context of Easy Language and its target groups. The article is structured as follows: *Metadata and Test Battery* section shows which metadata and neuropsychological information about the participants should be collected; the following sections focus on the online methods eye-tracking (*Eye-Tracking* section), EEG section and fMRI section and their application in the context of Easy Language research. In the final section, the strengths and problems of online methods in the context of Easy Language are summarized.

METADATA AND TEST BATTERY

Metadata

Since the target groups for Easy Language are heterogeneous within and across groups it is important to collect metadata concerning the participants. Similarly to other psycholinguistic experiments, general data such as age, sex, native language etc. are collected. This information is important because factors such as native language or educational level can impact the level of reading experience which in turn influences reading skills. However, additional data come into play when involving participants with special communication needs. Information on the kind and degree of disability is important in order to differentiate the target groups. Differentiating the target groups is necessary to analyze cognitive factors that impact reading ability. Furthermore, the degree of disability is one of the main factors influencing an individual's level of reading ability.

Comparable to other psycholinguistic experiments and tests, data collection has to be carried out on the basis of informed consent. In the case of persons not capable of giving consent (e.g., children, some of the people with cognitive disability or dementia), their legal representative will be asked to sign the form. In case they are able to give consent, information on the experiment has to be presented in Easy or Plain Language to ensure comprehensibility.

Neuropsychological Testing

As the target group of German Easy Language with its several subgroups is expected to be very heterogeneous concerning not only the recipients' abilities to process written language but also their cognitive performance, neuropsychological skills should be evaluated precisely when investigating language processing with regard to Easy Language.

In the course of variance explanation, the participants' neuropsychological performance is to be regarded as of the same importance as their metadata (as discussed above). Therefore, a neuropsychological test battery should be implemented in each experimental design using online methods to examine cognitive processing. A minimum of the following subtests and their constructs are recommended to be

[1]https://www.lag-abt-niedersachsen.de/uploads/migrate/Download/Infofralle.pdf.
[2]https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf.
[3]http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html.
[4]Fuchs, J., Schaeffer, M., Hansen-Schirra, S. Do adults with and without Intellectual Disabilities Benefit from German Easy Language? Eye-Tracking and Recall Studies on the Processing of Causal and Conditional Relations. Submitted Manuscript

taken into consideration. As the discussed tests are suitable for German native speakers they may need to be adapted to the examined language. Equivalent or similar tests that measure corresponding abilities should be used and are commonly available in neuropsychological assessment batteries for the relevant language.

Psychomotor Ability and Mental Flexibility

Psychomotor ability, speed of processing and mental flexibility can be assessed by the Trail Making Test (TMT-A and TMT-B). The TMT-A tests for psychomotor ability and speed of processing. The widely used diagnostic instrument originates in the Army Individual Test Battery (1944), where it was used for diagnosing attention disorders. The TMT is a paper and pencil test, where participants are asked to connect numbers (1–25) in ascending order by drawing a connective line without lifting the pen off the paper. The second part of the TMT (TMT-B) additionally takes mental flexibility into account as participants are asked to not only connect numbers but also letters in alternating order (numbers 1–13, letters A–L). For both parts, participants' performance is evaluated by the time taken to complete the task. In order to pass the test, participants are required to take no longer than 3 min for each part. In consequence, participants with a slow cognitive processing speed will likely fail the TMT-A and the TMT-B.

Working Memory

As the working memory capacity is an important factor in processing written language (e.g., Ober et al., 2019), it should be taken into investigation for each individual. One possible way to do so is through having participants repeat a series of numbers (Zahlenspanne). This is done in numerous test batteries (e.g., WAIS-R by Wechsler, 1981). In the test "Zahlenspanne" the reproduction of a number series by the participants is queried forwards (2–9 digits) in the first and backwards (2–8 digits) in a second part and allows conclusions about the participants' auditory memory span. The auditory processing abilities are particularly relevant for a successful acquisition of written language (Wise et al., 2010; Pezzino et al., 2019) and should be considered as highly relevant when assessing a neuropsychological profile of the target group. Considering that people with intellectual disability are reported to show a specific deficit in the phonological memory span (Schuchardt et al., 2011), working memory might take on an important role in the reading abilities and should be carefully surveyed. Psycholinguistic studies have found that the working memory span of unimpaired adults has an average size of six to eight items (digits, words or other units) depending on the phonological complexity of these (Miller, 1956; Hulme et al., 1995).

Word Fluency

Word fluency is considered a factor of verbal intelligence. It can be measured through the Regensburger Wortflüssigkeitstest (RWT - "Regensburger Word Fluency Test" by Aschenbrenner et al., 2000). Participants are given a category and are asked to orally list as many words as possible in 1 min. Three categories are considered in the RWT: Semantic (e.g., animals), lexical/phonemic (e.g., words beginning with the letter "p") and mixed semantic (e.g., alternating "fruits" and "sports"). By analyzing participants' ability to generate words, conclusions can be drawn about the size and organization of the mental lexicon as well as their cognitive strategies for successful and fast word retrieval (Whiteside et al., 2016).

Verbal Intelligence

For verbal intelligence the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B - "Multiple-Choice Vocabulary Intelligence Test" by Lehrl, 2005) can be used. Participants are asked to select one existing word within a word selection of non-words. In each line four, non-words are presented with one real word. The number of recognized words, out of the whole of 37 lines, then gives feedback on vocabulary size and verbal intelligence. It should be taken into account though, that because of the MWT-B being a written assignment, participants with low literacy skills might find special challenges solving the test.

A further analysis and correlation of the described neuropsychological testing as well as the metadata with processing data described in the further sections could possibly be used in the attempt of defining determinants for the reading and processing abilities of individuals with communicative impairments. This kind of data triangulation furthermore contributes to explaining the statistical variation among the target groups. Besides the factors described above, Hansen-Schirra and Maaß (2020: 29) suggest testing the target groups' abilities concerning hand-eye coordination, visual processing speed, visual-spatial abilities, general level of intelligence and reading speed. While those abilities can also be relevant, researchers working with Easy Language target groups should consider the duration of the experiments themselves and of the pretesting in order to avoid overstraining the participants mentally and/or physically. It is important to consider that certain challenges can occur when dealing with Easy Language target groups and to carefully select the tests suitable for your specific research question. The whole test battery should not be too extensive. In addition, some of the tests will possibly not be applicable due to the limitations in reading abilities and comprehension abilities or sensory limitations of several participants.

EYE-TRACKING

Eye-tracking is one of the most widely used techniques to investigate cognitive processes in reading (for an overview see e.g., Rayner, 1992; Rayner, 1998).

Most of the currently available eye-trackers are infrared video-based tracking systems mounted beneath or integrated into a display screen. The camera emits infrared light which is directed at the eyes. After entering the retina, a large proportion of the light is reflected and captured by image sensors. Subsequently, an eye gaze analysis software is used to calculate eye movements, gaze direction and fixation points. As it is assumed that the time the participant spends fixating on a word equals the time the word

is being processed (a phenomenon known as "eye-mind assumption," cf. Just and Carpenter, 1980), fixations are used as an indicator of participants' focus of attention. Therefore, the fixation position indicates, at least to a certain extent, which part of the sentence is currently processed. As a longer fixation duration is usually associated with a deeper or more effortful cognitive processing, the time the eyes remain on and return to this position is also indicative of the difficulty in extracting word information. Researchers are consequently able to identify passages, words or even letters that increase cognitive processing costs which in turn allows them to draw inferences about the underlying psychological processes during reading. To ensure that the comprehension process was successful, those drawn inferences from the eye-tracking experiment can then be checked against results from further comprehension tests, such as text-picture-matching or follow-up questions.

Since several studies have shown that eye movements are directly influenced by textual variables — for example, increased linguistic complexity leads to increased fixation duration and decreased saccade length — eye-tracking is especially suited to investigate the empirical validity of the postulated rules for Easy Language. In recent years, more and more researchers are taking advantage of this when evaluating the effectiveness of Easy Language rules.

Three eye-tracking studies are currently being conducted by the research group "Simply complex – Easy Language." Two of them investigate processing costs at the interface of morphology and lexis: Schiffl (2020) focuses on the effects of a target word's frequency, length and number of occurrences in the text, while Deilen (2020) infers differences in the cognitive processing of compounds that are segmented with a hyphen, segmented with a mediopoint or not segmented at all. As both studies involve the target groups of Easy Language — adults with cognitive impairment and students with prelingual hearing impairment – they face similar challenges conducting and carrying out the eye-tracking data acquisition (see Deilen and Schiffl, 2020). Eye-tracking was also used in the study conducted by Fuchs et al. (submitted manuscript) on the processing of causal and conditional relations by adults with and without intellectual disabilities.

Other researchers like Wellmann (2021) and Gutermuth (2020) also have implemented eye-tracking research with one or more of the Easy Language target groups to evaluate reading and comprehension processes. Similarly to Deilen (2020), Wellmann (2021) sets out to answer the question of how different segmentation signs (namely hyphen and mediopoint) affect the processing of compounds in Easy Language. Her study was conducted with representatives of the target group "learners of German as a second language." Gutermuth (2020) investigates the reception and processing difficulties of authentic texts with varying complexity levels (including Plain Language) for people with cognitive impairment, people with migration background and seniors. For her study she combined eye-tracking with tests of comprehensibility and recall.

One of the main advantages of using eye-tracking technology to investigate cognitive processing of Easy Language is that many of the currently available systems are mobile. Since there is no need for participants to leave their familiar surroundings, researchers can conduct their studies at different locations, thus reaching many of the heterogenous target groups of Easy Language. However, researchers investigating cognitive processing within the Easy Language target groups should be aware of several challenges when conducting their experiments. Many times, participants with cognitive impairment (due to mental disabilities, dementia, aphasia or other circumstances) will not be capable of the same requirements as unimpaired participants. To avoid mentally and/or cognitively overstraining the participants, conductors should carefully monitor the experiment's duration. Furthermore, participants with impairment cannot be expected to show the same ability in terms of understanding and acting in accordance to experimental instructions. Instructions should therefore be held rather simple than complicated and tasks should not involve complex steps or combinations (e.g., "if sentence x is true, press button 1; if not, press button 2"). In terms of self-control, participants with cognitive impairment might face difficulties, making it harder to sit quietly in front of a computer screen while restricting head and body movement. Even if instructions are well understood, participants with impairment might need more guidance and reminders in comparison to unimpaired adults. Also, participants without impairment, especially university students, will usually have an easier time dealing with the unnatural situation of (reading) experiments. Lastly, on a more practical level, physical impairments concerning the eyes (e.g., squint, nystagmus, thick glasses) seem to occur more often within the group of intellectually impaired people, making the calibration of the eye-tracking-system more difficult (cf. Splunder et al., 2006; Csakvari and Gyori, 2015).

EEG

The Electroencephalogram (EEG) has been a useful technique in neurophysiological, psycholinguistic and clinical linguistic research, as well as neuropsychological diagnostics. Since the 1920s, researchers measure electric activity on the human scalp by means of active electrodes. The positive or negative signals derive from summed postsynaptic polar shifts from pyramidal cells perpendicular to the cortical surface (cf. Luck, 2014). When a participant perceives auditory or visual stimuli, the neurotransmissions move to the scalp surface. Depending on the study, a differing number of active or passive electrodes pick up the voltage fluctuations. The potentials connected to a specific event are called event related potentials (ERPs) and appear as soon as a participant perceives a stimulus. Therefore, EEG has a high temporal resolution and is highly efficient in research dealing with the time course of stimuli processing.

To measure ERPs, active or passive electrodes are connected to an amplifier from which the signal is sent to a computer (Luck, 2014: 21). Because of the high sensitivity to muscle and eye movements, eye blinks, heart beat and power signal artefacts, eye electrodes are usually placed next to the eyes and experiments

take place in a soundproof cabin. The raw data has to be preprocessed before it can be analyzed. Processing steps include filtering of the frequency bandwidth, decomposition into independent components (ICA) so that muscle and eyeblink artefacts are erased from the signal, and baseline correction. Then, the data can be segmented into the relevant time windows associated with a critical stimulus. The analysis methods for ERPs differ from time-frequency-analysis. The extraction of ERPs happens by averaging trial amplitude values of same experimental conditions for each participant in a first step and grand-averaging amplitudes of conditions across all participants in a second step. The pairwise-comparisons between the conditions across all participants can then be undertaken by a cluster-based permutation test. Here, the amplitude values of the two conditions are compared and significant differences between conditions for each time point in each electrode position can be detected.

Whereas ERPs are sensitive to the time domain, time-frequency-analysis adds a further dimension to the observations. Here, the EEG signal is decomposed in five frequency bands of interest that range from <4 Hz (delta band) up to 80 Hz (gamma band, 30–80 Hz) and analyzed according to an event (event related oscillation, c.f. Tamm, 2005: 8–12).

The extracted ERPs can give insights into a wide range of brain responses towards stimuli. Participants' responses towards uncanonical sentence structures or morpho-syntactic violations (classically reported as a positive voltage shift 600 ms after the stimulus (Osterhout and Holcomb, 1992: 791) can be observed as well as participants' expectancy towards an auditory or visual stimulus. A word's frequency, familiarity or phonological complexity modulates the activation level and therefore is more or less expected in a sentence. These effects have been reported as enhanced negativity after 400 ms post onset an unexpected stimulus (cf. Kutas and Hillyard, 1980a: 103, Kutas and Hillyard, 1980b: 203–205, Kutas and Federmeier, 2011: 622–644 for a review). Over the past decades, a huge psycholinguistic research field has expanded observing components such as the P300, N400 and P600 creating a complex mass of theories on the interaction of context with the incremental integration of words into a sentence on all linguistic levels. By means of the components, the relationship between cognitive processing efficiency and a participant's comprehension of an Easy Language product can be examined (e.g., Van Petten and Luka, 2012 for a review on prediction). For instance, the use of words with higher frequency could be reflected in a decrease of the N400 component. This would imply that the brain required less processing costs in order to process the word on a semantic level. Another effect could be diminished brain activity due to syntactic complexity reduction when readers process Easy Language material. If there was no processing cost decrease when an adult control group had read Easy Language sentences, the syntactic structure of the stimulus should be revised. Increasingly, also time-frequency-analysis has been employed in language processing studies (Prystauka and Lewis, 2019: 5–31 for a review). For instance, the processing of semantic anomalies have been found in increased theta oscillations (3–7 Hz) (e.g., Bastiaansen and Hagoort, 2015: 2,100–2,103, and gamma band oscillation (>30 Hz) (e.g., Hald et al., 2006: 95–98) as well as decrease in beta power (16-19 Hz) (Wang et al., 2012: 2904-2906). Whereas a wide range of questions on interlingual comparisons, clinical research questions and research on second language learners has been observed (e.g., Hahne, 2001; Midgley et al., 2009; Barkley et al., 2015; Almor et al., 2017), to our knowledge no research has been undertaken exploring the processing of reduced or simplified language such as Plain English or German Easy Language using the EEG methodology so far. In Korean, Kang et al. (2017) investigated the influence of intelligence level (high or low) on the processing of visuo-spatial and language tasks with two difficulty levels. Measuring alpha, beta and theta band coherence, the results showed that individuals with different cognitive preconditions processed the stimuli of the two tasks differently (Kang et al., 2017: 51ff.). Even though the study is not a linguistic one on simplified Korean, it reveals that the results obtained in the time-frequency-domain potentially reflect specific linguistic aspects in a text that lead to increased processing costs for specific target groups.

Because of the well-established research on language processing through EEG, the method can serve as a useful tool for modeling complexity stages in processing simplified language. Theoretically, there is no restriction to be included in EEG experiments and all types of stimuli can be presented. Furthermore, the method can be combined e.g., with eye-tracking or fMRI. The extraction of ERPs and the analysis of the time-frequency domain make it possible to draw conclusions on very specific language phenomena and therefore built models on the complexity of language processing for simplified languages as well. When testing members of the target group, insights into interacting processes such as attention, memory and language processes can be observed with EEG. Potentially, predictors for the usefulness of simplified language can be detected.

However, several aspects need to be considered when planning an EEG study with target groups of Easy Language. Participants with cognitive impairments are more likely to have a lower attention span as compared to unimpaired groups and thus might have difficulties following the instructions. Also, movement artefacts are more likely for this group. Reading assessments, neuropsychological tests (see Chapter 2) and additional behavioral measures are crucial for those participants to control for effects associated with cognitive impairments. It is also important to consider possible methodological constraints. When presenting a whole text to participants in reading studies, eye movements during reading cause strong artefacts that distort the ERP recording. Therefore, text stimuli must be presented one word at a time, either in rapid serial visual presentation (RSVP) or self-paced reading mode. However, RSVP cannot be considered a natural reading situation as participants have no control over the input. Additionally, in order to get a better resolution of the ERP components, words are usually presented at slow rates (500–800 ms per word), while in natural reading, readers generally read five words per second. The prolonged duration of the word presentation may cause interpretive processes in the reader that would be reflected in

the ERPs (Camblin et al., 2007: 176–178). Self-paced reading is considered more natural than RSVP and has been employed in ERP studies (e.g. Ditman et al., 2007). Nevertheless, it may cause eye movement artefacts and may increase the probability of interpretive processes. Furthermore, it can be a problematic task for subjects with cognitive disabilities. To bypass this problem, auditory presentation is used in many studies. Some findings show that components are sensitive to input modality, for instance the N400. Holcomb and Neville (1990) found that the N400 occurred earlier and lasted longer with auditory than with written stimuli presentation. Additionally, the scalp distribution is different for the two modalities (Holcomb and Neville, 1990: 296–301). Therefore, it has to be ensured that the participants have sufficient reading skills so that the stimuli can be visually presented.

In general, it has to be considered that parts of the target group might face difficulties taking part in an EEG study. Even though EEG is a non-invasive method, the necessity of sitting in a soundproof booth without being allowed to move makes the situation ethically critical. For some target groups the demands might be bearable, e.g., for prelingually hearing impaired readers (cf. Malaia et al., 2020). However, for people with a high degree of disability, EEG is not feasible, since the length of experiments might be too effortful and the placement of the electrodes on the scalp is physically challenging. Also, the repetition of stimulus material demands a high amount of attention and concentration. These aspects have to be considered when planning studies on language forms for and with a specific target group.

FMRI

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique that uses MRI scanners to investigate changes in brain function. MRI is based on the use of a very strong magnetic field, which energetically excites hydrogen atoms in the body. The energy emitted later can be measured and localized. This allows anatomic structures to be depicted clearly. It does not involve radiation and therefore an MRI-survey is harmless and can be repeated frequently (cf. Huettel et al., 2009: 3–21, 35). In fMRI-experiments, a three-dimensional anatomical image of the head is constructed first using the (structural) MRI. The brain activity detected in the fMRI is transferred to this image as intensity-coded colour areas at a later stage (cf. Huettel et al., 2009: 91–97, 369–372). This determination of brain areas with increased activation is the aim of fMRI. It is a hemodynamic method, which means that changes of the concentration of oxygen in the blood are measured. This measurement is based on the so-called neurovascular coupling, which describes the relationship between the activity of nerve cells and subsequent changes in cerebral blood flow. Since the brain does not provide storage capacities for energy, any energy consumption must be compensated directly. Active neural regions must therefore be provided with energy carriers via the blood flow (cf. Huettel et al., 2009: 165–166, 176–178).

The main energy carrier in all metabolic cycles is oxygen, which is transported through the blood via hemoglobin. Increased activity in a region causes increased blood flow into that region. This changes the distribution of oxygenated and deoxygenated hemoglobin; there is even an excess of oxygen. Deoxygenated hemoglobin is paramagnetic and leads to greater inhomogeneity in the magnetic field (cf. Huettel et al., 2009: 201–211). This is the basis of the Blood Oxygenation Level-Dependent response (short: BOLD response), which associates the respective brain activity with the task performed by the participant of an experiment.

Hence, local changes in the hemodynamic function are measured while a participant performs an experimental task. In order to depict such brain functions, images are constructed that distinguish the activity levels from areas of the brain. These images are based on the functional contrast described above. It is important to note that in all cases, differences between two contrasts with statistically significant differences and no absolute activations are described (cf. Huettel et al., 2009: 11).

FMRI is a comparatively new method and has only been applied since the early 1990s. Nevertheless, it has become well established in linguistic investigations. Especially in the field of semantic and syntactic investigations, several studies have already been conducted. In these studies, the focus was not only on localizing production and processing (e.g., Dapretto and Bookheimer, 1999; Friederici et al., 2000; Friederici et al., 2003; Heim et al., 2003), but also on the differences in the demands on the brain in the processing of simple vs. complex constructions (e.g., Just et al., 1996; Röder et al., 2002; Müller et al., 2003). Yet, there are still many open questions concerning brain functions. Even analyses with the same object of study do not always yield comparable results. It should also be noted that many functions are based on distributed networks (e.g., the Dual-Stream-Model, cf. Hickok and Poeppel, 2007) and that a single brain region is involved in more than one function (e.g., Broca's area, cf. Fiebach et al., 2005). Accordingly, fMRI studies must always be interpreted critically in the sense that activated brain areas could also be attributed to other functions that may be triggered by the stimulus material, e.g., an active button press task involves motor regions or the working memory is activated during sentence processing (the so-called reverse inference, cf. also Huettel et al., 2009: 490–491).

As mentioned, several studies on complexity in language have been published, but no fMRI studies in the field of Easy Language research have been conducted yet. However, since the rules mentioned at the beginning (see Introduction) build on assumptions about cognitive demands and working memory performance, such investigations are necessary, especially on the syntactic level. Although some well-known studies are transferable (e.g., studies of word order such as Röder et al., 2002), further studies, e.g., on maximum sentence complexity and above all based on the Easy Language rules, can advance research. Similar to the components in the EEG (see EEG section), stimuli formulated in Easy Language may cause a decrease in the BOLD response (compared to more complex conditions), leading to the corresponding conclusion that these structures imply a lower demand in processing. However, it is important to note that attention, especially in a control group, can decrease after some

time. In order to keep it high, comprehension questions can be included, which can, for instance, be answered by a button press. In this way, data on the comprehensibility of the presented stimuli is obtained simultaneously.

It should be noted, however, that studies with at least some of the target groups are not feasible. The ethical acceptability is not clear, since an fMRI investigation requires some effort for the participant. Not only the length of the investigation, during which the participant is positioned in the scanner the whole time and is not allowed to move, but also the safety risks resulting from the strong magnetic field are problematic. Even if MRI does not pose any danger to the participants when used correctly, the strong and rapidly changing magnetic fields during scanning result in several safety challenges (cf. Huettel et al., 2009: 44). Hence, before each scan procedure there is an instruction concerning the risks, which must be understood and approved. It must therefore be ensured that this is really understood. Furthermore, the length of an experiment also affects concentration and attention. The shown tasks or stimuli must be presented repeatedly (30–50 reps/condition), which means a high demand on the participant. The type of presentation of the stimuli must also be adapted respectively (see also EEG section on EEG).

Nonetheless, fMRI studies focusing on Easy Language are important for the further development and validation of the existing rules. Experiments with unimpaired participants can also help to investigate current assumptions about the processing of transformed texts and to develop a neurobiologically-feasible model for the processing of Easy Language.

CONCLUSION

A major gap Easy Language research is currently facing consists in the empirical investigation of the effectiveness of the existing rules using online methods. The research community has to enhance its understanding of the comprehension processes triggered by Easy Language in order to be able to decide whether and in what way the existing rules have to be revised or not. In psycho- and neurolinguistics, comprehension processes are investigated via online methods like eye-tracking, EEG and fMRI. At present, first studies using online methods in the context of Easy Language research are being conducted with unimpaired participants and with different target groups. But to the best of our knowledge, these studies are restricted to the investigation of German Easy Language. The present article aimed at giving a first overview over the application of the online methods eye-tracking, EEG and fMRI in the context of Easy Language research and stressed the importance of collecting metadata and neuropsychological information about participants. The strengths of online methods lie in their potential to give insights into the real-time processing of language, in contrast to offline methods such as questionnaires, which are only capable of informing about comprehension products. Mobile systems, mostly eye-trackers, enable researchers to collect data in the field. However, researchers are also facing several challenges when using online methods: due to the special neuropsychological dysfunctions of most target groups, the length of the experiments has to be restricted to a reasonable extent. This can be problematic if several conditions requiring numerous stimuli are of interest. Moreover, researchers have to make sure that the instructions are well understood in particular by the target groups. Any data loss due to the specific dispositions of the target groups, e.g., due to impairments concerning the eyes causing difficulties with the calibration of an eye-tracker, must already be considered in the planning phase of an experiment. It is advisable to recruit more participants from target groups of Easy Language than in usual studies with unimpaired populations.

Researchers are called on to make their experiences with online methods in the context of Easy Language research available to the community so that this kind of research can be established and continuously improved. The present article makes a first step into this direction and is intended to provide (initial) guidance for designing such studies.

AUTHOR CONTRIBUTIONS

SH-S developed the conception and design of the manuscript. LB, SD, JF, A-KG, LS, and JS wrote the manuscript and edited the final version. AN supervised the project and the paper writing. All authors contributed to the paper and the final version of the manuscript has been approved for submission by all authors: they are accountable for the whole work.

FUNDING

This research was funded by the Gutenberg Council for Young Researchers (GYR).

REFERENCES

Almor, A., Nair, V. A., Boiteau, T. W., and Vendemia, J. M. C. (2017). The N400 in Processing Repeated Name and Pronoun Anaphors in Sentences and Discourse. *Brain Lang.* 173, 52–66. doi:10.1016/j.bandl.2017.06.003

Army Individual Test Battery (1944). *Manual of Directions and Scoring.* Washington DC: War Department, Adjutant General's Office.

Aschenbrenner, S., Tucha, O., and Lange, K. (2000). *Regensburger Wortflüssigkeits-Test (RWT). Göttingen, Bern.* Toronto, Seattle: Hogrefe, Verlag für Psychologie.

Barkley, C., Kluender, R., and Kutas, M. (2015). Referential Processing in the Human Brain: An Event-Related Potential (ERP) Study. *Brain Res.* 1629, 143–159. doi:10.1016/j.brainres.2015.09.017

Bastiaansen, M., and Hagoort, P. (2015). Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. *J. Cogn. Neurosci.* 27 (11), 2095–2107. doi:10.1162/jocn_a_00829

BITV 2.0 (2011). Barrierefreie-Informationstechnik-Verordnung, Anlage 2, Teil 2. Web Available at: http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html. (Accessed April 12, 2021).

Bock, B. M. (2019). *"Leichte Sprache" – Kein Regelwerk. Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeiSA-Projekt (Kommunikation – Partizipation – Inklusion 5).* Berlin: Frank & Timme.

Bredel, U., and Maaß, C. (2016a). *Leichte Sprache. Theoretische Grundlagen. Orientierung für die Praxis (Sprache im Blick)*. Berlin: Dudenverlag.

Bredel, U., and Maaß, C. (2016b). *Ratgeber Leichte Sprache. Die wichtigsten Regeln und Empfehlungen für die Praxis (Sprache im Blick)*. Berlin: Dudenverlag.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., and Swaab, T. Y. (2007). Processing New and Repeated Names: Effects of Coreference on Repetition Priming with Speech and Fast RSVP. *Brain Res.* 1146, 172–184. doi:10.1016/j.brainres.2006.07.033

Christmann, U. (2002). Methoden der Verstehens- und Verständlichkeitserhebung. *Z. Literaturwiss Linguistik* 32 (4), 76–97. doi:10.1007/BF03379326

Csákvári, J., and Gyori, M. (2015). Applicability of Standard Eye-Tracking Technique in People with Intellectual Disability: Methodological Conclusions from a Series of Studies. *Stud. Health Technol. Inform.* 217, 63–70. Print.

Dapretto, M., and Bookheimer, S. Y. (1999). Form and Content. *Neuron* 24, 427–432. doi:10.1016/s0896-6273(00)80855-7

Deilen, S. (2020). "Easy Language Research : Text and User Perspectives," in *Easy Language Research: Text and User Perspectives*. Editors S. Hansen-Schirra and C. Maaß (Berlin: Frank & Timme), 241–256. doi:10.26530/20.500.12657/42088

Deilen, S., and Schiffl, L. (2020). "Easy Language Research : Text and User Perspectives," in *Easy Language Research: Text and User Perspectives*. Editors S. Hansen-Schirra and C. Maaß (Berlin: Frank & Timme), 273–281. doi:10.26530/20.500.12657/42088

Ditman, T., Holcomb, P. J., and Kuperberg, G. R. (2007). An Investigation of Concurrent ERP and Self-Paced reading Methodologies. *Psychophysiology* 44, 927–935. doi:10.1111/j.1469-8986.2007.00593.x

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., and Friederici, A. D. (2005). Revisiting the Role of Broca's Area in Sentence Processing: Syntactic Integration versus Syntactic Working Memory. *Hum. Brain Mapp.* 24, 79–91. doi:10.1002/hbm.20070

Friederici, A. D., Meyer, M., and Cramon, D. Y. v. (2000). Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain Lang.* 75, 465–477. doi:10.1006/brln.2000.2438

Friederici, A. D., Rüschemeyer, S.-A., Fiebach, C. J., and Hahne, A. (2003). The Role of Left Inferior Frontal and superior Temporal Cortex in Sentence Comprehension: Localizing Syntactic and Semantic Processes. *Cereb. Cortex* 13, 170–177. doi:10.1093/cercor/13.2.170

Gillioz, C., and Zufferey, S. (2020). *Introduction to Experimental Linguistics (Cognitive Science Series)*. Hoboken, NJ: John Wiley & Sons.

Gutermuth, S. (2020). *Eine zielgruppenorientierte Rezeptionsstudie zu Leichter und Einfacher Sprache*. Berlin: Frank & Timme.One for All?.

Hahne, A. (2001). What's Different in Second-Language Processing? Evidence from Event-Related Brain Potentials. *J. Psycholinguistic Res.* 30, 251–266. doi:10.1023/A:1010490917575

Hald, L. A., Bastiaansen, M. C. M., and Hagoort, P. (2006). EEG Theta and Gamma Responses to Semantic Violations in Online Sentence Processing. *Brain Lang.* 96 (1), 90–105. doi:10.1016/j.bandl.2005.06.007

Hansen-Schirra, S., and Maaß, C. (2020). "Easy Language Research : Text and User Perspectives," in *Easy Language Research: Text and User Perspectives*. Editors S. Hansen-Schirra and C. Maaß (Berlin: Frank & Timme), 17–38. doi:10.26530/20.500.12657/42088

Heim, S., Opitz, B., Müller, K., and Friederici, A. D. (2003). Phonological Processing during Language Production: fMRI Evidence for a Shared Production-Comprehension Network. *Cogn. Brain Res.* 16, 285–296. doi:10.1016/S0926-6410(02)00284-7

Hickok, G., and Poeppel, D. (2007). The Cortical Organization of Speech Processing. *Nat. Rev. Neurosci.* 8 (5), 393–402. doi:10.1038/nrn2113

Holcomb, P. J., and Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-Related Brain Potentials. *Lang. Cogn. Process.* 5, 281–312. doi:10.1080/01690969008407065

Huettel, S. A., Song, A. W., and McCarthy, G. (2009). *Functional Magnetic Resonance Imaging*. Sunderland, Massachusetts: Sinauer Associates.

Hulme, C., Roodenrys, S., Brown, G., and Mercer, R. (1995). The Role of Long-Term Memory Mechanisms in Memory Span. *Br. J. Psychol.* 86 (4), 527–536. doi:10.1111/j.2044-8295.1995.tb02570.x

Inclusion Europe (2009). Informationen für alle. Europäische Regeln, wie man Informationen leicht lesbar und verständlich macht. Available at: https://www.lag-abt-niedersachsen.de/uploads/migrate/Download/Infofralle.pdf (Accessed April 10, 2021).

Just, M. A., and Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychol. Rev.* 87 (4), 329–354. doi:10.1037/0033-295x.87.4.329

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996). Brain Activation Modulated by Sentence Comprehension. *Science* 274, 114–116. doi:10.1126/science.274.5284.114

Kang, J.-S., Ojha, A., Lee, G., and Lee, M. (2017). Difference in Brain Activation Patterns of Individuals with High and Low Intelligence in Linguistic and Visuo-Spatial Tasks: An EEG Study. *Intelligence* 61, 47–55. doi:10.1016/j.intell.2017.01.002

Kutas, M., and Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980a). Event-related Brain Potentials to Semantically Inappropriate and Surprisingly Large Words. *Biol. Psychol.* 11, 99–116. doi:10.1016/0301-0511(80)90046-0

Kutas, M., and Hillyard, S. (1980b). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science* 207, 4427203–4427205. doi:10.1126/science.7350657

Lange, D. (2019). Der Genitiv in der "Leichten Sprache" - das Für und Wider aus theoretischer und empirischer Sicht. *Z. für Angew. Linguistik* 70 (1), 37–72. doi:10.1515/zfal-2019-2001

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B). 5. Unveränderte Auflage*. Balingen: Spitta Verlag.

Luck, S. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA, United States: MIT Press.

Maaß, C. (2020). *Balancing Comprehensibility and Acceptability (Easy – Plain – Accessible 3)*. Berlin: Frank & Timme. doi:10.26530/20.500.12657/42089Easy Language - Plain Language - Easy Language Plus : Balancing Comprehensibility and Acceptability

Maaß, C. (2015). *Leichte Sprache. Das Regelbuch (Barrierefreie Kommunikation 1)*. Münster: LIT.

Malaia, E. A., Krebs, J., Roehm, D., and Wilbur, R. B. (2020). Age of Acquisition Effects Differ across Linguistic Domains in Sign Language: EEG Evidence. *Brain Lang.* 200, 104708. doi:10.1016/j.bandl.2019.104708

Midgley, K. J., Holcomb, P. J., and Grainger, J. (2009). Language Effects in Second Language Learners and Proficient Bilinguals Investigated with Event-Related Potentials. *J. Neurolinguist.* 22 (3), 281–300. doi:10.1016/j.jneuroling.2008.08.001

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychol. Rev.* 63 (2), 81–97. doi:10.1037/h0043158

Müller, R.-A., Kleinhans, N., and Courchesne, E. (2003). Linguistic Theory and Neuroimaging Evidence: an fMRI Study of Broca's Area in Lexical Semantics. *Neuropsychologia* 41, 1199–1207. doi:10.1016/S0028-3932(03)00045-9

Netzwerk Leichte Sprache (2009). Die Regeln für Leichte Sprache. Available at: https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf (Accessed April 12, 2021).

Ober, T. M., Brooks, P. J., Plass, J. L., and Homer, B. D. (2019). Distinguishing Direct and Indirect Effects of Executive Functions on Reading Comprehension in Adolescents. *Reading Psychol.* 40, 551–581. doi:10.1080/02702711.2019.1635239

Osterhout, L., and Holcomb, P. J. (1992). Event-Related Brain Potentials Elicited by Syntactic Anomaly. *J. Mem. Lang.* 31, 785–806. doi:10.1016/0749-596X(92)90039-Z

Pezzino, A.-S., Marec-Breton, N., and Lacroix, A. (2019). Acquisition of Reading and Intellectual Development Disorder. *J. Psycholinguist Res.* 48 (3), 569–600. doi:10.1007/s10936-018-9620-5

Prystauka, Y., and Lewis, A. G. (2019). The Power of Neural Oscillations to Inform Sentence Comprehension: A Linguistic Perspective. *Lang. Linguist Compass* 13, 9. doi:10.1111/lnc3.12347

Rayner, K. (1992). *Eye Movements and Visual Cognition: Scene Perception and reading*. New York: Springer-Verlag.

Rayner, K. (1998). Eye Movements in reading and Information Processing: 20 Years of Research. *Psychol. Bull.* 124, 372–422. doi:10.1037/0033-2909.124.3.372

Röder, B., Stock, O., Neville, H., Bien, S., and Rösler, F. (2002). Brain Activation Modulated by the Comprehension of normal and Pseudo-word Sentences of Different Processing Demands: A Functional Magnetic Resonance Imaging Study. *NeuroImage* 15, 1003–1014. doi:10.1006/nimg.2001.1026

Schiffl, L. (2020). "Easy Language Research : Text and User Perspectives," in *Easy Language Research: Text and User Perspectives*. Editors S. Hansen-Schirra and C. Maaß (Berlin: Frank & Timme), 227–239. doi:10.26530/20.500.12657/42088

Schuchardt, K., Maehler, C., and Hasselhorn, M. (2011). Functional Deficits in Phonological Working Memory in Children with Intellectual Disabilities. *Res. Dev. disabilities* 32 (5), 1934–1940. doi:10.1016/j.ridd.2011.03.022

Tamm, Sascha. (2005). *Hochaufgelöste Zeit-Frequenz-Analysen Ereigniskorrelierter EEG-Oszillationen Mittels S-Transformation*. dissertation. Berlin: Freie Universität. doi:10.17169/refubium-11564

Van Petten, C., and Luka, B. J. (2012). Prediction during Language Comprehension: Benefits, Costs, and ERP Components. *Int. J. Psychophysiology* 83, 176–190. doi:10.1016/j.ijpsycho.2011.09.015

van Splunder, J., Stilma, J. S., Bernsen, R. M. D., and Evenhuis, H. M. (2006). Prevalence of Visual Impairment in Adults with Intellectual Disabilities in the Netherlands: Cross-Sectional Study. *Eye* 20, 1004–1010. doi:10.1038/sj.eye.6702059

Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., et al. (2012). Beta Oscillations Relate to the N400m during Language Comprehension. *Hum. Brain Mapp.* 33, 2898–2912. doi:10.1002/hbm.21410

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Wellmann, K. (2021). "Medio·punkt mit Binde-Strich? Eine Eyetracking-Studie," in *Leichte Sprache – Empirische und multimodale Perspektiven*. Editors A. Gros, S. Gutermuth, and K. Oster (Berlin: Frank & Timme), 23–42.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., et al. (2016). Verbal Fluency: Language or Executive Function Measure? *Appl. Neuropsychol. Adult* 23, 29–34. doi:10.1080/23279095.2015.1004574

Wise, J. C., Sevcik, R. A., Romski, M., and Morris, R. D. (2010). The Relationship between Phonological Processing Skills and Word and Nonword Identification Performance in Children with Mild Intellectual Disabilities. *Res. Dev. disabilities* 31 (6), 1170–1175. doi:10.1016/j.ridd.2010.08.004

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Borghardt, Deilen, Fuchs, Gros, Hansen-Schirra, Nagels, Schiffl and Sommer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 26 October 2021
doi: 10.3389/fpsyg.2021.703690

A Comparative Study of Educational Texts for Native, Foreign, and Bilingual Young Speakers of Russian: Are Simplified Texts Equally Simple?

Anna Dmitrieva[1,2], *Antonina Laposhina*[1]* *and Maria Lebedeva*[1]

[1] Language and Cognition Laboratory, Pushkin State Russian Language Institute, Moscow, Russia, [2] Faculty of Arts, University of Helsinki, Helsinki, Finland

Studies on simple language and simplification are often based on datasets of texts, either for children or learners of a second language. In both cases, these texts represent an example of simple language, but simplification likely involves different strategies. As such, this data may not be entirely homogeneous in terms of text simplicity. This study investigates linguistic properties and specific simplification strategies used in Russian texts for primary school children with different language backgrounds and levels of language proficiency. To explore the structure and variability of simple texts for young readers of different age groups, we have trained models for multiclass and binary classification. The models were based on quantitative features of texts. Subsequently, we evaluated the simplification strategies applied to readers of the same age with different linguistic backgrounds. This study is particularly relevant for the Russian language material, where the concept of easy and plain language has not been sufficiently investigated. The study revealed that the three types of texts cannot easily be distinguished from each other by judging the performance of multiclass models based on various quantitative features. Therefore, it can be said that texts of all types exhibit a similar level of accessibility to young readers. In contrast, binary classification tasks demonstrated better results, especially in the R-native vs. non R-native track (with 0.78 F1-score), these results may indicate that the strategies used for adapting or creating texts for each type of audience are different.

Keywords: simple language, simple Russian, young readers, simplification strategies, textbook analysis, textbook corpus, text simplification, Russian language

OPEN ACCESS

Edited by:
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

Reviewed by:
Leina Zhu,
University of Houston, United States
Tatiana Petrova,
Saint Petersburg State University, Russia

***Correspondence:**
Antonina Laposhina
anlaposhina@pushkin.institute

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

Received: *30 April 2021*
Accepted: *23 September 2021*
Published: *26 October 2021*

Citation:
Dmitrieva A, Laposhina A and Lebedeva M (2021) A Comparative Study of Educational Texts for Native, Foreign, and Bilingual Young Speakers of Russian: Are Simplified Texts Equally Simple? Front. Psychol. 12:703690. doi: 10.3389/fpsyg.2021.703690

INTRODUCTION

Modern data-based research on simple language and simplification is in critical need of sufficiently representative and reliable data—that is, texts that are samples of simple language. For the Russian language, this need is particularly acute. On the one hand, the concept of simple, easy, and plain Russian language has not been sufficiently investigated and is in its formative stages (Mustajoki et al., 2021). On the other hand, research on textual complexity in Russian is still in search of parameters that predict the complexity of comprehension more reliably and precisely than readability formulas (Laposhina, 2017; Solovyev et al., 2018). Psychophysiological studies of reading in the elementary school age confirm the influence of various factors, such as the frequency of words included in the text or discourse parameters of the text, on text comprehension, but at the moment

these parameters are not considered in standard readability formulas for Russian (Petrova, 2016; Korneev et al., 2018, 2019). Finally, in the field of automatic simplification, a promising and actively pursued task in natural language processing, the problem of scarce data resources for simplification in Russian is noted (Sakhovskiy et al., 2021).

Any simple language research needs a relevant collection of samples of simple or simplified language. Such samples, on the one hand, are texts intended for developing readers. For example, Brouwers et al. (2014) employed educational materials such as encyclopedic entries simplified for children to study the strategies of sentence simplification in informative and narrative texts. In Gala et al. (2020), literary and scientific texts, along with their simplified versions, were used to create a parallel corpus for French learners who struggle with reading. During the creation of the Newsela corpus (Xu et al., 2015), the same texts were simplified for children at four different school grade levels to create a high-quality dataset for text simplification. All of these corpora can be used for the creation and/or evaluation of automatic text simplification systems. As for the Russian language, the linguistic complexity of texts for children was studied on educational materials for Russian-speaking students at primary school (Laposhina et al., 2019) and secondary school (Solovyev et al., 2018; Vakhrusheva et al., 2021) and the collection of book previews labelled with one of two categories—children's or adult (Glazkova et al., 2021).

On the other hand, most of the research on simplification is based on texts that were created or adapted for adult foreign language learners. According to Crossley et al. (2011), simplified L2 reading texts are either adapted from authentic texts or written explicitly for the L2 reader. The authors of textbooks are guided by educational standards and regulations, methodological experience and intuition, and non-formalized textual ideas that are simple enough to understand and affordable for non-native language learners. Such materials are used for studying the properties and text comprehension of simplified texts (Crossley et al., 2014) or in creating and testing simplification systems (Arfé et al., 2014). For the Russian language, texts for L2 learners were used for building systems of automatic complexity estimation (Karpov et al., 2014; Laposhina et al., 2018), refining objective parameters of text complexity (Solovyev et al., 2019), and studying L2 adaptation strategies (Sibirtseva and Karpov, 2014; Dmitrieva et al., 2021).

At the intersection of these two categories of simple texts are educational texts created for young L2 learners. Such texts constitute a separate category of simple texts, which are under-researched; usually, simplification studies are based either on texts for children or on texts for L2 learners.

Moreover, on the figurative scale of language proficiency, another category of children stands out—namely children with unbalanced bilingualism/multilingualism, including heritage speakers. In studies of Russian language acquisition and Russian language teaching practice, this category of children is identified specifically (Kagan and Dillon, 2003; Polinsky and Kagan, 2007; Protassova, 2008; Kalenkova and Zhiltsova, 2018; Moskovkin, 2019), and educational and assessment materials for such children are created and labelled separately from standard Russian young speakers, on the one hand, and from young L2 learners, on the other (Lebedeva et al., 2021). However, the specifics of texts written specifically for this category of children, and how they differ in complexity from texts for their peers with a different level of language proficiency, have not yet been studied.

Thus, the focus of our study is on three groups of texts for children with different language proficiency in Russian. A detailed study on the arrangement and simplicity of such data is of significant importance for studying simplification strategies, and it may contribute to both research of text complexity and the field of language teaching.

RESEARCH QUESTIONS

The study aims to explore the simple Russian language presented in texts for children with different levels of language proficiency. Herein, we determine which simplification strategies are used to create simple texts for different groups of readers.

This study hypothesizes that the target group of simplification (children or second language learners) determines simplification strategies, so that simple texts for different groups of readers are modified differently.

In this study, we test the hypothesis on educational texts for children with the different settings of the Russian language acquisition such as follows: primary school children with Russian as a native language (hereafter R-native), their peers with Russian as a weaker language in unbalanced bilingualism (hereafter R-bilingual), and children who study Russian as a foreign language outside the Russian language environment (hereafter R-foreign).

Accordingly, this study aims to answer the following research question:

> Are there any specific simplification strategies in educational texts for children with different language backgrounds and levels of Russian language proficiency?

MATERIALS AND METHODS

Corpus Building

To answer the aforementioned research questions, we employed Text-Image Russian Textbook Corpus (TIRTEC) of texts from Russian language textbooks for children aged 7–11 years (corresponding to the age of primary school students in the Russian education system), intended for three groups of children based on their language proficiency and settings of language acquisition: R-native, R-bilingual, and R-foreign[1]. We followed the existing division of texts into the three target groups and relied on the methodological description of the target audience of the textbook indicated by the authors in the book annotation (e.g., "for bilingual 10-year-olds learners Russian at weekend schools").

Table 1 shows the volume and basic text characteristics of the three groups of texts randomly chosen from the TIRTEC corpus for the following experiment. Each group contains the same

[1]Textbook names and references are available at https://digitalpushkin.tilda.ws/tirtec.

TABLE 1 | Characteristics of the three subdomain of texts randomly chosen from the TIRTEC corpus for the following experiment.

	R-foreign	R-bilingual	R-native
Collection size			
Number of texts	1,100	1,100	1,100
Number of tokens	39,955	58,964	31,670
Vocabulary size (number of unique tokens)	8,846	12,760	10,919
Text source			
Simple fragment of authentic text	170	205	727
Fragment of authentic text adapted by textbook authors	61	41	30
Texts written specifically for this textbook	869	854	343
Basic text characteristics			
Mean sentence length (words)	5.84	7.3	7.56
Mean word length	4.67	4.86	5.14
Average number of punctuations per sentence	0.73	0.81	1.02

number of texts, 1,100, so that the classes were balanced for future experiments. The texts for the R-bilingual group contain the maximum number of words and many unique words, whereas the least number of words is found in the R-native texts. This is due to the peculiarities of the Russian school system, in which the Reading course has separate textbooks that were not included in the TIRTEC corpus, while textbooks for R-bilingual and R-foreign combine linguistic exercises and reading in one book.

Each domain includes texts from the different sources: fragments of authentic text (e.g., written by Pushkin A.); fragments of authentic text adapted by textbook authors (e.g., based on "The Tale of the Fisherman and the Fish" by Pushkin A.); and texts written specifically for this textbook. However, the proportion of these types differs among these three groups, which can also be an illustration of different strategies for simple text selection.

In terms of language proficiency, these groups should form an ascending scale of language users, from beginners (R-foreign) to proficient (R-native), according to their age; R-bilingual children are expected to occupy a middle position. This is confirmed by the average word and sentence length, and the average number of punctuation symbols per sentence.

Text Preprocessing

First, texts from coursebooks were digitized and annotated with meta-attributes manually. Before extracting text features for feature-based models, we cleaned the texts of noisy symbols and non-standard punctuation (for example, we replaced "?." with "?"). Before extracting some features, such as coverage by different word lists, we also lemmatized the texts with the Mystem 3.1 toolkit for Python (Segalovich, 2003). Sentence tokenization was performed with ru_punkt[2], an NLTK sentence tokenizer for Russian.

[2]https://github.com/Mottl/ru_punkt

Features Extraction

We identified a **set of quantitative features** that determine the difficulty level of the text, building on relevant research on automated readability assessment (Karpov et al., 2014; Reynolds, 2016; Laposhina et al., 2018, Sharoff et al., 2008). Our current study makes use of 95 features which can be divided into four groups.

1. **Length-based features** of texts are presented by average word and sentence length and the ratio of words longer than four syllables.
2. **Readability formulas.** We implement the 5 often used in modern Russian readability studies formulas:
   - Flesch–Kincaid readability tests
   - The Coleman–Liau index
   - Dale–Chall readability formula
   - Automated Readability Index(ARI)
   - Simple Measure of Gobbledygook (SMOG)

Almost all of them represent various combinations of mean word length in signs or syllables, sentence length, and constant coefficients.

3. **Lexical features** include:
   - Coverage by vocabulary lists for the learners Russian as a foreign language graded by the Common European Framework of Reference for Languages (CEFR) levels (Andryshina and Kozlova, 2012, 2015; Andryshina, 2017a,b). Since there are currently no such lists specifically for children, we used the version for adult learners of Russian
   - Coverage by frequency lists of Modern Russian Frequency Dictionary (Lyashevskaya and Sharov, 2009)
   - Coverage by the list of abstract words
   - Type/token ratio (TTR) is the ratio of different unique word stems (types) to the total number of words (tokens) that indicate lexical diversity in the text
   - Lexical density is calculated as the ratio of lexical items to the total number of words.
4. **Morphosyntactic features** represent the relative ratio of tokens with given morphosyntactic tags, so observed frequencies of POS tags were divided by the total amount of words in the text (e.g., the number of NOUN-tags divided by the total number of tokens), counts of cases were divided by the number of words that have cases. We used 50 morphosyntactic tags in total, e.g., percentage of nouns, prepositions, conjunctions, words in the genitive case, and the number of passive forms.

Features from groups 1, 3, 4 were extracted using Python programming language and the Mystem 3.1 toolkit. Readability formulas with constants optimized for Russian texts were taken from I. Begtin's Plain Russian project[3].

[3]https://github.com/infoculture/plainrussian

TABLE 2 | F1-scores in binary feature based classifiers.

Classifier	F1-score
R-native vs. non-R-native	0.78
R-foreign vs. non-R-foreign	0.72
R-bilingual vs. non-R-bilingual	0.68
R-native adult vs. R-native kids	0.9

Model

To study the possible relations between various features of texts and their domains, we employed both multiclass and binary classification, using Python and the scikit-learn library (Pedregosa et al., 2012) to build our models[4]. Scikit-learn allows for simple and efficient data analysis with the help of many built-in tools such as various statistical models. For the multiclass setup, we used multinomial logistic regression with a limited-memory Broyden-Fletcher-Goldfarb-Shannon (BFGS) solver ("LBFGS") for optimization. For binary tasks, we employed logistic regression with default parameters. We also scaled all features between 0 and 1 during preprocessing when working with our text metrics.

Model Testing

To test the adequacy of the model and extracted features, we trained a binary model on two groups of texts with obvious differences in simplicity and comprehensibility: texts from our corpus for children of primary school age vs. fragments from fiction books included in the high school curriculum (such as War and Peace by L. Tolstoy, and Oblomov by Goncharov), with a similar total word count. Both models showed high performance in the classification tasks: model showed an F_1-score of 90 (see **Table 2**). This demonstrates that the selected sets of features can distinguish texts by difficulty level, and also confirms the general presupposition that the texts we have selected actually are the examples of a simple language.

RESULTS

As can be seen in **Table 1**, one of the parameters by which the three groups differ is the source of the educational text. While simple fragments in R-native textbooks are mostly taken from children's and classical literature, many of the R-foreign and R-bilingual texts are written specifically for educational purposes. Authentic texts are most often presented in these books in small folklore genres: songs, riddles, and proverbs.

Regression Models

To estimate the homogeneity of texts within the three selected groups in terms of their linguistic features, we performed both multiclass and binary classifications based on text features described in section Features Extraction.

The multiclass model task was to predict a right target audience for the given text—R-native, R-foreign, or R-bilingual.

[4]The source code is available on https://github.com/Digital-Pushkin-Lab/SimplifiedRussianTextsForChildrenClassif.

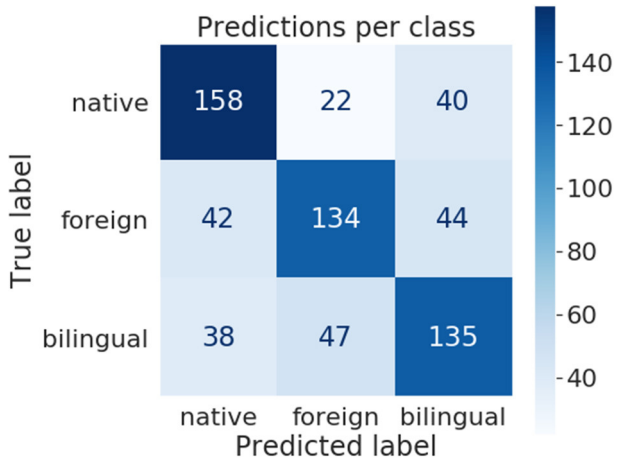

FIGURE 1 | Confusion matrix of multiclass model predictions based on set of text features.

It performed best on R-native texts with an F_1-score of 69, and worst on R-bilingual, with an F_1-score of 62, but these results are not satisfactory enough. As can be seen in **Figure 1**, most of the time, the model confused R-bilingual texts for R-foreign and *vice versa*. The classification task of predicting the target audience for educational text written for young learners of Russian proved to be difficult.

However, transforming the task into **several binary classifiers** improved accuracy. We performed a series of binary comparisons of these collections. In these one vs. the rest setups we tried to train the models to distinguish one particular class from the rest of the texts: for example, R-native texts from non-R-native (R-foreign and R-bilingual texts combined) texts. The numbers of instances in classes 1 and 0 were equal.

The best results were observed in the R-native vs. non-R-native comparison with a 78 F_1-score for class 0 and 77 for class 1 (see **Table 2**). The ROC curves for one-vs-rest setups were again best in the R-native classifier with an AUC score of 0.85, and worst in the R-bilingual classifier with an AUC score of 0.73.

The model error analysis shows that for all types of errors, the median value of the percentage of words from lexical minima turns out to be closer to the median value not of its correct category, but of the one determined by the model. For instance, texts that were marked R-bilingual by the model while actually being R-native tend to contain more vocabulary from the CEFR-graded lexical minima than R-native texts contain on average. And in texts marked R-foreign instead of R-native these numbers were even higher. This can indicate that lexical differences were one of the factors that confused the model. Readability proved to be among such factors as well. Some grammatical features, such as relative numbers of adjectives, nouns, verbs and adverbs among all words, also influenced wrong decisions of the model. For example, in R-native texts the relative number of adjectives is quite high on average. However, in R-native texts that were wrongly identified as R-bilingual this number is lower, and in R-native texts marked as R-foreign there were almost no adjectives at all. Finally, it is worth noticing that the model made more

TABLE 3 | Selected Kendall's tau correlations between the dependent variable (class) and various independent variables (features).

Domain	Most significant features	Kendall's τ
R-foreign	Relative numbers of verbs in past tense	−0.28
	Percentage of A1 vocabulary	0.28
	Percentage of A2 vocabulary	0.26
	Coleman's readability formula	−0.25
	Percentage of B1 vocabulary	0.25
	Relative numbers of verbs in perfective aspect	−0.24
R-bilingual	Number of unique words	0.19
	Number of words	0.19
	Text coverage by 5,000 most frequent Russian words list	0.18
	Relative amount of nouns	−0.13
	Lexical density	−0.12
	TTR	−0.12
R-native	Percentage of A1 vocabulary	−0.39
	Percentage of B2 vocabulary	−0.39
	Percentage of B1 vocabulary	−0.38
	TTR	0.33
	Coleman's readability formula	0.26
	ARI readability formula	0.25

P-value of all counts <0.05.

errors on texts from certain textbooks, which may indicate that these texts do not correspond to the proclaimed target audience. It is especially true for the most diffuse category, R-bilingual.

Correlations and Means

To analyze the effect of each text feature for the texts discrimination into three groups, we examined correlations on our data using Kendall's tau coefficient. This non-parametric test does not rely on assumptions about variable distributions. We assumed the text features to be independent variables, and the class of text (R-native, R-foreign, or R-bilingual) to be the dependent variable.

The results of hypothesis testing are shown in **Table 3**. We tested the correlations on binary problems; for example, the correlations of the features in the R-foreign section are calculated using a binary dependent variable, where 1 is R-foreign texts and 0 is non-R-foreign texts (the numbers of entries in each class are balanced).

It should be admitted that we did not observe any particularly strong correlations here. However, we note some peculiarities that may be associated with different strategies for creating and adapting education materials for these groups of children. First, the coefficients among these groups are different: the highest coefficients are observed in the R-native group, and the lowest one—in the R-bilingual, which may indicate the heterogeneity of this group. The top lines of the table for groups R-foreign and R-native are occupied by features based on lexical minima for adult L2 Russian learners. It may be a signal of a difference in understanding of simple basic vocabulary among these groups. R-native textbooks contain texts from Russian classical literature, prose about nature as well as children literature—this leads to the presence in books of specific vocabulary about nature and agriculture (e.g., оляпка "white-throated dipper," осина "aspen," элеамор "grain elevator"). At the same time, the text materials for the R-foreign group are more guided by designated lexical minima for L2 learners, which contain more everyday vocabulary. Meanwhile, in R-bilingual texts features based on lexical minima did not play a significant role. However, other lexical indicators came to the fore, such as words from frequency lists, lexical density, and lexical diversity.

A negative correlation between the number of verbs in the past tense and the R-foreign group (the more such verbs in the text, the less likely it is that the text belongs to the R-foreign group) may be due to the simplicity in grammatical forms: foreign students start using verbs from the present and future tense forms. The relative numbers of verbs in the perfective aspect, which do not have present tense forms in Russian, also speak in favor of this hypothesis. It can also be caused by the fact that textbooks for foreigners have a large number of examples of everyday communicative situations, in contrast to fiction texts for R-native, which is often turned to the past.

DISCUSSION AND CONCLUSION

The study revealed that the three types of texts cannot easily be distinguished from each other by judging the performance of multiclass models. Therefore, it can be said that texts of all types exhibit a similar level of accessibility to young readers. In contrast, the feature-based approach proved to be effective at binary tasks, especially in the R-native vs. R-foreign tracks. These results indicated that the strategies used for adapting or creating texts for each type of audience are different, which makes some groups of texts easier to distinguish. For instance, in R-foreign texts, more standardized words are used, and conversely, in R-native texts the vocabulary is richer, and more advanced grammar is used. The considerable difference between the R-native domain and the others can also be explained by the number of authentic texts in this part of the corpus, as opposed to the R-foreign and R-bilingual domains, in which texts written specifically for textbooks are common. Judging by the correlation analysis, it seems that texts intended for R-foreign learners contain fewer verbs in past tense forms, which may indicate different notions about the grammatical side of the text complexity. The most informative lexical features for R-native and R-foreign groups were those based on lexical minima for adult L2 Russian learners. This suggests that authors of educational texts for foreign children are largely guided by the requirements of the CEFR level system, although these requirements have not been accommodated to children studying Russian. The text coverage by lexical minima of R-bilingual text is higher than the R-native group, even considering that, for example, R-bilingual texts are longer on average. The R-bilingual group showed a low connection with the linguistic parameters of the text in the binary classification task (R-bilingual vs. not R-bilingual). Therefore, we can assume that this group is the most diverse, combining different strategies and views on text simplification.

The complexity and heterogeneity of this group of texts create significant limitations for the use of these materials as data for simplification outside the field of research on heritage speakers and bilinguals.

The experiments described above were limited to examining the differences between the three domains of the simplified Russian language. In future studies, it would be interesting to investigate the change in the comprehensibility level inside these domains, for example, from one school grade to another, and to observe whether the language of educational texts reflects a crucial restructuring in reading patterns that occurs around the third grade (Korneev et al., 2019).

Overall, the study found that the three observed domains can be ordered on a scale from the simplest (R-foreign with simpler grammar and standardized vocabulary) to the most complex (R-native with a richer vocabulary and more complicated grammar). Despite the fact that in the practice of Russian teaching educational materials for bilinguals are distinguished as a separate category, the quantitative linguistic analysis showed that the status of R-bilingual texts is ambiguous and they are the least classified area. The results of this linguistic study contribute to various areas of research on simple Russian and suggest directions for further research, including psychophysiological research aimed at studying which text parameters constitute complexity for different categories of young readers with different levels of Russian language proficiency.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://digitalpushkin.tilda.ws/tirtec.

AUTHOR CONTRIBUTIONS

AD: literature review, conducting preprocessing, classification model building and evaluation, data analysis, and writing the paper. AL: data collection and annotation, data analysis, interpretation of results, and writing the paper. ML: conception and design of the study, formulation of research goals and aims, literature review, interpretation of results, writing the paper, overall management, and coordination of the study. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been supported by the RFBR Grant No.17-29-09156.

ACKNOWLEDGMENTS

The authors thank their colleagues and students of the Pushkin State Russian Language Institute for their contribution to the corpus collection and annotation.

REFERENCES

Andryshina, N. P. (ed.). (2017a). *Lexical Minimum of Russian as a Foreign Language. Level B1. Common language (9th Edn.)*. St. Petersburg: Zlatoust.

Andryshina, N. P. (ed.). (2017b). *Lexical Minimum of Russian as a foreign language. Level B2. Common language (7th Edn.)*. St. Petersburg: Zlatoust.

Andryshina, N. P., and Kozlova, T. V. (2012). *Lexical Minimum of Russian as a Foreign Language. Level A1. Common Language, 4th Edn.* St. Petersburg: Zlatoust.

Andryshina, N. P., and Kozlova, T. V. (2015). *Lexical Minimum of Russian as a Foreign Language. Level A2. Common Language, 5th Edn.* St. Petersburg: Zlatoust.

Arfé, B., Oakhill, J., and Pianta, E. (2014). "The text simplification in TERENCE," in *Methodologies and Intelligent Systems for Technology Enhanced Learning*. Cham: Springer, 165–172.

Brouwers, L., Bernhard, D., Ligozat, A. L., and François, T. (2014). "Syntactic sentence simplification for French," in *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, 47–56. doi: 10.3115/v1/W14-1206

Crossley, S. A., Allen, D. B., and McNamara, D. S. (2011). Text readability and intuitive simplification: a comparison of readability formulas. *Read. Foreign Lang.* 23, 84–101.

Crossley, S. A., Yang, H. S., and McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Read. Foreign Lang.* 26, 92–113.

Dmitrieva, A., Laposhina, A., and Lebedeva, M. (2021). "A quantitative study of simplification strategies in adapted texts for L2 learners of Russian," in *Proceedings of the International Conference "Dialogue 2021"* (Moscow), 191–203.

Gala, N., Tack, A., Javourey-Drevet, L., François, T., and Ziegler, J. C. (2020). "Alector: a parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers," in *Language Resources and Evaluation for Language Technologies (LREC)*, 1353–1361

Glazkova, A., Egorov, Y., and Glazkov, M. (2021). "A comparative study of feature types for age-based text classification," in *Analysis of Images, Social Networks and Texts* (Cham: Springer International Publishing), 120–134.

Kagan, O., and Dillon, K. (2003). A new perspective on the teaching russian: focus on the heritage learner. *Heritage Lang. J.* 1, 76–90. doi: 10.46538/hlj.1.1.5

Kalenkova, O. N. and Zhiltsova, L. Y. (2018). Principy Formirovaniya I Analiz Obrazovatel'nyh Programm Po Russkomu Yazyku Dlya Detej-bilingvov, Prozhivayushchih Za Rubezhom [Principles of the Formation and Analysis of Educational Programs In Russian Language for Bilingual Children Living Abroad]. *Russian Lang. Abroad* 2, 75–81.

Karpov, N., Baranova, J., and Vitugin, F. (2014). "Single-sentence readability prediction in Russian," in *Proceedings of Analysis of Images, Social Networks, and Texts Conference (AIST)*, 91–100.

Korneev, A. A., Akhutina, T. V., and Matveeva, E. Y. (2019). Reading in third graders with different state of the skill: an eye-tracking study. *Moscow Univ. Psychol. Bull.* 2, 64–87. doi: 10.11621/vsp.2019.02.64

Korneev, A. A., Matveeva, E., and Akhutina, T. V. (2018). Chto my mozhem skazat' o formirovanii chteniya na osnove analiza dvizheniya glaz? [What can we say about the formation of reading based on the analysis of eye movement?]. *Human Physiol.* 44, 75–83. doi: 10.1134/S036211971802010X

Laposhina, A. N. (2017). "Analiz relevantnyh priznakov dlya avtomaticheskogo opredeleniya slozhnosti russkogo teksta kak inostrannogo (Analysis of relevant characteristics for automatic determination of the complexity of russian text as a foreign language)," in *Proceedings of the International Conference "Dialogue 2017", Student Session* (Moscow).

Laposhina, A. N., Veselovskaya, T. S., Lebedeva, M. U., and Kupreshchenko, O. F. (2019). "Lexical analysis of the russian language textbooks for primary school: corpus study," in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019"*, 351–363.

Laposhina, A. N., Veselovskaya, T. S., Lebedeva, M. Y., and Kupreshchenko, O. F. (2018). "Automated text readability assessment for Russian second language learners," in *Proceedings of the International Conference "Dialogue 2018"* (Moscow), 396–406.

Lebedeva, M., Veselovskaya, T., Kupreshchenko, O., and Laposhina, A. (2021). "Corpus-based evaluation of textbook content: a case of Russian language primary school textbooks for migrants," in *Facing Diversity in Child Foreign Language Education. Second Language Learning and Teaching*, eds. J. Rokita-Jaśkow, A. Wolanin (Cham: Springer). doi: 10.1007/978-3-030-66022-2_12

Lyashevskaya, O. N., and Sharov, S. A. (2009). *Chastotnyj slovar' sovremennogo russkogo yazyka (na materialah Nacionalnogo korpusa russkogo yazyka) [Modern Russian frequency dictionary (based on the data from the Russian National Corpus)]*. Moscow: Azbukovnik.

Moskovkin, L. V. (2019). "Russkij yazyk kak rodnoj, nerodnoj i inostrannyj: istoriya stanovleniya i sovremennoe upotreblenie terminov (Russian language as a native, non-native and foreign language: history of formation and modern use of terms)" in *Uchitelyu zarubezhnoj shkoly. Sbornik nauchno-metodicheskih statej [To the Teacher of a School Abroad. A Collection of Scientific and Methodological Articles.]*. St. Petersburg: Zlatoust, 97–102.

Mustajoki, A., Mihienko, Z., Nechaeva, N., Kairova, E., and Dmitrieva, A. (2021). "Easy language in Russia," in *Handbook of Easy Languages in Europe*, eds. C. Lindholm and U. Vanhatalo (Berlin: Frank and Timme), 439–467.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., and Louppe, G. (2012). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Petrova, T. (2016). "Eye movements in reading the texts of different functional styles: evidence from Russian," in *Intelligent Decision Technologies 2016. Smart Innovation, Systems and Technologies, vol 57* (Cham: Springer). doi: 10.1007/978-3-319-39627-9_25

Polinsky, M., and Kagan, O. (2007). Heritage languages: in the "wild" and in the classroom. *Lang. Ling. Compass* 1, 368–395. doi: 10.1111/j.1749-818x.2007.00022.x

Protassova, E. (2008). Teaching Russian as a heritage language in Finland. *Heritage Lang. J.* 6, 127–152.

Reynolds, R. (2016). "Insights from Russian second language readability classification: complexity-dependent training requirements, and feature evaluation of multiple categories," in *Proceedings of the 11th Workshop on the Innovative Use of NLP for Building Educational Applications, San Diego*, 289–300.

Sakhovskiy, A., Izhevskaya, A., Pestova, A., Tutubalina, E., Malykh, V., Smurov, I., et al. (2021). "RuSimpleSentEval-2021 shared task: evaluating sentence simplification for Russian," in *Proceedings of the International Conference "Dialogue 2021"* (Moscow), 607–617.

Segalovich, I. (2003). "A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine," in *Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications. MLMTA'03* (Las Vegas, Nevada), 273–280

Sharoff, S., Kurella, S., and Hartley, A. (2008). "Seeking needles in the web's haystack: finding texts suitable for language learners," in *Proceedings of the 8th Teaching and Language Corpora Conference, (TaLC-8)* (Lisbon, Portugal).

Sibirtseva, V. G., and Karpov, N. V. (2014). Automatic adaptation of the texts for electronic textbooks. Problems and perspectives (on an example of Russian). [Avtomaticheskaya adaptaciya tekstov dlya elektronnyh uchebnikov. Problemy i perspektivy (na primere russkogo yazyka)]. *Nová rusistika*. VII, 19–33.

Solovyev, V., Ivanov, V., and Solnyshkina, M. (2018). Assessment of reading difficulty levels in Russian academic texts: approaches and metrics. *J. Intell. Fuzzy Syst.* 34, 3049–3058. doi: 10.3233/JIFS-169489

Solovyev, V., Solnyshkina, M., Ivanov, V., and Batyrshin, I. (2019). Prediction of reading difficulty in Russian academic texts. *J. Intell. Fuzzy Syst.* 36, 4553–4563. doi: 10.3233/JIFS-179007

Vakhrusheva, A. Y., Solnyshkina, M. I., Kupriyanov, R. V., Gafiyatova, E. V., and Klimagina, I. O. (2021). Linguistic complexity of academic texts. *Issues J. Educ. Ling.* 40, 89–99. doi: 10.52575/2712-7451-2021-40-1-89-99

Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: new data can help. *Trans. Assoc. Comput. Ling.* 3, 283–297. doi: 10.1162/tacl_a_00139

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dmitrieva, Laposhina and Lebedeva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 October 2021
doi: 10.3389/fpsyg.2021.690593

Creole Prosodic Systems Are Areal, Not Simple

*Kofi Yakpo**

Department of Linguistics, The University of Hong Kong, Hong Kong, Hong Kong SAR, China

This study refutes the common idea that tone gets simplified or eliminated in creoles and contact languages. Speakers of African tone languages imposed tone systems on all Afro-European creoles spoken in the tone-dominant linguistic ecologies of Africa and the colonial Americas. African speakers of tone languages also imposed tone systems on the colonial varieties of English, French, Spanish, and Portuguese spoken in tonal Africa. A crucial mechanism involved in the emergence of the tone systems of creoles and colonial varieties is stress-to-tone mapping. A typological comparison with African non-creole languages shows that creole tone systems are no simpler than African non-creole tone systems. Demographic, linguistic, and social changes in an ecology can lead to switches from tone to stress systems and vice versa. As a result, there is an areal continuum of tone systems roughly coterminous with the presence of tone in the east (Africa) and stress in the west (Americas). Transitional systems combining features of tone and stress converge on the areal buffer zone of the Caribbean. The prosodic systems of creoles and European colonial varieties undergo regular processes of contact, typological change and areal convergence. None of these are specific to creoles. So far, creoles and colonial varieties have not featured in work on the world-wide areal clustering of prosodic systems. This study therefore aims to contribute to a broader perspective on prosodic contact beyond the narrow confines of the creole simplicity debate.

Keywords: simplification, prosodic system, stress, tone, creole, linguistic ecology, language contact, areal convergence

OPEN ACCESS

Edited by:
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

Reviewed by:
Umberto Ansaldo,
Head of School of Media, Creative Arts, and Social Inquiry, Curtin University, Australia
Anouschka Foltz,
University of Graz, Austria

***Correspondence:**
Kofi Yakpo
kofi@hku.hk

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

Received: *03 April 2021*
Accepted: *06 September 2021*
Published: *27 October 2021*

Citation:
Yakpo K (2021) Creole Prosodic Systems Are Areal, Not Simple. Front. Psychol. 12:690593. doi: 10.3389/fpsyg.2021.690593

1. INTRODUCTION

Creolization is said to involve the simplification of input structures (for a thematic overview, see Ansaldo et al., 2007). One such structure is tone, which has been argued to constitute a feature that gets lost or is starkly reduced during language contact and creolization (e.g., Heine, 1978, 220; Salmons, 1992; Sebba, 1997, 49; McWhorter, 1998, 793; Kusters, 2003, 343; Trudgill, 2010, 309; Sessarego, 2020, 3). I propose that creolization has neither led to the elimination nor the simplification of tone systems.

Instead, creoles feature prosodic systems ranging from tone to stress, and to mixed systems incorporating both. The same holds for European colonial varieties (varieties of English, French, Spanish, and Portuguese spoken in Africa and the Americas). These are generally left out of discussions about creole prosody but are essential for developing a general typology of prosodic contact outcomes.

The prosodic systems of creoles and colonial varieties (contact prosodic systems) have developed tone or stress systems in accordance with the linguistic factor of areal typology (dominance of tone vs. stress in the ecology), the cognitive factor of psycholinguistic dominance (recipient vs. source-language agentivity), and social factors in their specific linguistic ecologies (the demographic

proportion and social stratification of speakers of tone and stress-only languages) (Bordal Steien and Yakpo, 2020; Yakpo, 2020).

Further, there is an east-west, tone-stress continuum from Africa to the Americas. Contact prosodic systems in African ecologies and in isolated ecologies of the Americas (e.g., in the Amazonian region of Suriname) feature tone systems. In this, they reflect the prosodic proclivities of their adstrates (African languages presently spoken by the multilingual speakers of the creoles) and substrates (African languages once spoken alongside the creoles). The prosodic systems of most creoles and colonial varieties spoken in the Americas have, in turn, converged toward the stress-only systems of their European lexifiers (lexicon-providing languages) and superstrates (socially dominant languages, whether lexifier or not) but still maintain marginal tonal features (for detailed creolist definitions of adstrate, substrate, lexifier, and superstrate, see Yakpo, 2017a,b, 53, 227–229).

Afro-European creoles and colonial varieties therefore constitute no exception to the world-wide tendency of prosodic systems to cluster areally (Matisoff, 2001; Gussenhoven, 2004, 42–45; Clements and Rialland, 2007, 74). So far, creoles and colonial varieties have not featured in studies on the areal clustering of prosodic systems (e.g., Maddieson, 2013). This study therefore aims to contribute to a broader perspective on prosodic contact beyond the narrow confines of creole linguistics. The results of this study also complement and support the stratal-areal contact model proposed in earlier work (Yakpo, 2017a), which explains long-term contact outcomes in creoles spoken in the multilingual linguistic ecologies of Africa and the Americas.

Arguments that creole grammars and prosodic systems are simpler than those of non-creoles are based on the concept of 'bit complexity' (DeGraff, 2001, 284–285), which comes down to a simplistic measure of 'more overt material = more complex'. Even from the perspective of bit complexity, creole tone systems are *more complex* than those of the colonial varieties of English, French, Spanish, and Portuguese (section 3.1 and section 4). This is due to social factors that impede the same amount of innovation and areal diffusion of tonal features to the colonial varieties as to the creoles (section 6).

I identify three cognitive-typological mechanisms that drive the creation of contact prosodic systems in the encounter of tonal substrates and adstrates, and lexifiers and superstrates that make use of stress. These are stress-to-tone mapping, paradigmatization, and idiosyncratization. Neither the mechanisms themselves, nor their outcomes involve simplification. Instead, contact prosodic systems acquire their properties from 'typological matching' (Mufwene, 1996, 2001; Aboh and Ansaldo, 2007) between the features of the input languages in a specific linguistic ecology. Crucially, the acoustic and phonological realization of tone in the adstrates and substrates is matched with, and where compatible, grafted on the corresponding realization of stress in the lexifier.

A few additional definitions are in order before proceeding. In languages with stress, words or phrases are associated with metrical structure that is determined with respect to the position of a stressed syllable meeting two criteria. The first is obligatoriness: Every word or phrase has *at least* one syllable marked for primary stress, the highest degree of metrical prominence. The second criterion is culminativity: Every word or phrase has *no more than* one syllable marked for the highest degree of metrical prominence (Hyman, 2006, 231). The acoustic correlates of stress are language-specific but usually involve a combination of the cues of length, loudness, vowel quality, and pitch variations over the stress-bearing syllable. The pitch contours of utterances in languages that are non-tonal are composed of intonational pitch accents anchored to stressed syllables, and boundary tones associated with the edges of phrases and utterances. I henceforth refer to languages with stress, but no tone (some tone languages also have stress) to 'stress-only' languages. All the lexifiers and superstrates of the Afro-European creoles and colonial varieties (English, French, Spanish, Portuguese, and Dutch) are exclusively stress-only languages.

In languages with tone, pitch features are instead part of (at least some) morphemes and therefore part of the lexical realization of morphemes together with vowels and consonants (Welmers, 1959, 2; Hyman, 2001, 1367). The lexical tones of morphemes can therefore only be changed by specific rules akin to those which alter the segmental realization of morphemes. All the substrates and adstrates of the Afro-European creoles and colonial varieties treated here are tone languages.

Pitch is nevertheless employed in both stress-only and tone systems for the non-lexical purposes of intonation, that is, for marking phrase boundaries and expressing pragmatic functions like emphasis, focus, and asking questions (e.g., Downing and Rialland, 2016). However, stress-only systems use pitch distinctively *only* in these phrasal ways.

A concept that is equally important for the ensuing discussion is Van Coetsem's psycholinguistic metaphor of 'agentivity' (1988, 2000). The language with agentivity provides features to the contact language or variety. In the case of recipient language agentivity, features from a non-native source language are transferred to the speaker's dominant or native recipient language by *borrowing*. Source language agentivity represents the opposite case. The speaker uses a non-native recipient language, and features from her dominant or native source language are transferred to the recipient language by *imposition* (termed 'interference' in earlier work, see Weinreich, 1953; Thomason and Kaufman, 1988). Source language agentivity therefore manifests itself as substratal (in cases of shift from the source language to the recipient language) or adstratal (in cases of maintenance of the source and recipient languages) areal influence on the recipient language.

The Afro-European contact scenarios treated here all constitute cases of source language agentivity. The importance of this cannot be stressed enough. All sources that claim tone loss and reduction in language contact and creolization fail to make the distinction between the two transfer types of recipient language and source language agentivity. They therefore fail to identify the directionality of change in contact prosodic systems, thereby jumping to the logically flawed conclusion that stress trumps tone (section 5).

The paper is organized as follows. I first present an analysis of the contact prosodic systems of Pichi and Guyanese Creole

which occupy different spaces on the continuum of contact outcomes between African tone systems and European stress-only systems (section 2). I then provide evidence for an areal distribution of the prosodic systems of Afro-European creoles and colonial varieties along an east to west, Africa to the Americas axis (section 3). This distribution roughly corresponds to the presence of tone in Africa and stress in the Americas, with transitional prosodic systems in the Caribbean, the areal buffer zone between the two. I then identify three concrete mechanisms with potential for generalization, which were involved in the emergence of contact prosodic systems with tone (section 4). Finally, I compare features of tonal Afro-European creoles, which have been claimed to constitute simplification, with those of tonal non-creole languages in Africa (section 5). I show that there is no evidence for simplification in the tone systems of Afro-European creoles. The study is concluded with some general remarks on the role of social factors in the differential outcomes of prosodic contact (section 6).

2. AFRO-EUROPEAN CREOLES CAN HAVE TONE, STRESS, AND A MIX OF BOTH

Two case studies follow of the prosodic systems of the English-lexifier creoles Pichi (Equatorial Guinea) (section 2.1) and Guyanese Creole (Guyana) (section 2.2). The two languages belong to the linguistic family of Afro-Caribbean English-lexifier Creoles with shared ancestry in a (number of) 17th century protolanguage(s) in the Caribbean and West Africa (Hancock, 1986, 1987; Smith, 1987). Pichi and Guyanese Creole occupy different sections of an areal continuum of contact prosodic systems across the Afro-Atlantic (see section 3). Pichi has a tone system and Guyanese Creole has a mixed system featuring tone and stress.

2.1. A Creole With Tone: Pichi

The English-lexifier creole Pichi is spoken on the island of Bioko, Equatorial Guinea. A detailed description of the tone system of Pichi including acoustic evidence is provided by Yakpo (2019a, 37–60). The following sections summarize relevant aspects of the system.

2.1.1. Tones, Tone Patterns, and Minimal Pairs

Pichi has an 'equipollent' (Hyman, 2011a) 2-tone system with a High (H) and a Low (L) tone. This means that /H/ and /L/ are both lexically specified and phonologically activated, and subjected to tonal rules and processes (see section 2.1.2). There is no acoustic evidence for stress.

The Pichi prosodic lexicon is etymologically layered due to the mechanism of stress-to-tone mapping that converted English stress to tone (section 4.1). Most English-sourced words, which constitute the majority of the lexicon, feature an obligatory (at least one) and cumulative (at most one) H tone. The most frequent patterns are an /H/ over a monosyllabic word (54% of my corpus) e.g., *áks* 'ask', an /H-L/ sequence over a disyllabic word (23%), e.g., *húmàn* 'woman', and an /L-H/ sequence over a disyllabic word (14%), e.g., *grèví* 'gravy'. More diverse patterns without the restriction of obligatory and cumulative /H/ are present in a few English-sourced words, e.g., *ápás* /H-H/ 'after' (<Eng. *half past*), and African-sourced words, e.g., *nyɔ́ní* /H-H/ 'ant' (<Mende *yɔ́ní*), *òkóbó* /L-H-H/ 'impotent man' (<Yoruba *òkóbó*). A further pattern consists of an /L/ tone in English-sourced monosyllables whose etymons normally remain unstressed, e.g., *bìn* /L/ 'past tense marker' (<*been*), *dì* /L/ 'definite article' (<*the*), or African-sourced monosyllables with the same specification for /L/ tone, e.g., *nà* /L/ 'general locative preposition' (<Igbo *nà* 'general locative preposition'; also found as a reflex of Proto-Niger-Congo **na* in hundreds of African substrate and adstrate languages).

Ideophones and interjections of African or unknown origin feature more diverse word-tone patterns, often due to lexicalized duplication and triplication, e.g., *ékìé* /H-L-H/ 'expression of counter-expectation', *kɔ́ngkɔ́ngkɔ́ng* /H-H-H/ 'requesting entry', *ményéményé* /H-H-H-H/ 'whine in a childlike fashion', *gbògbògbò* /L-L-L/ 'hastily', *kàmúkàmú* /L-H-L-H/ 'sight of buttocks moving', and *súkútúpàmpà* /H-H-H-L-L/ 'in a cheap and mean fashion'.

Some monosyllabic roots are distinguished from each other by tone alone, see **Table 1**. In conformity with a general pattern, function words tend to be L-toned, while the corresponding content words are mostly H-toned.

Pichi also has a few disyllabic minimal pairs, see (1) in **Table 2**. We also find the maximal number of possible tone patterns over disyllabic words, see (2). A phrasal tonal minimal pair is given in (3), where *òpìn-yày* 'open-eye(s) = cultivated' has undergone the tonal derivation of compounding (see section 2.1.2). Abbreviations and glossing conventions are listed and explained at the end of this article.

2.1.2. Tonal Processes and Grammatical Tone

Pichi tonal processes are operative within prosodic domains of various sizes (see Yakpo, 2019a, 46–57 for details and pitch

TABLE 1 | Monosyllabic tonal minimal pairs in Pichi.

H tone	Gloss	L tone	Gloss
báy	'buy'	*bày*	'by'
bɔ́t	'hit with the head'	*bɔ̀t*	'but'
dé	'day; there'	*dè*	'IPFV'
dí	'this'	*dì*	'DEF'
gó	'go'	*gò*	'FUT; POT'
lɛ́k	'(to) like'	*lɛ̀k*	'like'
só	'like this; sew; show'	*sò*	'so'
wét	'wait'	*wèt*	'with'

TABLE 2 | Multisyllabic tonal minimal pairs in Pichi.

	Item	Gloss	Item	Gloss
(1)	*kàtá* /L-H/	'catarrh'	*kátà* /H-L/	'scatter'
	pàpá /L-H/	'father'	*pápà* /H-L/	'potato'
(2)	*fíbà* /H-L/	'fever'	*nyɔ́ní* /H-H/	'ant'
	wàtá /L-H/	'water'	*bàtà* /L-L/	'buttocks'
(3)	*ópìn yáy* /H-L H/	'open (an) eye'	*òpìn-yáy* /L-L-H/	'cultivated'

traces of examples provided in this section). Processes include tonal plateauing when the L-toned syllable of a disyllabic verb with an H-L pattern is hemmed in by the left-adjacent H and a right-adjacent H of a following object, as in *prɔ́mìs mí* /H-L H/ → *prɔ́**mís** mí* [H-**H** H] 'promise me'. Pichi also features downdrift (indicated by ↓H), which causes an H to be lowered by a preceding L tone, as in *yéstà**dé*** [H-L-↓**H**] 'yesterday'. In a series of adjacent H tones, we find downstep (also indicated by ↓H): Each H tone is lowered successively in relation to the preceding one, as in *wákà sén **sén sén*** [H ↓**H** ↓**H**] 'walk same same same = walk exactly in one line'. Pichi also features pitch or register raising for focal emphasis when all H tones of a focused constituent are raised a notch higher (also see 2.2.3, 3.1, and 4.3).

There are also instances of grammatical tone, i.e., processes restricted to the context of a specific morpheme or construction (Rolle, 2018). Tone floating and contour tone formation take place when the H-toned subjunctive marker (a complementizer) *mék* /H/ 'SBJV' occurs left-adjacent to the monosyllabic personal pronouns *à* /L/ '1SG.SBJ' and *è* /L/ '3SG.SBJ'. The final consonant of *mék* is generally not pronounced and this leads to a vowel hiatus and to further deletion of the vowel of *mék*. In the process, the H tone of *mék* is floated and linked to the L-toned syllable of the personal pronouns *à* and *è,* i.e., *mâ* /HL/ 'SBJV.1SG.SBJ' and *mê* /HL/ 'SBJV.3SG.SBJ'. The resulting portmanteau morphemes and contour tone are so common (the two words/tones are almost always merged), that the contour tone may be seen to be phonologized, i.e. *má* /H/ 'mother, madam' vs. *mâ* /HL/ 'SBJV.1SG.SBJ'.

The inflectional expression of the grammatical relations of subject, object, and possessive case involves the use of tonal suprafixation with personal pronouns (see Yakpo, 2019a, 128), see **Table 3**.

An example of case assignment in 1SG pronouns (object vs. possessive case) via tonal ablaut is given in (1). Note that in cases of clash between subject case and emphasis, the latter series of pronouns wins out, e.g., ***mí*** *nó sàbí* 'I [EMP] don't know' vs. *à nó sàbí* 'I don't know'.

(1)

È	*nó*	*gí*	***mí***	***mì***
3SG.SBJ	NEG	give	1SG.OBJ	1SG.POSS
mɔ̀ní	*yét.*			
money	yet			

'He hasn' t yet given **me my** money'.

The use of grammatical tone also characterizes tonal derivation in compounding and morphological reduplication. The H tone of the dependent is replaced by an L tone, while the head retains its original tone pattern; compare *wách* /H/ (to) 'watch' and *mán* /H/ 'man' → ***wàch**-mán* /**L**-H/ 'watchman', or ***wàch**-wách* /**L**-H/ 'to continuously/repeatedly watch'. An example involving a disyllabic dependent is *bérìn* /H-L/ 'bury' and *grɔ́n* /H/ 'ground' → ***bè**rìn-grɔ́n* /**L**-L-H/ 'burial ground' (also see Yakpo, 2012).

Thirdly, Pichi features a tone-conditioned suppletive allomorphy, a cross-linguistically rare or at least underreported phenomenon (Paster, 2006). Pichi has two pronominal variants that both instantiate (direct and indirect) object case. The variants are the clitic object pronoun *=àm* '3SG.OBJ' and the phonologically independent and emphatic pronoun *ín* '3SG.EMP'. The clitic *=àm* '3SG.OBJ' is the default form used in all licit contexts. Hence *=àm* is the only possible option if the host verb features a word-final consonant (2) or word-final H-toned vowel (3).

(2)

È	*gò*	*márèd=**àm**.*
3SG.SBJ	POT	marry=3SG.OBJ

S/he' ll marry him/her

(3)

À	*fít*	*ték*	*dì*	*wɔ̀tá*
1SG.SBJ	can	take	DEF	water
à	*tròwé=**àm**.*			
1SG.SBJ	throw=3SG.OBJ			

'I can take the water (and) pour it away'.

Examples (4) and (5) featuring the independent pronoun *ín* '3SG.EMP' are therefore ungrammatical:

(4)

**È*	*gò*	*márèd*	***ín**.*
3SG.SBJ	POT	marry	3SG.EMP

'S/he will marry him/her'.

(5)

**È*	*dɔ́n*	*tròwé*	***ín**.*
3SG.SBJ	PRF	throw	3SG.EMP

'S/he has thrown it away'.

The use of the allomorph *=àm* is, however, also ungrammatical if the host features a word-final L-toned vowel (6). Pichi tonotactics disallow string-adjacent identical tones in the same phonological word, hence in this case *ṾṾ >> ṾCṾ, V́Ṽ (string-adjacent H tones are also banned but this is not relevant here). The corresponding examples are *(6) >> (2), (3) (for additional layers of rules, see Yakpo, 2019b, 206–212). The restriction is therefore a manifestation of the Obligatory Contour Principle (OCP) (Leben, 1973). In order to avoid a breach of the OCP, the independent and emphatic pronoun *ín* is recruited when the verb features a word-final L-toned vowel (7).

(6)

**Yù*	*fíbà=**àm***	*bɔ̀kú.*
2SG	resemble=3SG.OBJ	a.lot

'You resemble him/her a lot'.

(7)

Yù	*fíbà*	***ín***	*bɔ̀kú.*
2SG	resemble	3SG.EMP	a.lot

'You resemble him/her a lot'.

TABLE 3 | Suprafixation with personal pronouns in Pichi.

Category expressed	Suprafix
Subject and possessive case	L tone
Object case and emphasis	H tone

2.2. A Creole With Stress and Tone: Guyanese Creole

The following description of the prosodic system of Guyanese Creole examines aspects of the analyses by Devonish (1989, 2002; and pers. comm.) and Devonish and Thompson (2010). My interpretation of the data is that Guyanese Creole has a mixed prosodic system featuring both lexical tone and stress (section 2.2.1 and section 2.2.2). Guyanese Creole additionally features residual tone in ideophones (section 2.2.1 and section 3.2).

2.2.1. Stress, Tone, Prosodic Patterns, and Minimal Pairs

Guyanese Creole makes use of tone and stress. The most reliable indicator of stress in Guyanese Creole is a quantity contrast: Stressed syllables are generally longer than unstressed ones. Stress appears to be assigned lexically, but the nature of stress placement is not elucidated fully in the sources. I therefore focus on the pitch-related aspects of the prosodic system. Devonish and Thompson (2010, 9) refer to Guyanese Creole as a 'restricted' tone language. I assume that the language has a privative contrast between /HL/ and Ø (i.e., zero or toneless) in the majority lexicon. Hyman (2011a, 191) employs the term 'privative' to characterize a binary contrast (/HL, Ø/ in the case of Guyanese Creole), in which only one tone (i.e., /HL/) is phonologically active, i.e., 'invoked by the language's constraints or rules'. The pitch traces contained in Devonish (2002) also indicate interpolation (gradual transitions between pitch peaks). Further, the presence of the HL contour is obligatory and culminative in lexical words, and there are positional restrictions on its occurrence. Tone systems like that of Guyanese Creole are also called 'sparse', a term I will use from now on (Gussenhoven, 2004, 34–35; Hyman, 2011b, 235).

The HL lexical tone is assigned independently of stress and may or may not coincide with the stressed syllable. There are various types of output prosodic patterns. One group consists of toneless (Ø) monosyllabic clitics that are realized as [L]. These form tonal minimal pairs in their output forms with other, segmentally identical words, specified for H(L) (the L of the contour is not realized in these monosyllables), e.g., *gó* [H] '*go*' vs. *gò* [L] 'FUT', and *bín* [H] 'to have been in a location' vs. *bìn* [L] 'PST'. Note the prosodic parallelism between the minimal pair *gò* vs. *gó* in **Table 1** (Pichi).

Guyanese Creole also has disyllabic tonal minimal pairs. Two examples follow (stressed syllable in bold, HL lexical tone indicated by a rising-falling circumflex): (1) ***pa**kît* /**Ø**-HL/ 'packet' → [H-H] vs. (2) ***pâ**kit* /**HL**-Ø/ 'pocket' → [HL-H] and (1) ***flo**wâ* /**Ø**-HL/ 'flour' → [H-H] vs. (2) ***flô**wa* /**HL**-Ø/ 'flower' → [HL-H]. For want of space, I shall not delve into the complex rules formulated by Devonish (2002, 86–95) as well as Devonish and Thompson (2010, 10–11) to account for the divergent realization of these word-tone patterns in phonological brackets to the right of the arrow. What is relevant is that the output tone patterns after the arrow show a pitch contrast between forms (1) and (2). Note that both sets of minimal pairs feature stress on the penultimate syllable, showing that the assignment of tone is independent of the assignment of stress.

2.2.2. Tonal Processes and Grammatical Tone

Guyanese Creole does not seem to have a similarly broad use of grammatical tone as Pichi. Compounding and reduplication nevertheless show parallels with Pichi (Devonish and Thompson, 2010, 13–57). Like in Pichi, the formation of compounds (1) and reduplications (2) (**Table 4**) involves tonal derivation: The lexical HL tone is deleted in the non-final component (the dependent), while the final component (the head) retains its lexical HL tone (Devonish and Thompson, 2010, 11). Note that in long vowels the contour is spread out across both moras hence *láàng* 'long'.

TABLE 4 | Tonal derivation of compounds and reduplications in Guyanese Creole.

	Component 1	Component 2	Compound/ Reduplication
(1)	*shâp* /HL/ 'shop'	*frônt* /HL/ 'front'	*shap-frônt* /Ø-HL/ 'shop front'
	blâk /HL/ 'black'	*pê.pa* /HL/ 'pepper'	*blak-pê.pa* /Ø-HL.Ø/ 'black pepper'
(2)	*láàng* /HL/ 'long'	*láàng* /HL/ 'long'	*laang-láàng* /Ø-HL/ 'long here and there'

The prosodic pattern of Guyanese Creole (and Pichi) compounds and reduplications is therefore the opposite of that found in British English, where the first component receives stress and the second is deaccentuated, i.e., ***shop**-front*. I have provided evidence elsewhere that the prosodic features of compounding found in Guyanese Creole, Pichi, and other Afro-Caribbean English-lexifier Creoles conform to an areal pattern found across West Africa (Yakpo, 2012).

2.2.3. Residual Tone

Guyanese Creole also features 'residual tone'. The term goes back to Berry (1972) and has been employed by some to characterize the occurrence of lexical, phrasal, or grammatical tone in specific semantic fields, and grammatical and pragmatic functions in Afro-European creoles otherwise characterized by stress systems (e.g., Todd, 1980; Granda, 1986; Smith and Adamson, 2006). In Guyanese Creole, residual tone is found in the formation of ideophones. The features of residual tone differ from those of the tone system described in section 2.2.1–2.2.2 in the following way: (1) There are two distinct tone heights, which differ from the HL contour described above, suggesting an /H, L/ or /H, Ø/ contrast; (2) There is no evidence for a quantity contrast (i.e., stress) in the ideophones covered here.

Ideophones depict sensory imagery pertaining to sensations like motion, visual appearance, texture, and feelings (Dingemanse, 2018). They also tend to be structurally marked cross-linguistically, for example through the presence of phonemes that are rare in other word classes of the same language, or the presence of lexical tone in a prosodic system otherwise characterized by stress, as in Guyanese Creole (All examples in this section are from Hubert Devonish, pers. comm. Tones in ideophones are marked). Example (8) contains the ideophone *pím.pím* /H.H/ 'remained quiet, did not respond verbally when a response might have been

expected', with two successive H tones. Example (9) features the ideophone *kìtàkàtà* /L-L-L-L/ 'hectically', which bears L tones throughout. The ideophone *brámbrámbrám* /H-H-H/ 'with a rumbling noise' (10) features three successive H tones. All three ideophones consist of (lexicalized) duplications or triplications, a common feature in African creoles and non-creoles (see section 2.1.1).

(8)

Mi	*tel*	*am*	*se*
1SG	tell	3SG.OBJ	QUOT
ii	*lai*	***pímpím***.	
3SG	lie	IDEO	

'I told him he was lying and my reproach was met with silence'.

(9)

Ii	*a*	*ron*	*chruu*
3SG	PROG	run	through
di	*hous*	***kìtàkàtà***.	
DEF	house	IDEO	

'She is running through the house hectically'.

(10)

Ii	*lik*	*dong*	*di*	*door*
3SG	hit	down	DEF	door
brámbrámbrám.				
IDEO				

'He hit down the door with a rumbling noise'.

Residual tone possibly also occurs in some degree-modifying adverbs, where an extra-high tone expresses focal emphasis together with the adverb. The multifunctional word *sotil* is a clause introducer with the meaning 'until' in time clauses like (11). When the clause introducer occurs at the end of a clause in an 'unfinished utterance', as in (12), it expresses emphasis and meanings like 'a lot' or 'excessively'. In the latter instance, *sótíl* always bears extra-high pitch on both syllables. The African English-lexifier creoles feature both uses of the corresponding form *(só)té(é)* as well, including the use of extra-high tone and final-vowel lengthening (see section 4.2) (for an example sentence in Pichi, see Yakpo, 2019a, 277). The phenomenon has also been described for Sranan and African creoles under the term 'register raising' (Smith and Adamson, 2006; see section 3.3 and section 3.4). Only an acoustic analysis can eventually clarify whether the extra-high tone in a context like (12) instantiates lexical tone or a purely intonational use of pitch.

(11)

Ii	*biit*	*di*	*eg*	***sotil***
3SG	bit	DEF	egg	until
ii	*spail*.			
3SG	spoil			

'She beat the egg until it spoiled'.

(12)

Ii	*biit*	*biit*	*biit*	*di*
3SG	bit	REP	REP	DEF
eg	***sótíl***.			
egg	excessively			

'She beat the egg repeatedly and excessively'.

3. AFRO-EUROPEAN CONTACT PROSODIC SYSTEMS SHOW AN AREAL DISTRIBUTION ACROSS THE ATLANTIC

The analyses in section 2 have shown the existence of tone and stress systems in the same linguistic family, as well as mixing between the two prosodic types. I will now argue that Afro-European creoles and colonial varieties show an areal distribution across the Atlantic basin (section 3.4), which is roughly coterminous with the presence of tone in the east (section 3.1) and stress in the west (section 3.3). Transitional systems combining features of tone and stress are found in the areal buffer zone of the Caribbean (section 3.2). When the social and linguistic composition of an ecology changes, contact languages and varieties can shift from tone to stress systems and vice versa (section 3.5).

3.1. African Creoles and Colonial Varieties, and American Maroon Creoles Have Tone Systems

Tone systems typify all creoles and European colonial varieties (the varieties of English, French, Spanish, and Portuguese) spoken in the tonal ecologies of Africa for which detailed phonological data is available. Tone is one of the most conspicuous typological features across Africa (Maddieson, 2005). Tone was therefore naturally imposed on the prosodic systems of creoles and colonial varieties spoken in Africa. Tone systems are also found in isolated Maroon creoles of the Americas, which probably retained tone systems from earlier times (Rivera Castillo and Faraclas, 2006 provide a first typological comparison of African and Maroon creole systems with African non-creole systems).

All African English-lexifier creoles have been described as tonal. Krio (e.g., Berry, 1970; Hancock, 1971; Nylander, 1984; Finney, 2004), Pichi (Yakpo, 2019a), and Nigerian Pidgin (Faraclas, 1996) have been analyzed in detail, showing the presence of equipollent 2-tone systems with fully specified H and L tones, fixed word-tone patterns and tonal minimal pairs. Most English-derived words have a culminative and obligatory H. Polysyllabic lexemes with more than one H or no H at all are fewer and are mostly found in words with an African etymology. All African English-lexifier creoles make use of grammatical tone. We find identical or similar instantiations of grammatical tone like tone deletion and replacement during compounding and reduplication in Pichi (section 2.1.2), Nigerian Pidgin (Faraclas, 1996, 251–252), Krio (Finney, 1993), Cameroon Pidgin (Nkengasong, 2016, 36–37), and Ghanaian Pidgin (Huber, 2003). Case functions in personal pronouns are expressed by tonal contrasts (e.g., Nigerian Pidgin, Faraclas, 1991). For the better studied languages Nigerian Pidgin and Pichi, word-level and phrase-level processes have been described including downstep, tone-spreading, deletion, polarization (the OCP-triggered assignment of an opposite, polar tone, to an adjacent morpheme/syllable), and pitch or register raising. Preliminary analyses of my field data suggest that most of the lexical, grammatical, and phrasal functions of tone identified for the other African

English-lexifier creoles are also found in Ghanaian Pidgin and Cameroon Pidgin.

Tone systems are also found in the insular Gulf of Guinea Portuguese-lexifier Creoles Forro (Maurer, 2008) and Angolar (Maurer, 1995), spoken in São Tomé, Lung'le (Agostinho and Hyman, 2021), spoken in Príncipe, and Fa d'Ambô, spoken in Annobón. Fa d'Ambô and Lung'Ie have been analyzed as languages with a privative /H, Ø/ contrast, based on the stress contrast of Portuguese. The value Ø is generally realized as [L]. In Fa d'Ambô, for example, we find stress-to-tone mapping between lexifier and creole forms such as the following (H-toned and stressed syllable in bold here and thereafter): *fa**la*** /Ø-H/ 'say', from Port. *fa**lar*** 'say', and ***mo**sa* /H-Ø/ 'woman', from Port. ***mo**ça* 'girl' (Zamora, 2010).

The analysis of Lung'Ie, in turn, shows a privative /H, Ø/ contrast between the three word-tone patterns /H/, /Ø-H/, and /H-Ø/. The prosodic lexicon is etymologically stratified. Portuguese-sourced nouns have a culminative H tone, e.g., ***pá**ta* /H-Ø/ 'duck' from Port. ***pa**ta*. African-sourced words are, by contrast, toneless, and bear L output tones, e.g., *ugbododo* /Ø-Ø-Ø-Ø/ → [L-L-L-L] 'precipice'. Agostinho and Hyman (2021, 88) explain the somewhat unexpected feature of the latter stratum by the resolution of the prosodic clash between the minority African lexicon (with non-culminative, non-obligatory H) and the Portuguese-sourced majority lexicon (with culminative H tone due to stress-to-tone mapping). Moreover, Lung'Ie has not been in much contact with tonal African adstrates for several centuries because of Príncipe's geographical isolation as an island. Idiosyncratic outcomes and innovations are to be expected during prosodic mixing due to differing social histories, and should not be seen as unique to creoles (see also Good, 2009). There is no evidence for stress in Lung'Ie (Agostinho and Hyman, 2021, 81–86) and due to the absence of similarly detailed acoustic analyses, it is difficult to substantiate claims that Forro and Fa d'Ambô employ stress in addition to tone (e.g., Traill and Ferraz, 1981; Zamora, 2010).

By contrast, the family of Upper Guinea Portuguese-lexifier creoles of Cape Verde (Kabuverdianu) (Swolkien, 2015), Guinea-Bissau (Kriyol) (Chapouto, 2014), and Senegal (also called Kriyol by its speakers) (Biagui, 2012) have all been analyzed as languages with stress, not tone. Lang (2009) and Jacobs (2010) provide convincing lexical and structural evidence for a founder role of non-tonal Wolof (Atlantic) at a crucial period in the development of Upper Guinea Creole in the 15th century. Other Atlantic languages that were probably represented in the creole founder population (e.g., Fula, Seereer, and Joola) and are still spoken alongside the creole in Guinea Bissau and Senegal also have stress-only systems. This aligns the Upper Guinea creoles prosodically with other non-tonal languages spoken in adjoining parts of West Africa. Further, non-tonal Portuguese has been spoken alongside these creoles by the descendants of founder populations for several centuries (Jacobs, 2010, 302–307), hence much longer than other European languages in Africa. Nevertheless, there is equally strong linguistic evidence for an input of the tone languages Manding (Mande) and Temne (Atlantic) into Proto-Upper Guinea Creole (Quint and Tavares, 2019). Alternatively, the Upper Guinea Creoles could therefore also have completed a shift from tone to stress due to the prolonged absence or marginalization of African tone languages in the ecology, just like most creoles of the Caribbean. The presence of residual tone in Cape Verdean has indeed been suggested by Macedo (1979, 132–134), though not corroborated by acoustic analyses.

Crucial support for the areal distribution of prosodic systems across the Afro-American Atlantic comes from the presence of tone systems in African varieties of English, French, Spanish, and Portuguese, i.e., of the very lexifiers of the creoles. West African varieties of European colonial languages like Nigerian English (Gussenhoven and Udofot, 2010) and Ghanaian English (Criper-Friedman, 1990) have been analyzed as privative systems with a two-way /H, Ø/ contrast and fixed word-tone patterns. In both varieties, English-sourced content words feature a culminative and obligatory H. The syllable with primary stress in the British English cognate receives an H, as in ***mem**ber* /H-Ø/ → [H-L]. Monosyllabic function words that are unstressed in British English are toneless and L-toned in the output, e.g., *of* /Ø/ → [L], *a* /Ø/ → [L], and *he* /Ø/ → [L]. Nigerian and Ghanaian English both also feature rightward H tone spreading in utterance-medial positions (e.g., ***mem**ber* /H-Ø/ → [H-H]), as well as downdrift.

Central African French and Equatorial Guinean Spanish have been analyzed as equipollent /H, L/ systems with fully specified tone and no stress (Bordal Steien and Yakpo, 2020). In Central African French, an /H/ is realized on the final syllable of every content word, thus replicating the most frequent position of phrasal stress at the word level in European French. Other syllables receive an /L/. Central African French has two fixed word-tone patterns, namely /L/ and /(L)H/, e.g., *ce* /H/ 'this', *le* /L/, *sentir* /L-H/ 'feel'). Equatorial Guinean Spanish, in turn, has four word-tone patterns, namely /L(-L)/, /(L-)H/, /(L-)H-L/, and /(L-)H-L-L/, e.g., *desde* /L-L/ 'since', *yó* /H/ '1SG.SBJ', *porqué* /L-H/ 'why?', *clase* /H-L/ 'class', *película* /L-H-L-L/ 'film'.

Central African French and Equatorial Guinean Spanish both have tonally distinguished minimal pairs in the category of function words, e.g., Equatorial Guinean Spanish *tú* /H/ '2SG.SBJ', *tu* /L/ '2SG.POSS'. The tone of Central African French personal pronouns is not only lexically specified but also unpredictable on the basis of French stress, e.g., *ils* /H/ '3PL.SBJ.M' vs. *il* /L/ '3SG.SBJ.M'. Equatorial Guinean Spanish also has H tone spreading and downdrift, e.g., *jóvenes* /H-L-L/ 'youths' → [H-H-H]. Tone systems apparently also characterize other African Romance varieties, among them the French varieties of Côte d'Ivoire (Boutin and Turscan, 2009) and Mali (Bordal and Skattum, 2014). The data on African varieties of Portuguese is not conclusive (e.g., Santos, 2019 for Angolan Portuguese). But it would be unusual if these varieties had stress-only systems, since most of them are spoken in tonal ecologies.

Besides the creoles and colonial varieties spoken in Africa, the Maroon creoles of the Americas for which we have conclusive data feature tone systems. Maroon creole languages are spoken by the descendants of Africans who liberated themselves from European enslavement and established independent polities in areas that remained geographically secluded until the 20th century. Maroon creoles therefore also remained relatively isolated from (non-tonal) European superstrates and creoles until quite recently (e.g., non-tonal Dutch and Sranan in Suriname).

Saramaccan (traditionally spoken in the Amazonian interior of Suriname) has a two-height contrast, fixed word-tone patterns, and tonal minimal pairs (Good, 2009). Most Portuguese- and English-sourced words feature a privative /H, Ø/ contrast. /H/ is borne by the syllable that bears stress in the lexifier; compare ***wó**mì* /H-Ø/ 'man' (<Port. ***ho**mem* 'man') and ***àkí*** **/**Ø-H/ 'here' (<Port. ***aqui*** 'here'). African-sourced words are fully specified for tone and feature an equipollent /H, L/ specification. In addition, a phonologized extra-H tone /!H/ is found in mostly African-sourced ideophones (Good, 2006, 20), thus constituting a third tone height (see 4.3). African words also have more diverse patterns, e.g., *lɛ̀gɛ̀dɛ̀* /L-L-L/ 'lie (noun)' and *tótómbòtí* /H-H-L-H/ 'woodpecker'. There are also numerous tonal processes in Saramaccan, including H tone spreading and raising, and plateauing (Rountree, 1972; Good, 2004). Like in the African English-lexifier creoles, personal pronouns are inflected by tonal ablaut to express case functions, e.g., *mì* '1SG.SBJ/POSS' vs. *mí* '1SG.OBJ/EMP' (McWhorter and Good, 2012, 42).

The related Surinamese Maroon creole Ndyuka also has a two-way height contrast including tonal minimal pairs, e.g., *tàkì* /L-L/ 'quotative complementizer' vs. *tákì* /H-L/ 'say' (Huttar and Huttar, 1994, 5). Compounds are created via the same tonal derivation as in the African English-lexifier creoles, i.e., *káw* /H/ 'chew' and *bón* /H/ 'bone' → *kàw-bón* /L-H/ 'chewed-bone(s)' (Huttar and Huttar, 1994, 373). The analysis by Hualde and Schwegler (2008) of the prosodic system of the Spanish-lexifier Maroon creole Palenquero spoken in the town of Palenque de San Basilio (Colombia) also indicates the presence of a tone system, when it is stated that 'accented' syllables *consistently* carry high (contour) pitch' (Hualde and Schwegler, 2008).

Instead of seeing the tone systems described in this section as exceptional they should be understood as typical instantiations of source language agentivity in ecologies dominated by tone languages. Such tone systems develop through the three mechanisms of stress-to-tone mapping, paradigmatization, and idiosyncratization (see section 4).

3.2. Some American Varieties Combine Stress With Tone

A second group of contact languages and varieties combines stress with tone in various ways. The resulting mixed systems include (sparse) tone systems of the type encountered in Guyanese Creole and Papiamentu (see below for the latter), in which tone and stress co-occur throughout the lexicon. Further, they extend to systems with fully specified 'residual tone' (Berry, 1972) in specific semantic fields and in specialized functions in prosodic systems otherwise characterized by stress alone, or both stress and sparse tone (e.g., Guyanese Creole, section 2.2).

Sranan (Suriname), for example, makes use of stress alone in the majority of its lexicon with characteristic effects like lengthening of stressed syllables, shortening of unstressed ones and consonant gemination, e.g., *pa**pa*** 'father' → [p**pa**], *wo**wo**yo* 'market' → [w**wo**yo] (van der Hilst, 1988, 51–54). Sranan ideophones, however, have fixed H or L tones, e.g., *píí* /H-H/ 'quietly', *pétépété* /H-H-H-H/ 'thoroughly' vs. *tjùbùm* /L-L/ 'with a plopping sound' (Smith and Adamson, 2006).

The use of lexical tone in addition to stress has also been posited for Tobagonian in the distinction between the grammatical and pragmatic functions of personal pronouns (James, 2003), e.g., *dèm* /L/ '3PL' vs. *dém* /H/ '3PL.EMP' (for parallels with Krio/Pichi, see section 4.2). In the absence of acoustic evidence that such uses of pitch in Tobagonian are indeed tonal, and not concomitants of stress and intonation, this is, however, difficult to verify.

Haitian Creole differs from its lexifier European French in that individual words all bear lexical stress on the final syllable (Cadely, 1994; for acoustic evidence, see Kalkhoff, 2018). In Haitian, the prosodic constituent is therefore not the accentual phrase as in European French, but the prosodic word, which is also the domain of attribution of lexical tone. In addition, the Haitian post-nominal determiner is stressed and consistently high-pitched, e.g., *mayi-**a*** 'maize-DET'. Kalkhoff (2018) proposes that this is in emulation of the H tone of the corresponding post-nominal determiner in Gbe, Haitian's main substrate cluster (Kalkhoff, 2018), i.e., *blǐ-**á*** 'maize-DET' (own knowledge). Further, in some basilectal varieties of Haitian (e.g., rural varieties that incorporate fewer features from the French superstrate than urban varieties), word stress is apparently replaced by word-final high pitch alone, thus mirroring the tone systems of African varieties of French (Brousseau, 2003, 132; see section 3.1). Sylvain (1936; cited in Gooden, 2003, 193) mentions the existence of tonal minimal pairs in distinguishing intensive from attenuative meanings in reduplications, e.g., *píké-píké* /H-H-H-H/ 'very pricking' vs. *pìkè-pìkè* /L-L-L-L/ 'slightly pricking'. There are thus indications that Haitian has residual tone, and has merged aspects of the stress-only system of its lexifier European French with the tone systems of its African substrates, prompting Kalkhoff (2018) to call it 'mixed'.

Besides Guyanese Creole (section 2.2), the Iberian-lexifier creole Papiamentu (Netherlands Antilles) is prosodically fully mixed in the sense that the acoustic properties of tone and stress co-occur and are generalized across the entire lexicon (Devonish, 1989, 60; Rivera-Castillo, 1998; Rivera-Castillo and Pickering, 2004; Remijsen and Van Heuven, 2005). Papiamentu has a tone system that combines lexical word stress with lexical tone and numerous tonal processes, including downstep and polarization. The vowels of stressed syllables are longer and louder than unstressed ones, and unstressed syllables tend to be more centralized (i.e., more schwa-like) than stressed ones. Some analyses postulate an /H, L/ equipollent system with full lexical specification of tones (Rivera-Castillo, 1998; Kouwenberg, 2004). According to Remijsen and Van Heuven (2005), Papiamentu has a privative /HL, Ø/ contrast. Syllables specified for Ø are realized as lower than the H of the HL lexical tone or they carry an LH intonational pitch accent that signals focus. Further, the lexical HL tone can, but need not coincide with the stressed syllable. The position of the lexical HL is also unpredictable in a large number of words. In Joubert (1991), there are over two hundred tone-stress minimal pairs, see ex. (1–3) below. Further, tone distinguishes (1) disyllabic verbs from (2) disyllabic nominals and is therefore also used for derivation (see Kouwenberg, 2004, 66–69 for more instances of grammatical tone). Examples

follow (stressed syllable in bold, HL lexical tone indicated by a rising-falling circumflex): (1) ***lo*rârâ /Ø**-HL/ 'to turn' (verb), (2) ***lo*rârâ /Ø-HL/** 'turned' (participle), (3) ***lô*ra /HL**-Ø/ 'parrot'.

Few Afro-European creoles have been studied as extensively with regard to prosody as Papiamentu. It is therefore possible that systems with stress and (sparse) tone as well as residual tone are far more common in the Americas than meets the eye. Some of the stress-only systems described in section 3.3 and others not mentioned here could therefore turn out to be mixed as well.

3.3. Other American Varieties Combine Stress With African Intonational Features

Many Afro-Caribbean creoles and American varieties of European colonial languages spoken by African-descended majorities feature stress-only systems without lexical and morphological tone. But 'suspicious' features raise the possibility of a tonal past and an areal switch from tonal or mixed to stress-only systems (see section 3.5 and 5.2 for further discussion). Relevant prosodic features of some of these creoles and colonial varieties are discussed in the following.

Gooden (2003) produces evidence that the English-lexifier Creole Jamaican makes use of lexical stress, not lexical tone. But the nature of pitch movements associated with word-level stress and intonational pitch accents is unlike that of its lexifier British English (e.g., an H*L on stressed English syllables vs. an H+L* on stressed syllables in Jamaican). Equally, compounding involves morphological stress placement on the rightmost morpheme like in Guyanese Creole (see section 2.2.2), which is reminiscent of compounding in tonal Pichi (see section 2.1.2), but unlike English, where the first morpheme is stressed. Further, the prosodic rhythm of Jamaican (Thomas and Carter, 2006) and Guyanese Creole (Devonish, 2002, 96–97) is syllable-timed. Syllable timing means that the duration of each syllable is more or less equal, unless there is some form of pragmatic marking. Syllable timing gives Jamaican an auditory impression that prompted earlier (English-speaking) observers to mistakenly classify Jamaican as tonal (e.g., Lawton, 1963). The prosodic rhythm of the lexifier British English is, by contrast, stress-timed. Phonetic effects to achieve optimal prosodic rhythm in British English are the lengthening of vowels in stressed syllables, the reduction of vowels in unstressed syllables, e.g., *police* [plis], as well as vowel laxing, e.g., *sane* [sein] vs. *sanity* [saniti] (Ciszewski, 1999, 30). Syllable timing, rather than stress-timing generally appears to be a hallmark of African tone languages (Gut et al., 2001) and is therefore very likely to be a tonal carry-over (see Bloomquist et al., 2015, for a summary of similar arguments with respect to Jamaican, Jamaican English, and African American English).

Many Caribbean creoles, whether they have been analyzed as tonal or not, also show pitch-related intonational features found across tonal Africa. These include utterance-level declination, which parallels downdrift and is widely attested in tonal Africa (Yip, 2002, 262–263), as well as 'register raising' (assignment of extra-high pitch to the entire relevant constituent, not just the stressed syllable) for focal emphasis (for Guyanese Creole, Trinidadian English Creole, and Bajan, see Sutcliffe, 2003; for Sranan, see Smith and Adamson, 2006; for Pichi, see Yakpo, 2019a, 55–57). A further, seemingly pan-Caribbean intonational feature with parallels in the tonal substrates is an utterance-final fall in wh-questions (Sutcliffe, 2003), which corresponds to 'lax question intonation', a Macro-Sudan areal feature of Africa (Güldemann, 2018, 481). Lax intonation has also been described for the Gulf of Guinea Portuguese-lexifier Creoles (Agostinho et al., 2019).

Colonial varieties of English, French, Spanish, and Portuguese spoken by African-descended majorities in the Americas also show intonational features that differ in often substantial ways from the colonial varieties spoken by European-descended populations. Caribbean varieties of English have a reputation for their 'melodic intonation', an auditory impression that is, again, occasioned by their greater degree of syllable-timing, un-English pitch contours associated with stress, and a greater range of pitch variation across stressed and unstressed syllables alike (see Wells, 1982).

Speakers of European French also think that Caribbean French has an 'accent chantant' because speakers of the latter place stress on individual words, not accentual phrases, again like Haitian. They also tend to stress the first syllable of multisyllabic words in addition to the last, and may stress clitics and prepositions, something that speakers of European French do not normally do (Pustka, 2007).

In the same vein, authors have commented upon, though often not described in detail, the peculiar prosodic characteristics of rural varieties of Spanish and Portuguese spoken in countries with large African-descended populations and by isolated communities of African origin. Popular (vernacular) Brazilian Portuguese, for example, has more utterance-internal pitch accents than European Portuguese with a frequent alternation between and H* and L*, and 'tonal events not linked to stressed syllables' (Frota and Vigário, 2000, 11).

Rao and Sessarego (2016) provide a detailed acoustic and phonological analysis of aspects of the prosody of Afro-Bolivian Spanish. They mention, among other features not found in other Bolivian Spanishes, an obligatory and fixed LH pitch contour or H level pitch over stressed syllables. In European and European-influenced American varieties word-level pitch contours can, by contrast, be significantly altered by intonation (see Hualde and Prieto, 2015 for an overview). Butera et al. (2020) arrive at a similar conclusion with respect to the Afro-Peruvian variety of Spanish spoken in the province of Chincha, Peru. There is cursory evidence for the existence of similar prosodic features in other American Spanish varieties as well, which require further substantiation (e.g., Choco Spanish and Congo, see Lipski, 2007).

The analyses of numerous American 'stress-only' varieties remain somewhat inconclusive. It is well possible that many feature residual tone or constitute mixed tone-stress systems as well. Either way, many of their distinct prosodic characteristics are very likely to result from the incorporation of pitch features of the tonal substrate languages once spoken by the African creators of these varieties.

3.4. The East to West, Africa to the Americas, Tone to Stress Areal Continuum of Prosodic Systems

Table 5 presents the areal east-west, Africa-Americas, tone-stress continuum of prosodic systems in the languages surveyed in section 3.1–3.3. Their classification is based on information contained in the sources cited there. The prosodic features *tone*, *residual tone*, *stress* in the headers of the three central columns are checked against the column captioned *languages*. The symbols + and – indicate the presence or absence of features. When in parentheses (+), evidence for the feature is anecdotal in the literature, i.e., not corroborated by acoustic evidence and

TABLE 5 | The areal continuum of Afro-European contact prosodic systems.

Group		Languages	Tone	Residual tone	Stress	Description of prosodic features
East (tone)	1	Tonal substrate and adstrate languages of Africa	+	–	–	(1) Mainly 2T and 2T3; equipollent (e.g., /H, L/) and privative (e.g., /H, Ø/) (2) Lots of word-level and phrasal tonal processes, incl. everything in Group 2 (3) Lots of grammatical tone, incl. everything in Group 2 (4) Lax question intonation (Macro-Sudan languages) and rising intonation (e.g., some Bantu languages); syllable timing
	2	African English-lexifier creoles (Krio, Pichi, Nigerian Pidgin, Cameroon Pidgin, Ghanaian Pidgin)	+	–	–	(1) 2T and 2T3 systems; equipollent /H, L/, privative /H, Ø/; /!H/ in pragmatically salient functions (degree words, ideophones)
		Gulf of Guinea Portuguese-lexifier creoles of Africa (Forro, Angolar, Lung'Ie, Fa d'Ambô)	+	–	–	(2) Downstep; H-tone raising, spreading, deletion, floating; OCP/polarization; pitch raising for emphasis
		Colonial varieties of European languages spoken in Africa (Nigerian English, Ghanaian English, Central African French, Equatorial Guinean Spanish)	+	–	–	(3) Compounding and reduplication; tonal inflection of personal pronouns; tone-conditioned allomorphy (Pichi); portmanteau morphemes with contour tones (African English-lexifier creoles)
		Maroon creoles of the Americas (Saramaccan, Ndyuka, Palenquero)	+	–	–	(4) Lax question intonation next to rising question intonation; syllable timing
	3	Papiamentu, Guyanese Creole, and possibly other Caribbean creoles	+	+	+	(1) 2T *and* stress; privative /H, Ø/; words with residual tone possibly have equipollent /H, L/; /!H/ in pragmatically salient functions (degree words, ideophones); contrastive stress (2) Downstep; H-tone raising, spreading, deletion, polarization (Papiamentu); pitch raising for emphasis (3) Compounding and reduplication; derivation (Papiamentu) (4) Lax question intonation next to rising question intonation; syllable timing
	4	Caribbean English-lexifier creoles (Sranan, Tobagonian, and possibly others), Caribbean French-lexifier creoles (Haitian, and possibly others)	–	+	+	(1) Contrastive word stress; no tone in the majority lexicon; residual tone with privative /H, Ø/ or equipollent /H, L/ in specific functions and fields (e.g. ideophones and reduplication) (2) Utterance-level and/or phrase-level pitch downtrends but no tonal downstep; pitch raising for emphasis (3) Possible grammatical (residual) tone in compounding and reduplication (4) Lax question intonation next to rising question intonation; syllable timing
	5	Caribbean English-lexifier creoles (Jamaican, Trinidad Creole English, Bajan, and others)	–	(+)	+	(1) Contrastive word stress; no tone in the majority lexicon; no firm evidence for residual tone but likely in some
		Upper Guinea Portuguese-lexifier Creoles (Kriyol, Kabuverdianu)	–	(+)	+	(2) Utterance-level and/or phrase-level pitch downtrend but no tonal downstep; pitch raising for emphasis
		Colonial varieties of European languages spoken by African-descended populations and many vernacular colonial varieties of the Americas (Caribbean French and English, Popular Brazilian Portuguese, Afro-Bolivian and Afro-Peruvian Spanish, Choco Spanish, and possibly others)	–	(+)	+	(3) Morphological stress (e.g., stress shift in compounds and reduplications); no grammatical tone (4) Lax question intonation next to rising question intonation; syllable timing
West (stress)	6	Stress-only lexifier and superstrate varieties of Europe and colonial (standard) varieties of European-descended populations of the Americas (e.g., European French and English, Argentinian Spanish, and others)	–	–	+	(1) Contrastive word-level and phrase level stress; no (residual) tone (2) Utterance-level and/or phrase-level pitch downtrend; no tonal downstep (3) Morphological stress; no grammatical tone (4) Rising question intonation; stress-timing (English, Dutch) and syllable-timing (Spanish, French)

detailed phonological analysis. The rightmost column provides details of the three checked features, with numbers (1)–(4) referring to the following characteristics discussed in section 2.1–2.2 and section 3.1–3.3: (1) type of prosodic system and tonal inventory, (2) phrase-level and utterance-level tonal or pitch-related processes, (3) aspects of grammatical tone, (4) aspects of intonation and prosodic rhythm. The *languages* column contains linguistic groupings and individual varieties. These are, in turn, grouped in the *Group* column in the following way:

The eastern pole (Group 1) at the top of **Table 5** is represented by the tonal substrates and adstrates of Africa. The western pole (Group 6) at the bottom is represented by the stress-only superstrates and lexifiers (English, French, Spanish, Portuguese, and Dutch) spoken in Europe and by largely European-descended populations in the Americas. Group 2–5 prosodic systems emerged from contact between Group 1 (tone) and Group 6 (stress) systems.

Table 5 shows a tone-stress cline from Group 1 to 6 languages with a gradual decrease in tonal features and a concomitant increase in stress-related features. Group 2 creoles and colonial varieties are exclusively tonal (section 3.1). Group 3 features the mixed systems of Papiamentu and Guyanese Creole (section 2.2 and 3.2) that combine stress and privative tone in all of their lexicon, additionally feature residual tone and many but not all of the tonal features of Group 2 languages detailed in the rightmost column. Group 4 languages (section 3.2) feature stress-only systems in most of their lexicon but there is evidence for residual tone in ideophones and some grammatical functions (e.g., compounding and reduplication). Group 4 languages share some pitch-related phrasal and intonational features reminiscent of Group 1 and 2 tone languages (downtrend, pitch or register raising, lax question intonation).

Group 5 languages have stress-only systems (section 3.3). The evidence for residual tone is anecdotal, hence (+) in the corresponding column. However, the evidence is more conclusive that Group 5 languages have incorporated African intonational features in which they overlap with Group 1–4 languages (e.g., lax question intonation in some languages, pitch or register raising, and syllable timing).

Groups 2–5 are idealized types and we should expect many more variations, gradations, and idiosyncrasies than captured by **Table 5**. It is also possible that future research reveals that many languages now in Group 4 and 5 have more tonal features than presently known.

3.5. Areal Switches Are Common in Creole Prosodic Systems

In section 3.1–3.4, I argued for the existence of an areal continuum of contact prosodic systems from Africa to the Americas. Two diachronic scenarios are thinkable on the basis of the demographic (African-descended majorities) and linguistic (tone-dominant ecologies) evidence in relation to the areal continuum summarized in section 3.4. One scenario would suggest that Group 4 and 5 languages in **Table 5** have always featured stress-only systems but incorporated substratal tonal features in their prosodic systems. However, I tend to think that Group 4 and 5 languages that evolved in ecologies with overwhelming African-descended majorities, once had full-blown tone systems like those of Group 2 (see section 3.1). Areal convergence with European stress-only lexifiers and superstrates and other stress-only languages in the ecology (e.g., the Indic languages of the Caribbean, see below) would have then led to the replacement of tone by stress systems and the retention of sparse and residual tone in some varieties.

In languages spoken in ecologies with somewhat less of a demographic dominance of African populations vis-à-vis European populations, as well as the right social factors (e.g., a slightly more porous social stratification of Africans and Europeans), stress and tone could have co-evolved right from the start (e.g., Papiamentu as well as Spanish and Portuguese colonial varieties of the Americas).

A switch from tone to stress has explicitly been claimed by Barth (2016) for Sranan (Suriname). Barth argues on the basis of historical phonology that Sranan once had a tone system like its closest relatives, the Maroon creoles Ndyuka and Saramaccan, and then lost tone through contact with Dutch. The survival of African-style systems of lexical and grammatical tone in the Maroon creoles makes it plausible that many other American creoles and colonial varieties started out as tonal and later shifted to stress-only prosodic systems (see Alleyne, 1980; Devonish, 1989). With the end of the European slave trade in the 19th century, the proportion of L1 speakers of African tone languages began to decline. Throughout the 20th century, socio-economic change led to the partial erosion of racialized social stratification, while formal education in the standardized European varieties of colonial languages was expanded (Yakpo, 2015; also Hackert, 2019, 225).

In the wake of such social transformations, the American ecologies came to be dominated by patterns of societal multilingualism involving creoles and the stress-only superstrates English, Spanish and Dutch. In several cases (Guyana, Suriname, Trinidad), non-tonal Indic adstrates also played an important role in the ecology. The influence of Bhojpuri, for instance, has contributed to changes in the pitch associated with stressed syllables in Trinidad Creole English (Gooden et al., 2009, 419–420). It is possible that source-language agentivity in Bhojpuri and other non-tonal languages of Trinidad besides English (e.g., Portuguese, see Ferreira, 2006) also contributed to the demise of tone in Trinidad Creole English.

The opposite areal switch from stress to tone is, by contrast, a possibility in the trajectory of Krio, the English-lexifier creole of Sierra Leone. In one account, Krio is seen as an offshoot of Western Maroon creole of Jamaica brought to Sierra Leone in the late 18th century by African-descended returnees (Smith, 2017). Jamaican is a stress-only language today (see section 3.3) and contemporary Krio is a tone language without stress (see section 3.1). If Western Maroon creole had already acquired a stress system in Jamaica by the time it arrived in Sierra Leone, then Proto-Krio would have jettisoned stress for tone due to adstratal influence from speakers of African tone languages. Most prominent among these adstrates were Yoruba, Gbe, Mende, Temne, and Manding (Hancock, 1971; for the historical background, see Huber, 1999, 59–74).

The commonness of switches between types of prosodic systems is corroborated by evidence from regions other than the Afro-Atlantic. The Tibeto-Burman family is presently split halfway between languages that employ tone and others that use stress. Tonal Tibeto-Burman languages are found in a prosodic linguistic area encompassing tonal Tai-Kadai, Hmong-Mien and Chinese languages (Ratliff, 2015).

The reverse switch involving 'tonoexodus' (Matisoff, 1973) is also attested. If Proto-Afro-Asiatic was tonal, as suggested by some (see Wolff, 2018), then the tonal Cushitic and Omotic subbranches of Afro-Asiatic retained tone during millennia of contact with tonal Nilo-Saharan. The Proto-Semitic subbranch of Afro-Asiatic therefore probably lost tone along the way. The Cushitic languages Kemant and Khamtanga therefore appear to have lost tone through contact with (Ethiopian) Semitic (Appleyard, 1991, 10). In northern Norway (Jahr, 1984; Bull, 1995) and southern Finland (Bruce, 2004), the superstrates Norwegian and Swedish, which employ both stress and tone (Kristoffersen, 2000), underwent substratal and adstratal transfer from the stress-only languages Sami and Finnish. The resulting contact varieties of Norwegian and Swedish have lost tone and feature stress-only systems.

There is therefore ample evidence that tonogenesis and tonoexodus are cyclical and complementary processes rather than one-way streets (see also Matisoff, 1973). There is no reason to exclude creoles from these cross-linguistic tendencies (see section 5.2 for further discussion).

4. CONTACT PROSODIC SYSTEMS WITH TONE EMERGE THROUGH THREE COGNITIVE-TYPOLOGICAL MECHANISMS

In the preceding sections, I have argued that the distribution of tone and stress is areal across the Afro-American Atlantic. Tone predominates in the east (Africa), stress in the west (Americas), and transitional systems cluster in the areal buffer zone of the Caribbean. Given that the tonal creoles and colonial varieties all have lexifiers with stress-only systems, it is useful to take a closer look at the emergence of tone systems during the encounter of non-tonal lexifiers and tonal substrates and adstrates.

Ratliff (2015, 258) cautions against *broad* explanations for tonogenesis in language contact because they fail to explain 'exactly how tones were either transferred to—or stimulated to develop in—previously atonal languages under contact'. Bordal Steien and Yakpo (2020) propose three *specific* cognitive-typological mechanisms in the genesis of contact prosodic systems with tone, namely: (1) stress-to-tone mapping (section 4.1), (2) paradigmatization (section 4.2), and (3) idiosyncratization (section 4.3).

4.1. Stress-to-Tone Mapping

Through the mechanism of stress-to-tone mapping speakers of tone languages create a tone system in the contact language or variety by building on perceptual analogies between the phonetic realizations of stress and tone (Bordal Steien and Yakpo, 2020, 23–26).

For one, high or rising pitch is a consistent correlate of stress besides duration, loudness, and vowel quality in the European lexifiers, e.g., in French (Jun and Fougeron, 2002), British English (Morton and Jassem, 1965), and Spanish (Hualde and Prieto, 2015). Speakers of tone languages are perceptually also more sensitive to pitch variations than to other acoustic cues of stress. Speakers of Mandarin Chinese selectively perceive the higher pitch of stressed syllables in English, rather than vowel length, loudness, and vowel quality. This makes pitch the primary cue for distinguishing stressed from unstressed syllables for native Chinese speakers (Wang, 2008). In other words, the pitch contour of a stressed syllable is reinterpreted as a tonal contour by tone language speakers and other cues of stress are ignored.

As a result, the position of H in the creoles coincides with primary stress placement in the cognate forms of the lexifier, e.g., Pichi, *go* → *gó* /H/, ***car****penter* → ***kyá****pìntà* /**H**-L-L/, ***en****ter* → ***én****tà* /**H**-L/, *for****get*** → *fɔ****gét*** /L-**H**/, *under****stand*** → *ɔ̀ndà****stán*** /L-L-**H**/. H tone in Pichi is therefore culminative and obligatory in English-sourced content words, just like primary stress placement is in English. Low tones are found on all syllables that do not bear stress in the corresponding English source word. Stress-to-tone mapping is attested in all the Afro-European creoles and colonial varieties covered in section 3.1 and other contact languages not mentioned so far, e.g., in the African Arabic-lexifier creoles Juba Arabic (Nakao, 2013) and Kinubi (Gussenhoven, 2006). Stress-to-tone mapping also characterizes the prosodic systems of Euro-Asian creoles and colonial varieties that arose from the encounter of stress-only superstrates and tonal substrates and adstrates (for an overview of several varieties, see Ng, 2011 and Lim, 2012; for a detailed study of Hong Kong English, see Wee, 2016).

4.2. Paradigmatization

The second mechanism in the creation of contact prosodic systems is paradigmatization (Bordal Steien and Yakpo, 2020, 26–28). Paradigmatization occurs by default when stress patterns of the non-tonal lexifier are mapped onto tone patterns in the contact languages and varieties. In the case of lexicon sourced from African tone languages, tone classes can be carried over into the contact language without prior stress-to-tone mapping, as in the case of Mende and Yoruba items in Krio described further below. But there may be contact-induced adaptation even with African-sourced tonal words, see e.g., 3.1, on Lung'Ie. The resulting tone classes therefore largely mirror corresponding stress and tone classes in the European and African input languages, respectively (see section 2.1.1).

Besides the replication of prosodic structures from source languages, paradigmatization may also regularize functional paradigms. The pronominal system of all African English-lexifier creoles is divided into two series. One series expresses subject and possessive case and invariably bears an L tone, e.g., *yù* '2SG.SBJ/POSS' (<*you*), *wì* '1PL.SBJ/POSS' (<*we*) and *dɛ̀m* '3PL.SBJ/POSS' (<*them*). The other series assumes the syntactic and pragmatic functions of object case and emphasis and is exclusively H-toned, e.g., *yú* '2SG.OBJ/EMP', *wí* '1PL.OBJ/EMP', and *dém* '3PL.OBJ/EMP' (e.g., *you*, *we*, and *them*). Hence,

it is *à sí **yú** yɛ́stàdé!* 'I saw you yesterday', but not **à sí **yù** yɛ́stàdé!* The English-sourced pronouns have therefore undergone paradigmatization in the creoles in order to fit into the L-toned and H-toned case paradigms, respectively. Paradigmatization in the tonal English-lexifier creoles of Africa shows numerous overlaps with that of the tonal English-lexifier creoles of Suriname (e.g. Saramaccan, see section 3.1), suggesting it was already present in the proto-creole(s).

Contact varieties whose lexifiers have fewer stress patterns show correspondingly fewer word-tone patterns. Phrase-final stress in European French has been converted into word-final stress in the tonal African varieties of French (Bordal Steien, 2015; Bordal Steien and Yakpo, 2020). Paradigmatization has only rendered two word-tone patterns in Central African French on the basis of the corresponding European French potential for stress placement (see section 5.2 for further discussion).

4.3. Idiosyncratization

The third mechanism, idiosyncratization, leads to the emergence of arbitrary word-tone patterns, paradigms and constructions. While paradigmatization creates tonally regular forms and paradigms, idiosyncratization therefore creates tonally irregular ones including idiosyncratic grammatical and syntactic tone rules. Idiosyncratization may occur through any combination of the factors of substratal and adstratal imposition, the operation of cross-linguistic tendencies (e.g., interactions of tone with consonant or syllable type), and language-specific constructionalization and grammaticalization, whether through contact or not.

An example follows of idiosyncratization in the word-tone patterns of content words. The African English-lexifier creoles have a considerable stock of English-sourced words in which H tone does not coincide with English primary stress. Examples from Krio/Pichi are ***wa**ter* → *wàt**á*** /L-**H**/, ***trou**sers* → *trɔ̀**sís*** /L-**H**/, ***pro**perty* → *prɔ̀pàt**í*** /L-L-**H**/, ***hos**pital* → *ɔ̀s**pí**tùl* /L-**H**-L/. Such words have undergone tone shift and no longer exhibit a prosodic parallelism with their English etymons (for the background to these changes, see Devonish, 2001). Tone shift therefore constitutes idiosyncratization vis-à-vis the English input.

A further example of idiosyncratization in the lexicon follows from Krio and Pichi. Both creoles have a set of degree-modifying and quantifying words, see **Table 6**. (1, 2) are English-sourced lexicalized reduplications (<*little, big*); (3) is of unknown origin [but see the back formation *sótíl* < **(so) till* in Guyanese Creole, ex. (11, 12)]; (4) is probably Igbo-sourced (<*sọ̀sọ̀* 'only'). The exclamation mark before the /!H/ tone signals that it is extra-high, hence a notch above the usual H tone register. H tone raising is conventionalized with these words, as is the lengthening of the final vowel in *sótéé* (3).

Idiosyncratization has therefore rendered a set of tonally arbitrary forms with respect to stress-to-tone mapping, and these are grouped in a specific semantic field. The use of /!H/ and lengthening for degree modification are probably both iconic processes (see Thompson, 2018, for a possibly universal prevalence of H tone in the sound symbolic lexicon) (also see section 5). But I have shown that the similar process of pitch or register raising for emphasis also occurs in other African and Caribbean English-lexifier creoles, see 2.1.2–2.2.3 and 3.1–3.3.

The presence of tone-conditioned suppletive allomorphy in Pichi is an example of the idiosyncratization of grammatical tone (see section 2.1.2), even if the conditioning feature of an obligatory tonal contour over adjacent tone-bearing units draws on an areally widespread model (Yakpo, 2019b, 217–218).

Contact tone systems can only incorporate the material provided by their input languages during stress-to-tone mapping. In combination with the two other mechanisms of paradigmatization and idiosyncratization, prosodic contact nevertheless leads to autonomous outcomes, particularly in the creoles in contrast to the European colonial varieties (see section 5.2).

TABLE 6 | Tonal idiosyncratization in lexical words (Krio and Pichi).

	Item	Tone pattern	Example
(1)	*lílí(lí)* 'little, tiny'	/!H!H/	*dì wàtá tú **lílílí*** 'the water (is) too little'
(2)	*bíbí* 'big, huge'	/!H!H/	*wán **bíbí** hós* 'a huge house'
(3)	*sótéé* 'excessively'	/!H!H!H/	*à rɔ́n **sótéé*** 'I ran excessively'
(4)	*sósó* 'only'	/!H!H/	***sósó** mɔ́nìn tɛ́n* 'really early in the morning'

5. CREOLE PROSODIC SYSTEMS ARE AREAL, NOT SIMPLE

I now revisit arguments for claims that tone is eliminated or simplified in creoles (see section 1). A comparison with tonal African non-creoles allows the conclusion that the tone systems of Afro-European creoles are neither particularly simple nor typologically divergent in other ways (section 5.1). Instead, creole prosody undergoes regular typological change and areal convergence (section 5.2).

5.1. Creole Tone Systems Are No Simpler Than African Non-creole Tone Systems

The hypothesis that creolization involves prosodic simplification has at least two subjacent assumptions: (i) Creolization is seen to involve the elimination of features that are difficult for adult learners (for a thematic overview, see Siegel, 2004), and tone is viewed as 'particularly hard to master during untutored second language acquisition' (Sessarego, 2020, 4).

This assumption is logically flawed, for it does not distinguish between the two psycholinguistic dominance relations of recipient language agentivity and source language agentivity (see Bordal Steien and Yakpo, 2020), and the two corresponding transfer types of borrowing and imposition (van Coetsem, 1988, 2000). Phonology (van Coetsem, 2000), and prosody in particular (Matras, 2009, 231–33), are the most stable domains of a natively acquired grammar. So if any domain gets transferred to a contact language from its input languages at all, *it will be prosody,* and in ecologies dominated by tone speakers, *it will be tone* (Bordal Steien and Yakpo, 2020). The assumption of tone loss is also

Eurocentric because it takes for granted that stress systems, which happen to characterize all European lexifiers and superstrates, constitute the fallback during prosodic contact.

(ii) Creoles can have tone systems, but these systems are assumed to be simpler than non-creole tone systems (see the sources cited in the opening paragraph of section 1). Assumption (ii) deserves some attention. In the face of irrefutable evidence for the existence of creole tone systems, it is less categorical than the tone loss hypothesis, yet aligns with the 'creoles-are-simpler-than-other-languages' hypothesis (for an overview and a critique, see Ansaldo and Matthews, 2007).

In the following, I address the characteristics of tone in Afro-European contact prosodic systems with respect to: (1) tonal inventories (i.e., the number of distinctive tones), (2) the existence of tonal minimal pairs, (3) the number of word-tone patterns, (4) and the nature of tonal processes and rules, comparing these with features of tonal non-creole languages of Africa.

Tone systems with two heights, whether equipollent /H, L/ or privative /H, Ø/, as in the Afro-European creoles and colonial varieties are the most common ones across a huge swath of West Africa and all of West Central Africa as far south as Angola (Wedekind, 1985; Hyman et al., 2020). These were the principal home regions to the millions of Africans enslaved and deported to the Americas by the Europeans (Eltis, 2001). Their languages therefore constituted the most important substrates to the creoles and colonial varieties that developed in the Americas and in the West African littoral region. A two-way contrast is also cross-linguistically the most common type beyond Africa (Maddieson, 2013) and tonogenesis almost always produces a binary contrast between two tone heights (Hyman et al., 2020).

Further, a significant number of relevant African substrate and adstrate languages feature restricted three-height systems termed 2T3 height systems by Hyman (2018, 208) (2 input vs. 3 output heights, e.g., /H, L, Ø/). In such systems, the presence of an additional (e.g., mid or extra-high) tone is, for example, conditioned through a constructional tone rule or the segmental structure of its tone-bearing unit and therefore predictable and not lexical in *sensu stricto* (as in Yoruba, see Akinlabi, 1985).

The conventionalization of extra-high tone in Krio and Pichi degree-modifying words is an example of such a restricted use of a third height in the creoles (see section 4.3). In contrast, systems with three and more *lexical* tones have marked regional distributions. In West Africa, these are principally found in the Kru languages and the adjacent contact zones, the Mabia (Gur) languages, and the Nigerian-Cameroonian plateau region. None of these areas were pre-dominant home regions of enslaved Africans as far as the historical records are concerned (see Eltis, 2001).

A second argument encountered for classifying creole and other contact prosodic systems as simpler is the supposedly low number of tonal minimal pairs. This is claimed by McWhorter (1998), an assertion contradicted in later work by the same author's listing of over twenty tonal minimal pairs in Saramaccan, many of which are multisyllabic (McWhorter and Good, 2012, 39). Matisoff (2001, 304) argues that largely monosyllabic languages are particularly 'tone (and phonation) prone' for marking semantic and grammatical distinctions, due to the sparsity and interdependence of segmental material in the word.

Conversely, the functional load of tone in the lexicon will be reduced in languages with longer words. It therefore comes as no surprise that tone languages in which two or more syllables predominate should be characterized by fewer tonal minimal pairs in the lexicon than logically possible. This is the case in Akan (Kwa, Ghana), for example, (e.g., *pàpá* 'father', *pápá* 'good', *pàpà* 'fan', **pápà*; own knowledge), the Narrow Bantu languages (Hyman et al., 2020), Cushitic (Mous, 2009), Chadic (Rolle, 2018), and, last but not least, the Afro-European creoles and colonial varieties.

Thirdly, African tone languages of all linguistic lineages feature restrictions on the distribution of tones and the number of word-tone patterns. Susu (Mande, Guinea) only has the three tone patterns H, H-L, and L-H over mono- and disyllabic nouns (nouns with more syllables are not very common) (Green et al., 2013). Two tone patterns, H and L-H, cover ninety per cent of the lexicon of Bambara (Mande, Mali) (Dumestre, 2003, 22).

Further, it is very common for African substrates and adstrates of Afro-European creoles to have lexical strata displaying specific prosodic behaviors, often as a result of the imposition of native prosody on loan lexicon, just like in the creoles. In the two-height system of Mende (Mande, Sierra Leone), loanwords bear a cumulative H on the penultimate syllable, irrespective of the original position of the H-toned or stressed syllable in the source language. Native vocabulary is not subject to such a restriction (Leben, 1977; cited in Clements and Ford, 1979:201). In the three-height tone system of Ewe (Kwa, Ghana, and Togo), European loanwords have a cumulative H on the stress-bearing syllable in the source language, other syllables are L-toned, e.g., *àbólò* 'bread' (<Port. ***bo****lo* 'cake', incl. nominal prefix *à-*), *àkɔ́ntà* /L-**H**-L/ 'arithmetic' (<Port. ***co****nta* 'account'), *sùkúù* /L-**H**-L/ 'school' (<Eng. *school*, incl. epenthetic vowel /ù/), ***dú****kù* /**H**-L/ 'scarf' (<Dutch *doek* 'scarf', incl. paragogic vowel /ù/). The tonal integration of European loanwords in Akan (Apenteng and Amfo, 2014) and Gã (Kropp Dakubu, 1999), two other languages of the Ghanaian littoral zone, proceeds along similar lines.

Kikongo (Narrow Bantu, Congo, and DRC) has a privative /H, Ø/ system and H tone is obligatory and cumulative in much of the lexicon. French loans bear a word-final cumulative H due to stress-to-tone mapping, just as in Central African French, e.g., *kàdó* /Ø-**H**/ 'present' (<Fr. *cad****eau***), *prɛ̀zì****dɛ́*** /Ø-Ø-**H**/ 'president' (<Fr. *prési****dent***). By contrast, word-final H is rare in the African-sourced lexicon, compare *ndúùmbù* /H-Ø-Ø/ 'spices', *èntsàngálà* /Ø-Ø-H-Ø/ 'basket' (Donnelly, 1982).

Besides loanwords, ideophones also display special phonological (and morphosyntactic) characteristics cross-linguistically (Dingemanse, 2018). In African non-creoles and creoles alike, ideophones exhibit additional tone heights (e.g., extra-high) and tone types (e.g., contour tones instead of level tones alone), as well as idiosyncratic tone patterns. In the 2-tone language Temne (Mel, Sierra Leone), ideophones generally take an H or LH (composite) tone where other word classes take an L or HL tone. This renders tonal minimal pairs like *gbâŋ*

/HL/ 'manufacture a tool' vs. *gbǎŋ* /LH/ 'extremely old (IDEO)' (Kanu, 2008).

In the two-tone system of Kisi (Atlantic, spoken in Guinea and Sierra Leone), contour tones are not restricted to final syllables in ideophones and are also found on initial syllables, unlike in other word classes, e.g., *kpîngmgbí* /HL-H/ 'darkly' (IDEO) (Childs, 1988). There are also many more H tones and extra-H tones than L tones in Kisi ideophones, which is the opposite of the distribution in non-ideophonic word classes, e.g., *kpáng* /!H/ 'tightly, carefully (IDEO)'.

Beyond the ideophonic lexicon, restrictions on the number, types and position of tones are common throughout tonal Africa (for an overview, see Downing, 2010). This is again no different from the creoles. A good many Bantu tone systems are characterized as sparse because they feature a privative /H, Ø/ contrast. The single /H/ toneme is obligatory and culminative, and its position predictable (Odden, 1988; Gussenhoven, 2004, 34–35; Hyman, 2011b, 235). But restrictions are also found in systems with full specification of lexical tones. In Akan, vowel height and the place of articulation of consonants determine the distribution of H and L tones in disyllabic verbs, e.g., *pìrá* /L-H/ 'hurt' (first vowel is high and followed by a sonorant) vs. *kásà* /H-L/ 'speak' (first vowel is non-high) (Welmers, 1973, 118).

A fourth argument for simplification is the claim that there is no grammatical tone in creoles (McWhorter, 2005, 13–14). 'Grammatical tone' is poorly delimited in the first place, making it difficult to distinguish from tone sandhi phenomena, intonation, phrasal tonology, even lexical tone (Rolle, 2018, 3), and the terminology is unclear (Rolle, 2020, 71). If we assume the definition of grammatical tone as a tonological operation restricted to the context of a specific morpheme or construction (Rolle, 2018), there is a vast range of phenomena that can be subsumed under the label in Africa (e.g., polar tone in Yoruba, see Akinlabi and Liberman, 2000; and constructional tone in Kalabari, see Harry and Hyman, 2014). Some African tone languages have much grammatical tone, others less (Hyman et al., 2020). In comparison, tonal Afro-European creoles neither feature abundant nor particularly sparse grammatical tone. Uses discussed in this study include the expression of pronominal case functions by tonal minimal pairs in all tonal English-lexifier creoles, e.g., in Pichi (section 2.1.2) and in Saramaccan (McWhorter and Good, 2012, 42).

Tonal inflection in the pronominal paradigm in addition to segmental inflection for number and case is common in West African tone languages, including Ewe (Duthie, 1996, 53), Akan (Dolphyne, 1996, 109–110), and Edo (Uchihara, 2010). I have shown that creole grammatical tone also extends to tonal derivation in compounding and reduplication, constructional tone and OCP, as well as tone fusion and contour tone formation (see section 2.1.2).The creoles also have phrasal tone rules of tone spreading, downdrift, dissimilation, assimilation, and tone insertion, all of which are shared with African tone languages (Hyman and Schuh, 1974; Yip, 2002). These rules have been investigated in detail for creoles like Pichi (see section 2.1 and the references there), Saramaccan (Good, 2004, 2006, 2009), and Lung'Ie (Agostinho and Hyman, 2021, 77–80). There is good reason to assume that similar rules exist in other tonal Afro-European creoles of Africa and the Americas.

In sum, the prosodic systems of Afro-European creoles do not differ from the systems of African non-creole languages in any consistent and sufficient way to qualify them as simpler in terms of tonal inventories, the existence of tonal minimal pairs, the number of tone patterns, and the nature of tonal processes and rules. Instead, creole tone systems fit snugly into the areal patterns attested in countless variations throughout tonal Africa.

5.2. Creole Prosodic Systems Undergo Regular Typological Change and Areal Convergence

Contact prosodic systems acquire their properties from a typological matching exercise between features of the input languages in a specific linguistic ecology (Mufwene, 1996, 2001; Aboh and Ansaldo, 2007). The acoustic and phonological realizations of tone imposed by the adstrates and substrates can initially only be grafted on the prosodic patterns of the lexical material provided by lexifier stress patterns.

The more typologically compatible the input systems are, the more tonal features will be utilizable. In other words, if a creole evolved from contact between the Mande language Dan (Côte d'Ivoire, Liberia) with 5 tone heights and the neighboring Kru language Krahn with 3 tone heights, tone-to-tone mapping should allow a higher number of tonal features to be utilized. Claims that tone is reduced in function or lost in African contact languages that have emerged from contact between *tone languages* should be taken with a grain of salt, due to the absence of detailed analyses (e.g., Heine, 1978, 221 and the sources cited there). The only comprehensive study of the prosody of the Kikongo cluster (Congo, RDC, Angola), for example, concludes that the Bantu-based contact language Kikongo-Kituba 'has merely accelerated certain tendencies inherent in the Kikongo pitch system' (Donnelly, 1982, 343).

The prosodic systems that emerge under the typological constraints of stress-to-tone mapping will therefore not look like the tone system of a language like Guébie (Kru, Côte d'Ivoire), with its above-average number of tonal features utilized by West African standards. Guébie has five contrasting tone levels and abundant grammatical tone (Sande, 2018). Clements and Rialland (2007, 70–74) argue that in all areas of West and East Africa with a similarly high functional load of tone, lexical and grammatical tone was not borrowed from neighboring languages. Instead there was an areal diffusion of monosyllabicity, the structural prerequisite for such a functional proliferation of tone. None of the tonal creoles of Africa have a predominantly monosyllabic template, so it is not surprising that they have 2T (two heights) or 2T3 (2 input, 3 output heights), not five-tone systems. However, there is no reason to assume that diachronic change of a creole through areal pressure toward monosyllabicity should not produce additional tone heights and more active grammatical tone than is already the case.

Afro-European creoles are 'late arrivals' in their respective linguistic ecologies in the sense that they owe part of their lineage

to exogenous Indo-European lexifiers. The prosodic systems of Afro-European creoles therefore continue to align themselves over time with other languages in their respective ecologies according to regular areal dynamics (for an overview of these dynamics, see Raymond, 2017, 3–5). Since convergence takes time, typological inconsistencies may persist in the prosodic systems of creoles vis-à-vis those of their tonal (Africa) or stress-only (Americas) areal cohabiters.

For example, a large part of the lexicon of Ghanaian Pidgin is sourced from English and therefore features a cumulative H due to stress-to-tone mapping. By contrast, the Ghanaian Pidgin adstrates Akan, Gã, and Ewe have a much smaller European-sourced lexicon with a cumulative H (see section 5.1). Such an inconsistency between Ghanaian Pidgin and the non-creoles in the Ghanaian ecology is not caused by creole distinctiveness, but by differences in the size of the European-derived prosodic lexicon. Likewise, the persistence of (residual) tone in Caribbean creoles and areally unusual prosodic layering in an African creole like Lung'Ie (African-sourced words are toneless) means that these languages have not (yet) fully aligned themselves with their adstrates and/or superstrates. These differences are gradual, and can progressively narrow down due to continuing areal convergence, or stabilize as innovations that bring additional typological diversity to an ecology.

Such a role of time depth in the areal diffusion of prosodic features is fundamentally different from the idea that creoles are young and have not yet had time to accumulate (tonal and morphological) complexity in their grammars (e.g., Bickerton, 1988, 274–278; Lightfoot, 2006, 7; McWhorter, 2007, 4–5). The latter view is a trope rooted in 19th century linguistic evolutionism (for an epistemological deconstruction, see Krämer, 2013; McElvenny, 2021). The results of this study instead suggest that creoles and colonial varieties undergo regular cycles of shift from one part of the typological spectrum (e.g., tone) to another (e.g., stress) and vice versa, without an *a priori* assumption of simplification or complexification.

In the scenarios covered here, such shifts are contact-driven and result in prosodic convergence between unrelated and typologically dissimilar languages cohabiting the same ecology, and prosodic divergence between related languages inhabiting different ecologies. Prosodic convergence and divergence are reflected particularly well in the range of prosodic systems found in the family of Afro-Caribbean English-lexifier Creoles, with its large geographical spread across diverse linguistic ecologies in Africa and the Americas.

6. THE IDEA OF CREOLE SIMPLICITY IS A CHIMERA

The prosodic systems of Afro-European creoles and colonial varieties form an areal continuum across the Afro-Atlantic from Africa to the Americas, roughly corresponding to tone in the east and stress in the west. Transitional systems are found in the areal buffer zone of the Caribbean, where tone and stress-only systems have converged in various ways. Numerous pitch-related phenomena found in American creole languages and colonial varieties of the Americas that exclusively feature stress today suggest the existence of tone or mixed tone-stress systems before the shift to stress-only systems.

I have identified and described three mechanisms involved in the emergence of contact prosodic systems with potential for generalization to prosodic contact scenarios beyond the Afro-Atlantic. These are stress-to-tone mapping, paradigmatization, and idiosyncratization. The label 'simple' neither captures these mechanisms themselves, nor the tone systems they engender.

The argument that creoles are simpler than non-creoles is based on the notion of 'bit complexity' (DeGraff, 2001, 284–285), which boils down to 'more overt material = more complex'. Bit complexity has been criticized as a simplistic and arbitrary criterion of little heuristic value for measuring linguistic complexity (Aboh and DeGraff, 2017, 14–20; Newmeyer, 2021).

Nevertheless, even from the perspective of bit complexity, *creole tone systems are more complex* than those of the colonial varieties of English, French, Spanish, and Portuguese spoken in Africa (section 3.1 and section 4). This is due to social factors that impede the same amount of innovation and areal diffusion of tonal features to the colonial varieties as to the creoles. The colonial varieties are heavily standardized, are usually acquired in classrooms, and are predominantly used in formal settings. The proportion of speakers who regularly use the colonial varieties is small and limited to social classes with access to secondary and tertiary education (for French in Africa, see e.g., Mufwene, 2011). The natural evolution of European colonial varieties, including that of their prosodic systems, is therefore severely constrained (Yakpo, 2020, 133–134). By contrast, the creoles have been evolving without state-sanctioned standardization and are primarily spoken by urban working class and rural populations, many of whom have little formal education and limited exposure to the colonial varieties. The creoles could therefore acquire many more autonomous prosodic features through areal diffusion from substrate and adstrate languages than the colonial varieties.

The idea of creole simplicity is a chimera. Research should rather focus on the roles played by genealogy, areal typology, cognition, and social factors in shaping the fascinating diversity of specific language contact outcomes, as I have attempted here with respect to prosodic 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

The studies involving human participants were reviewed and approved by the Human Research Ethics Committee, The University of Hong Kong. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

KY designed the work, collected and analyzed the primary data, assembled secondary data from other sources, conducted the qualitative analyses, and wrote the manuscript.

FUNDING

Research for this study was conducted with the support of the Research Grants Council, University Grants Committee of the Government of Hong Kong (GRF 17608819) and the Seed Fund for Basic Research (grant no. 201910159234) of the University of Hong Kong. This manuscript was written during a Humboldt Research Fellowship for Experienced Researchers (2020–2021) at the Humboldt University of Berlin.

ACKNOWLEDGMENTS

I am indebted to the two reviewers of this article as well as Ana Lívia Agostinho, Jeff Good, Shelome Gooden, Carlos Gussenhoven, Larry Hyman, Douglas Pulleyblank, and Guri Steien for their generous comments on a draft version of this article. I thank Hubert Devonish for providing crucial linguistic data on Guyanese Creole and Joseph Farquharson for sending me vital literature from his library. I am also grateful to Isabel Wetzel, Evans Koskei, Violet Akoth Owiso, and Jeremias Koskei for creating the enchanting environment in which this piece could be written.

REFERENCES

Aboh, E. A., and Ansaldo, U. (2007). "The role of typology in language creation: a descriptive take," in *Deconstructing Creole* Typological Studies in Language, eds U. Ansaldo, S. Matthews, and L. Lim (Amsterdam: John Benjamins), 39–66. doi: 10.1075/tsl.73.05abo

Aboh, E., and DeGraff, M. (2017). A null theory of creole formation based on Universal Grammar. *Oxf. Handb. Univers. Gramm.* 2017, 401–458.

Agostinho, A. L., and Hyman, L. M. (2021). Word Prosody in Lung'Ie: One System or Two? *Probus* 33, 57–93. doi: 10.1515/prbs-2021-0002

Agostinho, A. L., Araujo, G. A., de, Santos, E. F., and dos. (2019). Partícula interrogativa e pitch-accent frasal nas perguntas polares em fa d'Ambô. *Bol. Mus. Para. Emílio Goeldi Ciênc. Humanas* 14, 857–872. doi: 10.1590/1981.81222019000300009

Akinlabi, A. M. (1985). *Tonal Underspecification and Yoruba Tone*. Ibadan: University of Ibadan.

Akinlabi, A., and Liberman, M. (2000). "The tonal phonology of Yoruba clitics," in *Clitics in phonology, morphology and syntax*, eds B. Gerlach and J. Grizenhout (Amsterdam: John Benjamins), 31–62. doi: 10.1075/la.36.04aki

Alleyne, M. C. (1980). *Comparative Afro-American: an Historical-comparative Study of English-based Afro-American dialects of the New World*. Ann Arbor: Karoma Publishers.

Ansaldo, U., and Matthews, S. (2007). "Deconstructing creole: The rationale," in *Deconstructing Creole Typological Studies in Language*, eds U. Ansaldo, S. Matthews, and L. Lim (Amsterdam: John Benjamins), 1–18. doi: 10.1075/tsl.73.02ans

Ansaldo, U., Matthews, S., and Lim, L. (eds) (2007). *Deconstructing Creole*. Amsterdam: John Benjamins.

Apenteng, M. A., and Amfo, N. A. A. (2014). The form and function of English loanwords in Akan. *Nord. J. Afr. Stud.* 23, 22–22.

Appleyard, D. L. (1991). The role of tone in some Cushitic languages. *York Pap. Linguist.* 15, 5–32.

Barth, E. (2016). Tone loss in Sranan Creole: re-thinking contact driven change. *Pap. Hist. Phonol.* 1, 76–89. doi: 10.2218/pihph.1.2016.1694

Berry, J. (1970). A Note on the Prosodic Structure of Krio. *Int. J. Am. Linguist.* 36, 266–267. doi: 10.1086/465120

Berry, J. (1972). "Some observations on "residual tone" in West Indian English," in *Paper presented at the Conference on Creole Languages and Educational Development*, (Trinidad: University of the West Indies).

Biagui, N. B. (2012). *Description générale du créole afro-portugais parlé à Ziguinchor (Sénégal)*. Dakar: Universite Cheikh Anta Diop.

Bickerton, D. (1988). "Creole languages and the bioprogram," in *Linguistics: The Cambridge survey. Vol II: Linguistics Theory: Extensions and Implications*, ed. F. Newmeyer (Cambridge: Cambridge University Press), 268–284. doi: 10.1017/cbo9780511621055.016

Bloomquist, J., Green, L. J., Lanehart, S. L., and Thomas, E. R. (2015). "Prosodic Features of African American English," in *The Oxford Handbook of African American Language*, eds J. Bloomquist, L. J. Green, and S. L. Lanehart (Oxford: Oxford University Press), doi: 10.1093/oxfordhb/9780199795390.013.12

Bordal Steien, G. (2015). "Traces of the lexical tone system of Sango in Central African French," in *Prosody and Language in Contact*, ed. E. Delais-Roussarie (Heidelberg: Springer), 29–49. doi: 10.1007/978-3-662-45168-7_3

Bordal Steien, G., and Yakpo, K. (2020). Romancing with tone: On the outcomes of prosodic contact. *Language* 96, 1–41. doi: 10.1353/lan.2020.0000

Bordal, G., and Skattum, I. (2014). "La prosodie du français en Afrique – traits panafricains ou traits de la langue première? Le cas de locuteurs natifs de quatre langues: sango, bambara, wolof et tamasheq," in *La phonologie du français contemporain: normes, périphéries, modélisation. Mélanges pour Chantal Lyche*, eds J. Durand, G. Kristoffersen, B. Laks, and J. Peuvergne (Paris: Presses universitaires de Paris Ouest), 119–152.

Boutin, B. A., and Turscan, G. (2009). "La prononciation du français en Afrique: la Côte d'Ivoire," in *Phonologie, variation et accents du français*, eds J. Durand, B. Laks, and C. Lyche (Paris: Hermès), 131–151.

Brousseau, A.-M. (2003). "The accentual system of Haitian Creole: The role of transfer and markedness values," in *Phonology and morphology of creole languages*, ed. I. Plag (Tübingen: Max Niemeyer Verlag), 123–146.

Bruce, G. (2004). "An intonational typology of Swedish," in *Speech Prosody 2004*, (Japan: Nara University).

Bull, T. (1995). Language maintenance and loss in an originally trilingual area in North Norway. *Int. J. Sociol. Lang.* 115, 125–134.

Butera, B., Rao, R., and Sessarego, S. (2020). "Afro-Peruvian Spanish intonation: A case of contact-induced language change," in *Issues in Hispanic and Lusophone Linguistics*, ed. R. Rao (Amsterdam: John Benjamins), 207–226. doi: 10.1075/ihll.28.08but

Cadely, J.-R. (1994). *Aspects de la phonologie du créole haïtien*. Saarbrücken: universitaires européennes.

Chapouto, S. M. C. (2014). *Contributo Para a Descrição de Aspetos Fonológicos e Prosódicos Do Crioulo Guineense*. Coimbra: University of Coimbra.

Childs, T. G. (1988). The Phonology of Kisi ideophones. *J. Afr. Lang. Linguist.* 10, 165–190.

Ciszewski, T. (1999). The English stress system: conditions and parameters. *Acta Neophil.* 1999, 24–34.

Clements, G. N., and Ford, K. C. (1979). Kikuyu tone shift and its synchronic consequences. *Linguist. Inq.* 10, 179–210.

Clements, G. N., and Rialland, A. (2007). "Africa as a phonological area," in *A linguistic geography of Africa*, eds B. Heine and D. Nurse (Cambridge: Cambridge University Press), 36–85. doi: 10.1017/cbo9780511486272.004

Criper-Friedman, L. (1990). The tone system of West African coastal English. *World Englishes* 9, 63–77.

DeGraff, M. (2001). On the origin of creoles: A Cartesian critique of neo-Darwinian linguistics. *Linguist. Typol.* 5, 213–310.

Devonish, H. (1989). *Talking in tones: A study of tone in Afro-European Creole languages*. Kingston: Caribbean Academic Publications.

Devonish, H. (2001). "On the Sierra Leone—Caribbean connection: Hot on the trail of 'tone-shifted items' in Anglo-West African," in *Due Respect: Essays on English and English-related Creoles in the Caribbean in Honour of Professor Robert Le Page*, ed. P. Christie (Kingston: University of the West Indies Press), 184–205.

Devonish, H. (2002). *Talking rhythm stressing tone: the role of prominence in Anglo-West African Creole languages*. Kingston: Arawak Publications.

Devonish, H., and Thompson, D. (2010). *A concise grammar of Guyanese Creole (Creolese)*. München: Lincom Europa.

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa J. Gen. Linguist.* 3:4. doi: 10.5334/gjgl.444

Dolphyne, F. A. (1996). *A comprehensive course in Twi (Asante), for the non-Twi learner*. Accra: Ghana Universities Press.

Donnelly, K. G. (1982). *A Comparative Study of Tone and Intonation in Seven Kongo Dialects*. Ph. D. thesis. London: SOAS University of London.

Downing, L. J. (2010). "Accent in African Languages," in *A Survey of Word Accentual Patterns in the Languages of the World*, eds H. van der Hulst, R. Goedemans, and E. van Zanten (Berlin: De Gruyter), 381–428. doi: 10.1515/9783110198966.1.381

Downing, L. J., and Rialland, A. (2016). *Intonation in African Tone Languages*. Berlin: De Gruyter Mouton, doi: 10.1515/9783110503524

Dumestre, G. (2003). *Grammaire fondamentale du bambara*. Paris: Ed. Karthala.

Duthie, A. S. (1996). *Introducing Ewe linguistic patterns: a textbook of phonology, grammar, and semantics*. Accra: Ghana Universities Press.

Eltis, D. (2001). The Volume and Structure of the Transatlantic Slave Trade: A Reassessment. *William Mary Q.* 58:17. doi: 10.2307/2674417

Faraclas, N. G. (1991). "The pronoun system in Nigerian Pidgin: a preliminary study," in *English around the world: the social contexts*, ed. J. Cheshire (Cambridge: Cambridge Univ. Press), 509–518. doi: 10.1017/cbo9780511611889.035

Faraclas, N. G. (1996). *Nigerian Pidgin*. London: Routledge.

Ferreira, J.-A. S. (2006). Madeiran Portuguese Migration to Guyana, St. Vincent, Antigua and Trinidad: A Comparative Overview. *Port. Stud. Rev.* 14, 63–85.

Finney, M. A. (1993). "Tone assignment and compound formation in Krio," in *Actes du congrès annuel de l'ACL (Association Canadienne de Linguistique) 1993 / Proceedings of the 1993 CLA (Canadian Linguistics Association) annual conference* Toronto working papers in linguistic, ed. C. Dyck (Toronto: University of Toronto).

Finney, M. A. (2004). "Tone Assignment on Lexical Items of English and African Origin in Krio," in *Creoles, Contact and Language Change: Linguistics and Social Implications*, eds G. Escure and A. Schwegler (Amsterdam: John Benjamins), 221–236. doi: 10.1075/cll.27.11fin

Frota, S., and Vigário, M. (2000). "Aspectos de prosódia comparada: ritmo e entoação no PE e no PB," in *Actas do XV Encontro da Associação Portuguesa de Linguística*, ed. Associação Portuguesa de Linguística (Braga: APL), 533–555.

Good, J. (2009). A twice-mixed creole? Tracing the history of a prosodic split in the Saramaccan lexicon. *Stud. Lang.* 33, 459–498. doi: 10.1075/sl.33.2.09goo

Good, J. C. (2004). Tone and accent in Saramaccan: Charting a deep split in the phonology of a language. *Lingua* 2004, 575–619. doi: 10.1016/s0024-3841(03)00062-7

Good, J. C. (2006). "The phonetics of tone in Saramaccan," in *Structure and variation in language contact*, eds A. Deumert and S. Durrleman-Tame (Amsterdam: John Benjamins), 9–28. doi: 10.1075/cll.29.03goo

Gooden, S. (2003). *The phonology and phonetics of Jamaican Creole reduplication*. Columbus: Ohio State University.

Gooden, S., Drayton, K.-A., and Beckman, M. (2009). Tone inventories and tune-text alignments: Prosodic variation in 'hybrid' prosodic systems. *Stud. Lang.* 33, 396–436. doi: 10.1075/sl.33.2.07goo

Granda, G (1986). Retenciones africanas en la fonética del criollo portugues de annobón. *Rev. Filol. Románica* 4, 111–123. doi: 10.2307/j.ctv1ddctnf.7

Green, C. R., Anderson, J. C., and Obeng, S. G. (2013). Interacting tonal processes in Susu. *Mandenkan* 50, 61–84. doi: 10.4000/mandenkan.876

Güldemann, T. (2018). "Areal linguistics beyond contact, and linguistic areas of Afrabia," in *The languages and linguistics of Africa* The World of Linguistics (WOL), ed. T. Güldemann (Berlin: De Gruyter Mouton), 448–545. doi: 10.1515/9783110421668-004

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2006). Between stress and tone in Nubi word prosody. *Phonology* 23, 192–223. doi: 10.1017/s0952675706000881

Gussenhoven, C., and Udofot, I. (2010). "Word melodies vs. pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English," in *Proceedings of Speech Prosody 2010* (Chicago).

Gut, U., Urua, E.-A., Adouakou, S., and Gibbon, D. (2001). "Rhythm in West African tone languages: a study of Ibibio, Anyi and Ega," in *Proceedings of Typology of African Prosodic Systems 2001* (Bielefeld: Bielefeld University), 159–165.

Hackert, S. (2019). The perfect in English-lexifier pidgins and creoles: A comparative study. *J. Pidgin Creole Lang.* 34, 195–242. doi: 10.1075/jpcl.00039.hac

Hancock, I. F. (1971). *A study of the sources and development of the lexicon of Sierra Leone Krio*. Ph. D. thesis. London: SOAS University of London.

Hancock, I. F. (1986). "The domestic hypothesis, diffusion and componentiality: an account of Atlantic Anglophone creole origins," in *Substrata versus universals in creole genesis*, eds P. Muysken and N. Smith (Amsterdam: John Benjamins), 71–102. doi: 10.1075/cll.1.06han

Hancock, I. F. (1987). "A preliminary classification of Anglophone Atlantic creoles, with syntactic data from thirty-three representative dialects," in *Pidgin and creole languages: essays in memory of John Reinecke*, ed. G. G. Gilbert (Honolulu: Univ. of Hawai'i Press), 264–333.

Harry, O. G., and Hyman, L. M. (2014). Phrasal Construction Tonology: The case of Kalabari. *Stud. Lang.* 38, 649–689. doi: 10.1075/sl.38.4.01hym

Heine, B. (1978). Some generalizations on African-based pidgin languages. *Afr. Übersee* 61, 219–229.

Hualde, J. I., and Prieto, P. (2015). "Intonational variation in Spanish: European and American varieties," in *Intonational variation in Romance*, eds S. Frota and P. Prieto (Oxford: Oxford University Press), 350–391. doi: 10.1093/acprof:oso/9780199685332.003.0010

Hualde, J. I., and Schwegler, A. (2008). Intonation in Palenquero. *J. Pidgin Creole Lang.* 23, 1–31. doi: 10.1075/jpcl.23.1.02hua

Huber, M. (1999). *Ghanaian Pidgin English in its West African context: a sociohistorical and structural analysis*. Amsterdam: John Benjamins.

Huber, M. (2003). "Verbal reduplication in Ghanaian Pidgin English: origins, forms and functions," in *Twice as meaningful: reduplication in pidgins, creoles, and other contact languages*, ed. S. Kouwenberg (London: Battlebridge Publ), 139–154.

Huttar, G. L., and Huttar, M. L. (1994). *Ndyuka*. London: Routledge.

Hyman, L. H. (2001). "Tone systems," in *Language Typology and Language Universals. An International Handbook. Vol. 2 Handbücher zur Sprach- und Kommunikationswissenschaft*, eds M. Haspelmath, E. König, W. Oesterreicher, and W. Raible (Berlin: Walter de Gruyter), 1367–1380.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology* 23, 225–257. doi: 10.1017/S0952675706000893

Hyman, L. M. (2011a). Markedness, faithfulness, and the typology of two-height tone systems. *UC Berkeley PhonLab Annu. Rep.* 7, 189–199.

Hyman, L. M. (2011b). "Tone: Is it different?" in *The handbook of phonological theory*, eds J. A. Goldsmith, J. Riggle, and A. C. L. Yu (Malden, MA: Wiley-Blackwell), 197–239.

Hyman, L. M. (2018). "Towards a Typology of Tone System Changes," in *Tonal Change and Neutralization*, eds H. Kubozono and M. Giriko (Berlin: De Gruyter Mouton), 203–222. doi: 10.1515/9783110567502-008

Hyman, L. M., and Schuh, R. G. (1974). Universals of tone rules: evidence from West Africa. *Linguist. Inq.* 5, 81–115.

Hyman, L., Sande, H., Lionnet, F., Rolle, N., and Clem, E. (2020). "Sub-Saharan Africa," in *The Oxford Handbook of Language Prosody*, eds C. Gussenhoven and A. Chen (Oxford: Oxford University Press), 183–194. doi: 10.1093/oxfordhb/9780198832232.013.11

Jacobs, B. (2010). Upper Guinea Creole: Evidence in favor of a Santiago birth. *J. Pidgin Creole Lang.* 25, 289–343. doi: 10.1075/jpcl.25.2.04jac

Jahr, E. H. (1984). Language contact in Northern Norway: Adstratum and substratum in the Norwegian, Sami and Finnish of Northern Norway. *Acta Boreal.* 1, 103–112.

James, W. (2003). "The role of tone and rhyme structure in the organisation of grammatical morphemes in Tobagonian," in *Phonology and Morphology*

of Creole Languages, ed. I. Plag (Tübingen: Max Niemeyer Verlag), 165–192. doi: 10.1515/9783110929560.165

Joubert, S. M. (1991). *Dikshonario Papiamentu-Hulandes–Handwoordenboek Papiaments-Nederlands*. Willemstad: Fundashon di Leksikografia.

Jun, S.-A., and Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus* 14, 147–172.

Kalkhoff, A. M. T. (2018). *Using corpus data for testing a working hypothesis about Haitian Creole prosody*. Potsdam: University of Potsdam.

Kanu, S. M. (2008). Ideophones in Temne. *Kans. Work. Pap. Linguist.* 30, 120–134.

Kouwenberg, S. (2004). The Grammatical Function of Papiamentu Tone. *J. Port. Linguist.* 3, 55–69. doi: 10.5334/jpl.11

Krämer, P. (2013). Creole exceptionalism in a historical perspective – from 19th century reflection to a self-conscious discipline. *Lang. Sci.* 38, 99–109. doi: 10.1016/j.langsci.2013.02.003

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kropp Dakubu, M. E. (ed.) (1999). *Ga-English Dictionary with English-Ga Index*. Accra: Black Mask.

Kusters, W. (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.

Lang, J. (2009). *Les langues des autres dans la créolisation: théorie et exemplification par le créole d'empreinte wolof à l'île Santiago du Cap Vert*. Tübingen: Gunter Narr Verlag.

Lawton, D. L. (1963). *Suprasegmental Phenomena in Jamaican Creole*. Michigan: State University.

Leben, W. R. (1973). *Suprasegmental phonology*. Spokane, WA: Garland Pub.

Leben, W. R. (1977). *Studies in the suprasegmental representation of tone*. Cambridge, MA: MIT Press.

Lightfoot, D. (2006). *How New Languages Emerge*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511616204

Lim, L. (2012). Revisiting English prosody (some) New Englishes as tone languages? *English World-Wide* 30, 218–239. doi: 10.1075/eww.30.2.06lim

Lipski, J. M. (2007). *Castile and the hydra: the diversification of Spanish in Latin America*. Pennsylvania,PA: The Pennsylvania State University.

Macedo, D. P. (1979). *A linguistic approach to the Capeverdean language*. Michigan: University Microfilms.

Maddieson, I. (2005). "Tone," in *The World Atlas of Language Structures*, eds M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (Oxford: Oxford University Press), 58–61.

Maddieson, I. (2013). "Tone," in *The World Atlas of Language Structures Online (WALS)*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Matisoff, J. A. (1973). "Tonogenesis in Southeast Asia," in *Consonant Types and Tone Southern California Occasional Papers in Linguistics*, ed. L. M. Hyman (Los Angeles: UCLA), 71–95.

Matisoff, J. A. (2001). "Genetic versus contact relationship: Prosodic diffusability in South-east Asian languages," in *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, eds M. W. Dixon Robert and A. Y. Ailhenvald (Oxford: Oxford University Press), 291–327.

Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.

Maurer, P. (1995). *L'angolar: un créole afro-portugais parlé à São Tomé: notes de grammaire, textes, vocabulaires*. Hamburg: Buske Verlag.

Maurer, P. (2008). "A first step towards the analysis of tone in Santomense," in *Roots of Creole structures: weighing the contribution of substrates and superstrates Creole language library*, ed. S. M. Michaelis (Amsterdam: John Benjamins), 253–261. doi: 10.1075/cll.33.13mau

McElvenny, J. (2021). Language complexity in historical perspective: the enduring tropes of natural growth and abnormal contact. *Front. Commun.* 6:621712. doi: 10.3389/fcomm.2021.621712

McWhorter, J. (2005). *Defining Creole*. Oxford: Oxford University Press.

McWhorter, J. (2007). *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.

McWhorter, J. H. (1998). Identifying the creole prototype: vindicating a typological class. *Language* 74, 788–818. doi: 10.2307/417003

McWhorter, J. H., and Good, J. (2012). *A Grammar of Saramaccan Creole*. Berlin: De Gruyter Mouton.

Morton, J., and Jassem, W. (1965). Acoustic Correlates of Stress. *Lang. Speech* 8, 159–181. doi: 10.1177/002383096500800303

Mous, M. (2009). *The typology of tone in Cushitic. in (Paper presented at the 6th World Congress of African Linguistics (WOCAL 6)*. Cologne: University of Cologne.

Mufwene, S. S. (1996). "Creole genesis: a population genetics perspective," in *Caribbean language issues: old and new (Festschrift for Mervyn C. Alleyne)*, ed. P. Christie (Kingston: University of the West Indies Press), 163–196.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Mufwene, S. S. (2011). Culture: la langue française doit nourrir son homme. *Altern. Int.* 50, 43–43. doi: 10.3917/ai.050.0043

Nakao, S. (2013). "The prosody of Juba Arabic: Split prosody, morphophonology and slang," in *African Arabic: Approaches to dialectology*, ed. M. Lafkioui (Berlin: De Gruyter Mouton), 95–120.

Newmeyer, F. J. (2021). Complexity and Relative Complexity in Generative Grammar. *Front. Commun.* 6:614352. doi: 10.3389/fcomm.2021.614352

Ng, E.-C. (2011). Chinese meets Malay meets English: origins of Singaporean English word-final high tone. *Int. J. Biling.* 16, 83–100.

Nkengasong, N. (2016). *A grammar of Cameroonian Pidgin*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Nylander, D. K. (1984). The relation between the middle tone and "Empty Category Principle" violations in Krio. *Stud. Afr. Linguist.* 15, 163–175.

Odden, D. (1988). "Predictable tone systems in Bantu," in *Autosegmental studies on pitch accent* Linguistic models, eds H. van der Hulst and N. Smith (Dordrecht: Foris Publ), 225–251.

Paster, M. E. (2006). *Phonological Conditions on Affixation*. Berkeley,CA: University of California.

Pustka, E. (2007). Le français régional émergent en Guadeloupe. *Bull. PFC Phonol. Fr. Contemp*. 2007, 261–271.

Quint, N., and Tavares, K. (2019). The common African lexical core of the Upper Guinea Creoles and its historical significance. *J. Ibero-Roman. Creoles* 2019, 115–161.

Rao, R., and Sessarego, S. (2016). On the intonation of Afro-Bolivian Spanish declaratives: Implications for a theory of Afro-Hispanic creole genesis. *Lingua* 174, 45–64. doi: 10.1016/j.lingua.2015.12.006

Ratliff, M. (2015). "Tonoexodus, Tonogenesis, and Tone Change," in *The Oxford Handbook of Historical Phonology*, eds P. Honeybone and J. Salmons (Oxford: Oxford University Press), 245–261. doi: 10.1093/oxfordhb/9780199232819.013.021

Raymond, H (ed.) (2017). *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press.

Remijsen, B., and Van Heuven, V. J. (2005). Stress, tone and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology* 22, 205–235.

Rivera Castillo, Y., and Faraclas, N. (2006). The emergence of systems of lexical and grammatical tone and stress in Caribbean and West African Creoles. *Sprachtypol. Universalienforschung* 2006, 148–169.

Rivera-Castillo, Y. (1998). Tone and stress in Papiamentu: the contribution of a constraint-based analysis to the problem of creole genesis. *J. Pidgin Creole Lang.* 13, 297–334.

Rivera-Castillo, Y., and Pickering, L. (2004). Phonetic correlates of stress and tone in a mixed system. *J. Pidgin Creole Lang.* 19, 261–284.

Rolle, N. (2018). *Grammatical tone: Typology and theory*. Available online at: https://digitalassets.lib.berkeley.edu/etd/ucb/text/Rolle_berkeley_0028E_18224.pdf (accessed March 21, 2021).

Rolle, N. (2020). "Many Hands or Many Cooks? The Functional Load of Tone in Tonal Languages," in *Invited talk at the Zentrum für allgemeine Sprachwissenschaft (ZAS), Berlin, Germany, 8 October*. Available online at: https://nicholasrolle.com/s/Rolle-ZAS-BBL-2020-v3.pdf (accessed March 28, 2021).

Rountree, S. C. (1972). Saramaccan tone in relation to intonation and grammar. *Lingua* 29, 308–325.

Salmons, J. (1992). *Accentual change and language contact: Comparative survey and a case study of early Northern Europe*. Stanford: Stanford University Press.

Sande, H. (2018). Cross-word morphologically conditioned scalar tone shift in Guébie. *Morphology* 28, 253–295. doi: 10.1007/s11525-018-9327-1

Santos, V. G. (2019). *Aspectos prosódicos do português angolano do Libolo: entoação e fraseamento*. Ph. D. thesis. São Paulo: Universidade de São Paulo, doi: 10.11606/T.8.2020.tde-03032020-174301

Sebba, M. (1997). *Contact languages: pidgins and creoles*. Houndmills: MacMillan Press.

Sessarego, S. (2020). Not all grammatical features are robustly transmitted during the emergence of creoles. *Humanit. Soc. Sci. Commun.* 7, 1–8. doi: 10.1057/s41599-020-00611-x

Siegel, J. (2004). Morphological simplicity in Pidgins and Creoles. *J. Pidgin Creole Lang.* 19, 139–162. doi: 10.1075/jpcl.19.1.06sie

Smith, N. (1987). *The genesis of the Creole languages of Surinam*. Ph. D. thesis. Amsterdam: Univer sity of Amsterdam.

Smith, N. (2017). "Krio as the Western Maroon creole language of Jamaica, and the /na/ isogloss," in *Creole Language Library* 53, eds C. Cutler, Z. Vrzić, and P. Angermeyer (Amsterdam: John Benjamins), 251–274. doi: 10.1075/cll.53.11smi

Smith, N., and Adamson, L. (2006). Tonal phenomena in Sranan. *Sprachtypol. Universalienforschung* 59, 211–218.

Sutcliffe, D. (2003). "Eastern Caribbean suprasegmental systems: A comparative view, with particular reference to Barbadian, Trinidadian, and Guyanese," in *Contact Englishes of the Eastern Caribbean* Varieties of English Around the World, eds M. Aceto and J. P. Williams (Amsterdam: John Benjamins), 265–296. doi: 10.1075/veaw.g30.15sut

Swolkien, D. (2015). *The Cape Verdean Creole of São Vicente: Its Genesis and Structure*. Coimbra: Universidade de Coimbra.

Sylvain, S. (1936). *Le créole haitien: morphologie et syntaxe*. Wetteren: Imprimerie de Meester.

Thomas, E. R., and Carter, P. M. (2006). Prosodic rhythm and African American English. *Engl. World-Wide* 27, 331–355. doi: 10.1075/eww.27.3.06tho

Thomason, S. G., and Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Los Angeles, LA: University of California Press.

Thompson, A. L. (2018). Are tones in the expressive lexicon iconic? Evidence from three Chinese languages. *PLoS One* 13:e0204270. doi: 10.1371/journal.pone.0204270

Todd, L. (1980). Review of Pidgin and Creole Linguistics, by Albert Valdman. *J. Linguist.* 16, 140–143.

Traill, A., and Ferraz, L. I. (1981). The interpretation of tone in Principense Creole: abstract phonological analysis. *Stud. Afr. Linguist.* 12, 205–215.

Trudgill, P. (2010). "Contact and Sociolinguistic Typology," in *Handbook of Language Contact*, ed. R. Hickey (Malden, MA: Wiley-Blackwell), 299–319.

Uchihara, H. (2010). *Pronominals in Edo: a descriptive study*. Available online at: file:///Users/linguist/Downloads/Final_Paper_Pronominals_in_Edo_A_descrip.pdf (accessed May 10, 2021).

van Coetsem, F. (1988). *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris Publications.

van Coetsem, F. (2000). *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag Winter.

van der Hilst, E. (1988). *Skrifi Sranantongo bun, leysi en bun tu [Write Sranantongo, read it well too]*. Paramaribo: Author.

Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study*. Available online at: https://dspace.library.uvic.ca//handle/1828/1282 (accessed March 29, 2021).

Wedekind, K. (1985). Thoughts when drawing a map of tone languages. *Afr. Arbeitspapiere AAP* 1, 105–124.

Wee, L.-H. (2016). Tone Assignment in Hong Kong English. *Language* 92, e67–e87.

Weinreich, U. (1953). *Languages in Contact*. New York,NY: Linguistic Circle of New York.

Wells, J. C. (1982). *Accents of English 3: Beyond the British Isles*. Cambridge: Cambridge University Press.

Welmers, W. E. (1959). Tonemics, morphotonemics, and tonal morphemes. *Gen. Linguist.* 4, 1–9.

Welmers, W. E. (1973). *African language structures*. Berkeley: University of California Press.

Wolff, H. E. (2018). *Afro-Asiatic languages*. Available online at: https://www.britannica.com/topic/Afro-Asiatic-languages (accessed June 26, 2021)

Yakpo, K. (2012). "Reiteration in Pichi: Forms, functions and areal-typological perspectives," in *The morphosyntax of reiteration in creole and non-creole languages Creole Language Library (CLL)*, eds E. A. Aboh and A. Zribi-Hertz (Amsterdam: John Benjamins), 251–284. doi: 10.1075/cll.43.08yak

Yakpo, K. (2015). "Code-switching and social change: Convergent language mixing in a multilingual society," in *Codeswitching between structural and sociolinguistic perspectives Linguae et Litterae*, eds G. Stell and K. Yakpo (Berlin: De Gruyter), 259–287. doi: 10.1515/9783110346879.259

Yakpo, K. (2017a). Towards a model of language contact and change in the English-lexifier creoles of Africa and the Caribbean. *Engl. World-Wide* 38, 50–76. doi: 10.1075/eww.38.1.04yak

Yakpo, K. (2017b). "Unity in diversity: The homogeneity of the substrate and the grammar of space in the African and Caribbean English-lexifier Creoles," in *Language contact in Africa and the African diaspora in the Americas. In honor of John V. Singler Creole language library*, eds C. Cutler, Z. Vrzic, and P. Angermeyer (Amsterdam: John Benjamins), 225–250. doi: 10.1075/cll.53.10yak

Yakpo, K. (2019a). *A grammar of Pichi*. Berlin: Language Science Press. doi: 10.5281/zenodo.2546450

Yakpo, K. (2019b). Inheritance, contact, convergence: Pronominal allomorphy in the African English-lexifier Creoles. *Engl. World-Wide* 40, 201–225. doi: 10.1075/eww.00028.yak

Yakpo, K. (2020). "Social factors," in *The Routledge Handbook of Language Contact Routledge Handbooks in Linguistics*, eds E. Adamou and Y. Matras (London: Routledge), 129–146. doi: 10.4324/9781351109154-10

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zamora, A. (2010). *Gramática descriptiva del fá d'ambô*. Barcelona: Ceiba Ediciones.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yakpo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

=	Clitic morpheme boundary
1	First person
2	Second person
3	Third person
2T	Two tone system
2T3	2 input vs. 3 output tone heights
H	High tone
L	Low tone
M	Mid tone
HL	High-low contour tone
LH	Low-high contour tone
H-L	Separates tone-bearing units (syllables), e.g., *kátà* /H-L/ 'scatter'
!H	Extra-high tone
↓H	Downdrifted or downstepped high tone
/H/	Input (phonological) tone
[H]	Output (phonetic) tone
/HL/ → [H]	Input /HL/ becomes output [H]
H*	High intonatonal pitch accent
L*	Low intonational pitch accent
ó	High level tone
ò	Low level tone
ô	High-low (falling) contour tone
ǒ	Low-high (rising) contour tone
C	Consonant
COMP	Complementizer
DEF	Definite article
EMP	Emphatic
Eng.	English
F	Feminine gender
FUT	Future tense
Fr.	French
IDEO	Ideophone
INDF	Indefinite article
IPFV	Imperfective aspect
M	Masculine gender (in glosses)
NEG	Negative
OBJ	Object case
OCP	Obligatory Contour Principle
PL	Plural number
Port.	Portuguese
POSS	Possessive case
POT	Potential mood
PROG	Progressive aspect
PRF	Perfect tense-aspect
PST	Past tense
QUOT	Quotative complementizer
REP	Repetition
SBJ	Subject case
SBJV	Subjunctive mood
SG	Singular number
V	Vowel
V́	H-toned vowel
V̀	L-toned vowel

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 15 November 2021
doi: 10.3389/fpsyg.2021.730112

Vocabulary: Common or Basic?

*András Kornai**

SZTAKI Institute for Computer Science, Budapest, Hungary

Neither linguistics nor psychology offers a single, unified notion of simplicity, and therefore the simplest "core" layer of vocabulary is hard to define in theory and hard to pinpoint in practice. In section 1 we briefly survey the main approaches, and distinguish two that are highly relevant to lexicography: we will call these common and basic. In sections 2 and 3 we compare these approaches, and in section 4 we point the reader to Kolmogorov complexity, unfamiliar as it may be to most working psychologists, lexicographers, and educators, as the best formal means to deal with core vocabulary.

Keywords: core vocabulary, basic vocabulary, word meaning, definition, word frequency, computational lexicography

1. BACKGROUND

Researchers and educators have a clear intuitive sense of text simplicity, and there appears to be complete agreement that simplicity is a strong contributing factor in mastering the reading task for low literacy readers, both non-native speakers and normal language learners (Watanabe et al., 2009; Paetzold, 2016); and for people suffering from language disorders such as autism, aphasia, or dyslexia (Parr, 1993; Evans et al., 2014). Unfortunately, neither linguistics nor psychology offers a single, unified notion of simplicity, and therefore the simplest "core" layer of vocabulary is hard to define in theory and hard to pinpoint in practice (Borin, 2012). Standard measures of simplicity, also known as "readability formulas" such as SMOG, F-K, Dale-Chall, etc. (see Zamanian and Heydari, 2012 for a recent survey) tend to concentrate on easily, automatically measurable factors such as the length of words and sentences.

Beyond sentence length, greater emphasis on syntactic complexity caused by the use of coordination, subordination, pronominalization, passive voice, and relative clauses, is a relatively recent area of research (Alva-Manchego et al., 2020). In this paper we will concentrate on the contribution of the vocabulary, taken to include morphological complexity as well, at the expense of syntactic measures. This is justified both by pure information-theoretic considerations (Kornai, 2019) and by functional Magnetic Resonance Imaging (fMRI) studies (Fedorenko et al., 2020).

We begin with a general survey of approaches to simplicity in the physical, biological, computational, psychological, and cognitive sciences, especially as language can be investigated from all these viewpoints. The most general approach to simplicity is to set up a scale with polar opposites 'simple' and 'complex', and some measure of where a particular entity falls on that scale. Since the basic scheme of scalar comparison is common to all these approaches, the difference must be lodged in the measure itself, and our list concentrates on these. The main variants are as follows.

1. Ordinal measures: Perhaps the single most popular measure in psychology and survey research is the Likert scale, typically 5 points which in our case would be "very simple, simple, neither particularly simple not particularly complex, complex, very complex," but more detailed (7, 9, or 11 point) scales are used quite often, and Pearse, 2011 concluded that even 21 points could be helpful to the researcher. All speakers of English know that *chew* is simpler than *masticate*, and it is this knowledge that a Likert-style survey brings to light whatever granularity we impose. The subjective 'intuitive' sense that we spoke of above is real, and this is the method to quantify it. We may try to model this knowledge in terms of other factors (e.g., Anglo-Saxon or Latinate origin of

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, Sevilla University, Spain

Reviewed by:
Matt Absalom, The University of Melbourne, Australia
Francisco J. Salguero-Lamillar, Sevilla University, Spain

***Correspondence:**
András Kornai
kornai@sztaki.hu

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

Received: *24 June 2021*
Accepted: *20 September 2021*
Published: *15 November 2021*

Citation:
Kornai A (2021) Vocabulary: Common or Basic? Front. Psychol. 12:730112. doi: 10.3389/fpsyg.2021.730112

words), but surveys are a fundamental data source, akin to grammaticality judgments, in that they constitute the explicanda for an entire field.

2. Counting: The assumption that entities with fewer parts are simpler than those with more parts can be traced back to the very beginnings of philosophy, the pre-Socratics. Democritus is very clear that atoms are the simplest possible things, and Plato's *Theaetetus* where knowing the wagon is equated to knowing its parts implies an epistemic commitment to a counting approach. Counting, length in particular, plays a huge role in readability formulas.

3. Developmental measures: Entities that appear earlier in the course of development, be it ontogenic or phylogenic, are considered simpler than those that appear later. In language development, this idea has to be taken with a grain of salt, as there are highly complex entities that appear very early on just because they are very frequent/salient. Clearly, the form *mama* correlates with a far more complex and dynamic collection of sense data, spanning more modalities, than the form *light*, whose meaning will therefore be simpler on any account, yet no infant learns *light* before *mama*.

4. Algorithmic complexity, description length: When component parts are not simply listed, but come with observable token frequencies/probability weights, we treat the uniform (equiprobable) distribution as the most simple, and treat additional parameters as additional complexity. This naturally leads to Kolmogorov complexity (Li and Vitnyi, 1997) and the strongly related Minimum Description Length (MDL) paradigm (Rissanen, 1978; Vitanyi and Li, 2000), which favors the shortest model over equally descriptive longer ones. Anticipating our conclusions somewhat, the analysis provided in sections 2 and 3 will furnish the empirical basis for our overall conclusion presented in section 4, that it is only this theory that can provide the right conceptual underpinnings for dealing with core vocabulary.

Historically, work on readability is closely tied to (3), and started with finding the most frequent words (Thorndike, 1921, 1931; Thorndike and Lorge, 1944) with an express pedagogical purpose, both for L1 and L2 learning. While the key assumption behind this work, that learning one word is about as hard as learning another, has stood the test of time, learnability has mushroomed into a large field of research, and even a brief overview is beyond the scope of this paper—see Klare (1974) and Paasche-Orlow et al. (2003) for informed but somewhat dated summaries, and for the more contemporary approach of bringing machine learning techniques to the task, see e.g., Pilán et al., 2014; Morato et al., 2021.

Here we take the central idea to mean simply that effort is best spent on the words that will cover the overall distribution best, i.e., on the most common ones. Remarkably, this means that serious effort needs to be spent on *function words*, because these are disproportionately present in the high frequency range. For example, in the speech portion of the British National Corpus (BNC, V2.0) that we will use in section 2 there are 9.6 m tokens for 61 k word types, and among the top 100, which jointly account for 5.6 m (58.6%) of the tokens, we find only 22 content words, half of which are prepositions. In many other languages, the same effort that in English we dedicate to function words would have to be spent on case endings and other morphological markers.

In more analytic languages like English the task of learning the grammar is intimately bound up with the task of learning the vocabulary, while in more synthetic languages like Latin the two tasks are somewhat easier to separate. Since about 85% of the information content is carried by the words (see Kornai, 2019 Ch.1.3 for discussion), comprehending them will be central to learning any language irrespective of typological differences, a fact already established in the classic (Thorndike, 1917). Importantly, recent fMRI work has established that the world's languages are functionally localized to the same brain network (Ayyash et al., 2021) so restricting this study to English does not significantly diminish the generality of our conclusions.

To fix terminology, we will refer to the frequency-based approach as aiming at *common* vocabulary, and the comprehension-based approach as aiming at *basic* vocabulary, without pre-judging the issue of how this relates to (1–4) above. At first blush, the goal of pocket dictionaries (typically 500–5,000 words) and collegiate dictionaries (typically 20–50 k) is very similar: they select a subset of the vocabulary that will provide maximum coverage in the statistical sense. This is a modern-day version of (2): we keep the word count fixed, and maximize the probability mass that can be covered by so many words[1]. It is only at the unabridged sizes that another goal, explaining what a word means, becomes evident: we look up *anaphylactic* in the dictionary precisely because it is so rare that we haven't seen it before but want to know what it means. For such an explanation to work, it is necessary to use words on the right hand side (rhs) of the definition that are, in sense (1) at least, simpler than the definiendum. We may define *masticate* via *chew*, but not the other way round, even though the two words are synonymous.

The largest contemporary effort focusing on explaining everything in simple terms is the Simple English Wikipedia https://simple.wikipedia.org, based on the principles of Ogden (1930), though not entirely faithfully (Yasseri et al., 2012). Ogden selected 850 basic words: 600 nouns, 150 adjectives, and a 100 verbs "which put the others into operation and make them do their work in statements." His method of selection was reductive, eliminating words such as *puppy* as long as *young* and *dog* were available. In this example, both words on the rhs are simpler (1). This, as we shall see here, is not fully sufficient: we also need to guarantee that the method of combining the elements that appear on the rhs is also simple.

In this particular case the method of combining *young* with *dog* is conjunction, obviously an elementary step, but let us inspect how Webster's 3rd (Gove, 1961) deals with *anaphylactic*: 'of, related to, affected by, or accompanying anaphylaxis'. Certainly at this stage the reader has gained very little comprehension. The true import of this definition, that *anaphylactic* is the adjectival form of *anaphylaxis*, is accessible only to the linguistically sophisticated reader—all that ordinary

[1]For historically accuracy we should mention that Thorndike's *Word Book*s actually attempted to balance considerations of text frequency (TF) and document frequency (DF), really aiming at maximum coverage over a variety of genres, anticipating (Spärck Jones, 1972) to a remarkable extent.

users see is that they must look up this other word. When they do, they find

> hypersensitivity (as to foreign proteins or drugs) that is marked by a tendency to intense systemic reaction and that results from specific sensitization following one or more usu. parenteral contacts with sensitizing agent and seen chiefly in experimental animals but manifested in man in acute serum sickness and in severe or fatal reactions to later administrations of certain drugs (as penicillin).

This is hardly reassuring. Even if we ignore the difficult words and phrases (senzitization, parenteral contact, sensitizing agent, serum sickness, …), for which the reader will have to consult the dictionary again and again, substituting this rhs is the earlier definition to obtain 'of, related to, affected by, or accompanying hypersensitivity (as to foreign proteins …' is nontrivial. Is it *of*? Is it *related to*? Is it *accompanying*? Is it *affected by*? All the above?

To genuinely aid comprehension, the dictionary must make the rhs simpler (1) than the definiendum, and must also guarantee that by iterated lookup this property is preserved. A major step in this direction is to restrict the rhs to a basic stratum, and to avoid the need for iterated lookup by strictly enforcing this restriction, as was done in the Longman Dictionary of Contemporary English (LDOCE, Procter, 1978). But even LDOCE permits a single indirection step (e.g., *deprecatory* is defined with the aid of *apologetic*) and gives no guidelines as to the necessary syntactic changes that must accompany such a substitution. For example, *Saturn* is defined as "the PLANET which is 6th in order from the sun and is surrounded by large rings" and at *planet* we find "a large body in space that moves round a star, esp. round the sun." But if we mechanically substitute this rhs back in the definition of Saturn, we obtain "the a large body…" rather than "the large body…." Humans of course eliminate one of the colliding articles "the a" as a matter of course, but for a computational system the changes such substitutions trigger are not at all trivial.

Besides the Ogden list, and many other concept lists that share the explanatory goal of basic vocabularies (see List et al., 2016 for a modern system that unifies many), there is another important source for the basic approach, sometimes with overt claims for simplicity (3), but more often with the goal of uncovering genetic and areal relationships among languages. Perhaps the best known are the Swadesh (1950) and Swadesh (1955) lists, widely used in glottochronological studies to this day. Instead of "foundationality" in the sense that in principle every other word should be explained based on them, the Swadesh lists aim at "accessibility" in the sense that words corresponding to the concepts in question should not be too hard to identify in any language. There is no life without water, so words for *river, lake,* or *swim* are likely present. A quarter of the Swadesh list is devoted to natural objects, natural phenomena, and body parts, but if our goal is to define other words it is entirely inadequate. Consider the word *random*. Using the Longman defining vocabulary, we have "happening or chosen without any definite plan, aim, or pattern." None of the rhs words appear on the Swadesh list, and it is not even clear how we could build definitions of them.

2. HOW COMMON IS BASIC?

Here we compare vocabulary lists based on these two approaches both to see what they have in common and to uncover the salient differences. Since spoken language precedes written both ontogenically and phylogenically, we will use only the spoken segment of the BNC. While contemporary English discourse often revolves around culture-specific issues that have no direct counterpart in other languages and cultures, this is still a better proxy for approximating less resourced languages and pre-literate usage than other major corpora based on written materials.

To avoid the issue of function words, we remove the most frequent 100 of these. In speech, this list includes 17 terms that are either filled pauses *mhm erm mm ah Er Mm er*; clearly phatic *actually alright bloody look okay quite really yeah*; or both *oh ooh*. The possessive suffix *'s* is tokenized separately by the Stanza NLP package[2] we used in the analysis, leaving us with 82 ordinary function words (see **Appendix A**). We also remove from the frequency count the six most common punctuation marks.,?!-; because these, in keeping with the convention that is standard in computational linguistics, are treated by Stanza as separate tokens. In total, frequent function words and punctuation are responsible for 60.6% of the tokens, with filler and phatic elements constituting 3.6%, and punctuation 11.9%. In what follows, all percentages refer to the remaining 39.6% (4.5 m tokens) of content words as 100%, though more rare function, phatic, and punctuation tokens are still present in small numbers.

We will consider five basic lists. Of these, the most ambitious is the natural semantic metalanguage (NSM) list (Goddard and Wierzbicka, 2014), in that it contains very few words, yet aims at being fully foundational, in principle offering a basis for defining every word sense in every language by combinations of a few dozen semantic primitives. In section 4 we will look more closely at the definition of *soul* offered in a cross-cultural case study (Wierzbicka, 1989). The Swadesh list already uses word combinations to distinguish word senses, e.g., *right* 'correct' versus *right* 'side', but what is a rather arbitrary disambiguation device for Swadesh, becomes a central organizing principle of NSM, which employs a variety of sophisticated syntactic constructions to define new phrases using the word list.

Next comes the (Swadesh, 1955) list, which would have good resources for function words: 22 of out of our 100 function words are listed by Swadesh, comprising 10.6% of his list. To obtain comparable numbers across basic vocabularies, we remove these here, even those two, *right* and *say*, which were clearly intended by Swadesh in the contentful, rather than the phatic sense. This is not to say that phatic skills are irrelevant for (early) language development, but corpus linguistic resources to study the issue are sadly lacking, especially as transcriptors have a strong tendency to normalize much of this out of the written corpora—studies such as (Bazzanella, 1990) are few and far between.

The `4lang` defining vocabulary (Kornai, 2022) is a medium-size vocabulary (732 words, see **Appendix B**) aiming both at

[2]https://stanfordnlp.github.io/stanza

foundationality and at controlled syntax: definitions are written in a language that has its own formal grammar (and yacc parser) that regulates the manner in which elements can combine. It was obtained from the earlier (Kornai, 2019) `4lang` list by systematic removal of word senses definable in terms of the remaining elements (Ács et al., 2019).

Another medium-size list is Basic English. After removal of stopwords, there remain 799 elements. Ogden (1944) was very cognizant of the differences between ordinary language use and the use of specialist vocabulary: by design, Basic English requires an additional 100 words of General Science, and 50 from each discipline he considered (physics/chemistry, geology, mathematics/mechanics, biology, business, economics). Limitations of the basic vocabulary in expressing the meaning of specialist words will be discussed in section 3.

Our last example of a basic system of words is the Longman Defining Vocabulary (LDV), 2,112 items once the function words are removed. This is the only list that is actually proven to have the power to act as foundation: LDOCE defines over 82k word senses, and there is little doubt that in a larger dictionary the authors could go further. Actually, the core LDV also contains a fair amount of (not always productive) English morphology: the prefixes *counter- dis- en- fore- im- in- ir- mid- mis- non- re- self- un- vice- well-*; and the suffixes *-able -al -an -ance -ar -ate -ation -dom -ed -ee -en -ence -er -ery -ess -est -ful -hood -ible -ic -ical -ing -ion -ish -ist -ity -ive -ization -ize -less -like -ly -ment -ness -or -ous -ry -ship -th -ure -ward -wards -work -y*. Stanza detects inflection (-s, -ind, -ed, -en) even in irregular cases like *go/went*, which makes the coverage statistics presented in **Table 1** more realistic.

In addition to the five original lists, we considered their union ($\cup$), and those that appeared in at least 3 of the 5 ($\cap_3$). These are not intended as a lexicographic proposal to somehow synthesize a better list: obviously the union is redundant as a basic list, and the foundationality of the majority intersection is not guaranteed. That said, they will be useful in drawing out some conclusions. The UG5 (Up-Goer Five, an XKCD comic by Randall Munroe) list, used as basic but derived as common[3] is deferred to section 3.

First, the larger a list the better the coverage: at 2,112 content words the LDV already takes care of about 2/3 of content tokens in the spoken part of the BNC. Since the basic lists were not designed by Thorndike's methodology, this cannot simply be attributed to 'skimming off the top' of the Zipf distribution, but the tendency is clear for growing lists sizes. The last column of **Table 1** shows the 'density' of a list, which shows how much of the weight that could maximally be captured by the top *n* elements is actually captured. Compared to the coverage offered by the most frequent 53 or 185 elements, the actual NSM and Swadesh lists cover only about 30–40% of the best attainable probability mass. For the medium-size `4lang` and Ogden lists, density is higher: these capture about 45-48% of what a common list of the same size would have captured. Finally, a relatively large list like the LDV or the union of the five lists is almost as good as a frequency list, capturing 79% of the theoretical maximum. This number is all the more remarkable given that the UG5 list, which was obtained on a different corpus of English by simply taking the top 1,000 (of which we ignore the function and phatic elements) gets only 86.5% on the BNC spoken materials.

TABLE 1 | Coverage of basic vocabularies.

List	Size	W/o fw	Weight (%)	Avg wt (%)	Density (%)
NSM	78	53	13.3	0.251	41.0
Swadesh	207	185	15.7	0.085	30.9
4lang	732	714	31.2	0.044	45.9
Ogden	850	799	33.4	0.042	48.1
LDV	2,190	2,112	64.4	0.030	78.7
$\cup$	2,390	2310	68.5	0.030	82.7
$\cap_3$	464	428	30.4	0.071	50.0
UG5	1,000	913	61.7	0.068	86.5

The first column is the original size, the second gives the size after removal of function words. Weight is the probability mass of content tokens in the BNC spoken section. See text for the last two columns.

Second, the smaller the list the more general the terms. Even the rarest terms in NSM, *below* and *above*, occur several hundred times each. In contrast, LDV contains 1,146 terms that occur less often than any of the NSM terms, including several like *admittance, adverb, gasoline* that occur only once in the spoken BNC, and some like *cowardly* or *nobleman* which do not occur there at all. The next to last column of **Table 1** shows the average contribution of a list word to the probability mass. The more basic a list, the larger this average contribution turns out to be, indicating not so much the selection of high frequency words as tighter control in terms of excluding really low-frequency ones.

3. HOW BASIC IS COMMON?

In a broad sense, the results of section 2 vindicate both Thorndike and Ogden. Proponents of Thorndike's approach could say: just get the first 1,500 most frequent words, and you covered all the basic vocabulary, since if you covered the NSM list you are done. Proponents of Ogden's approach could say: that is really wasteful, you are using a 1,500 words to accomplish something you could get done by a few dozen.

The pedagogical concern of Ogden and Thorndike is evident, but neither of them could have anticipated how much the goalposts have moved. Today, our interest is not just with L1 and L2 learners, but also with computers: a clear goal of AI, first set by Turing (1950), is to have intelligent conversations with machines. We aim at far more than the ability to deceive a human (Shieber, 2007), the custom-designed Winograd challenge (Levesque et al., 2012) and the updated WinoGrande challenge (Sakaguchi et al., 2020) exercise many semantic facilities. For readers not familiar with this work, here is a typical paired test question:

> The large ball crashed right through the table because it was made of styrofoam *What was made of styrofoam, the ball or the table?*
> The large ball crashed right through the table because it was made of steel. *What was made of steel, the ball or the table?*

[3]https://splasho.com/blog/2013/01/17/a-bit-more-about-the-up-goer-five-text-editor

In addition to the obvious grammatical prerequisites, the task exercises not just encyclopedic knowledge (steel is hard, styrofoam is fragile), but also a generic conceptual scheme, that normally it is hard things that crush through fragile ones and not the other way round.

To see how well common vocabulary can be used to define specialist words, we will briefly survey the 300 entries offered in the spirit of Randall Munroe's *Up Goer Five*[4] explaining terms like *syntax* using "only the 1,000 words people use the most often." The thousand most frequent words were derived from written sources, the Wiktionary contemporary fiction frequency list,[5] and as such, it is well resourced in function words (covers 82 of our 100), but far from ideal for content words (86.5% density, see the last line of **Table 1**). Since morphology is largely taken care of by the Automatically Generated Inflection Database,[6] in principle the UG5 vocabulary could work well for explaining technical work such as summarizing PhD theses and for defining specialist words. But there are several recurring problems.

First, the use of *idiomatic English*. Consider "…interesting because that gives us a real leg up in finding out how the mind works"—readers unfamiliar with the English idiom *to give a leg up* will not be able to figure out what is being said here.

Second, *using multiple senses*. For example, the original XKCD cartoon uses *space* both in the sense 'the area beyond the Earth where the stars and planets are' and, for a helium pressurization tank described as "more funny voice air (for filling up space)" in the sense 'the amount of an area, room, container etc that is empty or available to be used'.

Third, *associative descriptions*. "funny voice air" works well as an associative hint for helium, at least for those familiar with helium speech. "the kind of air that once burned a big sky bag" also works well for hydrogen, but only for those aware of the Hindenburg disaster.

Fourth, *nonce compounding*. With a bit of luck, everybody can figure out that "train-food" means fuel. But what are "idea-paper, air-light, pretend-box" or "fire rock"?

Fifth, *circumlocution*. We may be able to figure out that "a jumping animal that lives in the water and makes noise" is a frog (even though frogs don't live in water), but what is "the stuff that comes out of the animal with white and black spots"?

Sixth, *lack of naming*. A very large proportion of the specialist vocabulary refers to technical concepts that have a reserved meaning or directly reserve (create) a new meaning for a non-technical term. To learn about liquid oxygen "cold air for burning" we first need to learn about liquefying and fractioning gases: "wet and very cold" air would mean something entirely different in everyday language.

The first two problems are easily remedied by a system that does more than mechanically check the description against a word list. The third one actually leverages the preexistence of the kind of world knowledge that it aims at creating. Actually, nonce compounds and circumlocutions have the same mechanism, when they work, and they fail precisely when the outside knowledge is for some reason hard to access.

Ogden's approach was to leave room for 50 specialist words in each field of science he considered. Unfortunately linguistics, psychology, or cognitive science was not one of them, and for this reason we also omitted the specialist vocabulary of `4lang`, which includes grammatical terms like *agent, patient, instrument, …* and logical terms such as *cause, part-of, …* since these are never used in the BNC in the technical sense.

4. CONCLUSIONS, FURTHER WORK

In the final analysis, we see lack of naming not as a problem but as a solution. For an example from the same corpus[7], consider the following: "Everyone knows how to add numbers together. Right? But sometimes we want to use things that are not numbers and that is hard. We wish we were adding numbers instead. So we came up with a thing called a "group". We wrote down all the things that numbers do when you add them. And we said: if something does all the things that numbers do when you add them, then that thing is a "group" …"

Once we permit definitions, we may really begin to explain things. Everyone knows how to add numbers together. Right? This is called *addition*. But sometimes we want to use things that are not numbers and that is hard. This is called *symbolic computation*. So we came up with a thing called a "group". We wrote down all the things that numbers do when you add them. These are called *group axioms*. And we said: if something does all the things that numbers do when you add them (this is called *satisfying the group axioms*), then that thing is a "group"!

If things can be named, we are able to do away with the puzzle-solving aspect entirely, except for natural kinds (Quine, 1969). The fact remains that one either knows that the "animal of central Asia that looks like a cow with long hair" (LDOCE) is a *yak* or one can accept this as the definition of 'yak', there being no competing central Asian animal that would fit the rest of the definition. Once you have *milk* defined as "a white liquid produced by cows or goats that is drunk by people" (LDOCE), you no longer need to play clever games about the animal with white and black spots. The humorous effect of the original Up Goer Five comic and the subsequent *1,000 words of science* entries lies in great part in the puzzle-solving, but if our goal is actually to convey information, especially to those who don't already have it, *adding recursive definition of new words and phrases is a must*.

On the whole, when we speak of simple language, we generally mean both simple vocabulary and simple grammar. Here we concentrated on vocabulary, offering only a few tentative remarks in regards to grammar. Yet it is clear that to a certain extent these two are fungible: we can tighten the vocabulary at the expense of longer definitions. As an example, let us consider the NSM definition of a cross-culturally salient, albeit non-scientific concept, *soul*. (Line numbering added to the original definition in Wierzbicka, 1989, p 43.)

```
1. one of the two parts of a person
```

[4]https://xkcd.com/1133

[5]https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Contemporary_fiction

[6]http://wordlist.aspell.net

[7]https://tenhundredwordsofscience.tumblr.com/archive

```
2. one cannot see it
3. it is part of another world
4. good beings are part of that world
5. things are not part of that world
6. because of this part, a person can be
a good person
```

Notice the syntactic complexities in this definition. By using *the*, (1) already presupposes a theory: a person has two parts. The soul is one of these two. To simplify matters, we will call the other, which remains unnamed throughout the definition, the body, especially as this word is already part of the very limited NSM vocabulary. (2) is a simply conjoined statement that the soul is invisible, not any different from any ordinary definitional clause, e.g., that glass is transparent, or elephants are large. (3), however, introduces a new entity, `another world`, which again comes with an existential presupposition of there being one (ordinary) world relative to which this world counts as "another." (4) and (5) serve to define the other world, and we note that it takes a great deal of syntactic sophistication to recover the `that world` of these clauses as the *other world* of clause (3), while `this part` in clause (6) is resolved as the definiendum *soul*.

Also implicit in the definition is some general compilation of things, a *world*. (This is problematic only because we don't have an NSM dictionary of English.) The point to be noted is that we see the same generic conceptual scheme ONE–OTHER invoked twice: once for parts of a person, and the second time for worlds. We have argued elsewhere that this conceptual scheme is tied to the meaning of *other* (see Kornai, 2022 Figure 1.3) but whatever solution one might propose, it takes significant discourse representational resources (Kamp, 1981; Heim, 1982) to keep these two instances separate.

Syntactic complexities aside, this is remarkably close to the LDOCE definition of *soul* "the part of a person that is not physical, and that contains their character, thoughts, and feelings. Many people believe that a person's soul continues to exist after they have died," which also accounts for the doctrine, seen in many religions, of the immortality of souls (but does not make this an essential feature of the definition). The underlying theories are also similar in asserting the non-physical nature of the soul, and in positing it as the locus of goodness (character). In fact, LDOCE offers a different sense 'the special quality or part that gives something its true character' as in *Seafood is the soul of Provencal cousine*.

To summarize, a notion of *core* vocabulary that is useful for psychologists, linguists, and educators alike must synthesize the definitional simplicity (basic) and the high occurrence (frequent) aspects. Of the approaches we surveyed in section 1 it is only (4), Kolmogorov complexity, that is capable of doing this. To guarantee fungibility, we will say that the complexity of a defined term such as *group* will be equated to the complexity of its definition "a thing that satisfies the group axioms." This way, introducing and using defined terms incurs no extra penalty. To make shorter definitions simpler, we use a counting measure (2) that counts all the primitives at the same unit value. We also add a coordination penalty c to various clauses, roughly speaking by counting the commas in the definition.

There remain several important questions for further work. Do we wish to count conceptual schemas, such that *other* presupposes *one*, or that hard things crush fragile things, as part of some lexical entries, or do we amortize these over many instances where they are used? How do we count the complexity of function words and bound morphemes, entirely ignored in this study? The answers are of necessity tied to the model of syntax and morphology chosen, and unless we make strides in universal syntax and morphology, we may have to rely on language-specific stopgap measures.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is made available by the authors at https://kornai.com/VCB.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by MILAB, the Hungarian artificial intelligence national laboratory.

ACKNOWLEDGMENTS

I am indebted to the referees for suggestions that led to considerable improvements in the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.730112/full#supplementary-material

REFERENCES

Ács, J., Nemeskey, D. M., and Recski, G. (2019) "Building word embeddings from dictionary definitions," in *K + K = 120: Papers Dedicated to László Kálmán and András Kornai on the Occasion of Their 60th Birthdays*, eds K. M. Beáta Gyuris and G. Recski (Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences (RIL HAS)).

Alva-Manchego, F., Scarton, C., and Specia, L. (2020). Data-driven sentence simplification: survey and benchmark. *Comput. Linguist.* 46, 135–187. doi: 10.1162/coli_a_00370

Ayyash, D., Malik-Moraleda, S., Galle, J., Affourtit, J., Hoffman, M., Mineroff, Z., et al. (2021). *The Universal Language Network: A Cross-Linguistic Investigation Spanning 45 Languages and 11 Language Families*. Cold Spring Harbor Laboratory. Available online at: https://www.biorxiv.org/content/early/2021/07/29/2021.07.28.454040

Bazzanella, C. (1990). Phatic connectives as interactional cues in contemporary spoken italian. *J. Pragmat.* 14, 629–647. doi: 10.1016/0378-2166(90)90034-B

Borin, L. (2012). "Core vocabulary: A useful but mystical concept in some kinds of linguistics". In: *Shall we play the Festschrift game? Essays on the occasion of Lauri Carlson's 60th birthday*. eds. S. Diana, L. Krister, and N. Wanjiku (Cham: Springer), 53–65.

Evans, R., Orasan, C., and Dornescu, I. (2014). "An evaluation of syntactic simplification rules for people with autism," in *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (Gothenburg), 131–140.

Fedorenko, E., Blank, I. A., Siegelman, M., and Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition* 203:104348. doi: 10.1016/j.cognition.2020.104348

Goddard, C., and Wierzbicka, A. (2014). *Words and Meanings: Lexical Semantics across Domains, Languages and Cultures*. Oxford: Oxford University Press.

Gove, P. B. (ed.). (1961). *Webster's Third New International Dictionary of the English Language, Unabridged*. Springfield, MA: G. & C. Merriam.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases* (Ph.D. thesis). University of Massachusetts, Amherst, MA.

Kamp, H. (1981). "A theory of truth and semantic representation," in *Formal Methods in the Study of Language*, eds J. Groenendijk, T. Jansen, and M. Stokhof (Amsterdam: Mathematisch Centrum), 277–322.

Klare, G. R. (1974). Assessing readability. *Read. Res. Q.* 10, 62–102.

Kornai, A. (2019). *Semantics*. Cham: Springer Verlag.

Kornai, A. (2022). *Vector Semantics*. Cham: Springer Verlag.

Levesque, H., Davis, E., and Morgenstein, L. (2012). "The winograd schema challenge," in *Proceeding of the 13th International Conference on Principles of Knowledge Representation and Reasoning* (Rome), 8–15.

Li, M., and Vitnyi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications*. Cham: Springer.

List, J.-M., Cysouw, M., and Forkel, R. (2016). "Concepticon: a resource for the linking of concept lists," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (Portoroz: European Language Resources Association (ELRA)), 2393–2400.

Morato, J., Iglesias, A., Campillo, A., and Sanchez-Cuadrado, S. (2021). Automated readability assessment for spanish e-government information. *J. Inf. Syst. Eng. Manag.* 6:em0137. doi: 10.29333/jisem/9620

Ogden, C. (1930). *Basic English: A General Introduction With Rules and Grammar*. Trench: Kegan Paul.

Ogden, C. (1944). *Basic English: A General Introduction With Rules and Grammar*. Psyche miniatures: General Series. Trench: Kegan Paul.

Paasche-Orlow, M. K., Taylor, H. A., and Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readabilit y. *New Engl. J. Med.* 348, 721–726. doi: 10.1056/NEJMsa021212

Paetzold, G. H. (2016). *Lexical simplification for non-native English speakers* (Ph.D. Thesis). University of Sheffield.

Parr, S. P. (1993). *Aphasia and literacy: the application of practices associated with literacy teaching to the assessment of reading and writing disorders in adult aphasia* (Ph.D. Thesis). University of Central England, Birmingham.

Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: the case of a 21-point scale. *Electr. J. Bus. Res. Methods* 9, 159–171. Available online at: https://academic-publishing.org/index.php/ejbrm/article/view/1278/1241

Pilán, I., Volodina, E., and Johansson, R. (2014). "Rule-based and machine learning approaches for second language sentence-level readability," in *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications* (Baltimore, MD: Association for Computational Linguistics), 174–184.

Procter, P. (1978). *Longman Dictionary of Contemporary English, 1st Edn*. Burnt Mill: Longman.

Quine, W. (1969). "Natural kinds," in *In Ontological Relativity and Other Essays*. (New York, NY: Columbia University Press).

Rissanen, J. (1978). Modeling by the shortest data description. *Automatica* 14, 465–471.

Sakaguchi, K., Bras, R. L., Bhagavatula, C., and Choi, Y. (2020). "Winogrande: an adversarial winograd schema challenge at scale," in *Proceedings of 34th, AAAI Conference on Artificial Intelligence* (New York, NY), 8731–8738.

Shieber, S. M. (2007) The Turing test as interactive proof. 41(4):686–713.

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *J. Document.* 28, 11–21.

Swadesh, M. (1950). Salish internal relationships. *Int. J. Am. Linguist.* 16, 157–167.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *Int. J. Am. Linguist.* 21, 121–137.

Thorndike, E. L. (1917). Reading as reasoning: a study of mistakes in paragraph reading. *J. Educ. Psychol.* 8, 323–332.

Thorndike, E. L. (1921) *The teacher's word book*. New York Teachers College, Columbia University.

Thorndike, E. L. (1931). *A Teacher's Word Book*. New York, NY: New York Teachers College, Columbia University.

Thorndike, E. L. and Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York, NY: Teachers College Bureau of Publications.

Turing, A. (1950). Computing machinery and intelligence. *Mind* 59, 433–460.

Vitanyi, P. M. B., and Li, M. (2000). Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Trans. Inf. Theory* 46, 446–464. doi: 10.1109/18.825807

Watanabe, W. M., amd Vinicius Rodriguez Uzeda, A. C. J., de Mattos Fortes, R. P., Pardo, T. A. S., and Aluiso, S. M. (2009). "Facilita: reading assistance for low-literacy readers," in *Proceedings of the 27th ACM SIGDOC International Conference* (Bloomington, IN), 29–36.

Wierzbicka, A. (1989). Soul and mind: linguistic evidence for ethnopsychology and cultural history. *Am. Anthropol.* 91, 41–58.

Yasseri, T., Kornai, A., and Kertsz, J. (2012). A practical approach to language complexity: a Wikipedia case study. *PLoS ONE* 7:e48386. doi: 10.1371/journal.pone.0048386

Zamanian, M., and Heydari, P. (2012). Readability of texts: state of the art. *Theory Pract. Lang. Stud.* 2, 43–53. doi: 10.4304/tpls.2.1.43-53

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kornai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpsyg.2021.702652

Comparison of Preschooler Verbal and Graphic Symbol Production Across Different Syntactic Structures

Gat Savaldi-Harussi and Leah Fostick*

Department of Communication Disorders, Ariel University, Ariel, Israel

The present study focuses on the impact of graphic symbols used in Augmentative and Alternative Communication (AAC) on clause construction. It is not yet well-understood to what extent communication produced via graphic symbols differs from verbal production. This study attempts shed light on the impact of the graphic symbol modality on message construction beyond individual differences, language knowledge, and language-specific patterns by providing a direct comparison between children's verbal and graphic symbol production. Nineteen typically developing Hebrew-speaking children aged 4–5 years were presented with 16 short videos of actions and were asked to express what they saw verbally and by choosing among graphic symbols displayed on an iPad communication board. The 570 clauses produced by the children were coded and analyzed. A significant difference was found in favor of verbal speech across different syntactic structures in terms of utilization of the target lexicon, syntactic complexity, and expected target word order. These results are consistent with the existing literature for English. Implications for AAC practices are discussed, highlighting the notion that using graphic symbols to represent spoken language may not reflect actual linguistic knowledge and that adequate, explicit instruction is necessary for graphic representation of more complex linguistic structures.

Keywords: expressive use of graphic symbols, clause construction, augmentative and alternative communication, graphic symbol modality, native speakers, transitive and non-transitive verbs, language representation

OPEN ACCESS

Edited by:
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

Reviewed by:
Ruth Berman,
Tel Aviv University, Israel
Elena Theodorou,
Cyprus University of Technology, Cyprus

****Correspondence:***
Gat Savaldi-Harussi
gatsh@ariel.ac.il

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

Received: *29 April 2021*
Accepted: *01 November 2021*
Published: *25 November 2021*

Citation:
Savaldi-Harussi G and Fostick L (2021) Comparison of Preschooler Verbal and Graphic Symbol Production Across Different Syntactic Structures. Front. Psychol. 12:702652. doi: 10.3389/fpsyg.2021.702652

INTRODUCTION

A diverse population uses Augmentative and Alternative Communication (AAC) services, including children and adults with developmental and acquired disabilities whose ability to use natural speech is affected by severe speech or language difficulties (Smith, 1996, 2006; Binger and Light, 2008). AAC is an area of clinical practice that provides tools and techniques to supplement or replace speech, including the use of unaided communication (e.g., gestures, facial expression) and "aided communication" such as graphic symbols displayed on communication devices to represent spoken language (American Speech-Language-Hearing Association, 2020). AAC is often utilized within the context of multimodal communication, which involves selecting the preferred mode of communication (e.g., aided or unaided) that allows the most efficient self-expression. Among those who use AAC and are not yet literate, graphic symbols are the primary communication modality (Von Tetzchner and Grove, 2003).

In the typical, natural course of communication development, children in many cultures are thought to extend their spoken communication by developing external, visually based symbols to

communicate information (e.g., numbers, alphabet letters, and pictorial signs). Psycholinguistics research has found that external symbols accelerate the communication of knowledge and play a crucial role in enhancing human intelligence (Lee and Karmiloff-Smith, 1996; Mavrou et al., 2013). The automatic tendency to use multimodal communication, such as a combination of natural speech along with manual signs and external symbols, has become a principal practice in AAC (Loncke, 2014).

Symbolization, a fundamental component of AAC, is a vast area of research whose definition is pertinent to this study. Symbolization includes two elements: a signifier and a signified. The signifier is defined as something that stands for the signified: be it an idea, person, or object. In cognitive psychology, various terms such as "symbol," "sign," "icon," and "notation" are used to refer to signifiers (Lee and Karmiloff-Smith, 1996). Peirce's theory of signs (1965–1966) suggests that while a sign is the smallest unit of meaning, symbolic signs are arbitrary (e.g., words), and thus the relation between signifier and signified is based on convention. In contrast, iconic signs (e.g., pictures) refer to signifiers that resemble the signified, and indexical signs are those in which the relation between signifier and signifies is based on cause and effect (e.g., smoke and fire) (Atkin, 2014).

In the AAC field, the term "graphic symbol" or "symbol" is used for pictures and graphic representations that are signifiers of ideas the person wishes to convey (Loncke, 2014; Pampoulou and Fuller, 2020). "Graphic symbols" can be part of a symbol system with rules about building the pictures (e.g., Bliss words, Blissymbols) or a symbol set without internal principles for symbol formation (Fuller et al., 1992; Loncke, 2014; Pampoulou and Fuller, 2020). Graphic symbols often take the form of line drawings such as Picture Communication Symbols (PCS), SymbolStix©[1] (Clark, 1997), or Widgit©[2] (Kennedy, 2004) symbols. A set of graphic symbols aiming to represent the spoken language-for the purpose of communication- may include iconic, symbolic, and indexical signs depending on the target referent.

The ability to decode visual forms (such as graphic symbols) depends to a large extent on biological, cognitive, and cultural factors (Lee and Karmiloff-Smith, 1996). Children begin to understand that a symbol (signifier) stands for something else at around the age of two and, by the age of 6–7 years, understand most conventional notation systems (visual forms). That said, there is variation in the pace of development across different visual representations (e.g., drawing vs. written language) (Lee and Karmiloff-Smith, 1996). Indeed, an individual who uses graphic symbols to communicate must have well-developed internal visual representational skills and understand that a symbol is an object by itself and at the same time refers to something else (Loncke, 2014).

The term *iconicity* refers to the representation value of the graphic symbol's *image*, ranging from transparent (i.e., the graphic symbol displays the word's exact meaning, such as a picture of a house to represent the word "house") to translucent (i.e., an indirect relationship between the symbol and word, such as a horizontal line and a dot on top to represent the word "on") (Loncke, 2014). Graphic symbols in AAC are designed to be as transparent as possible, visually representing the target referent.

Considering the characteristics and constraints of visual forms such as the graphic symbols used in AAC, one of the big questions in the AAC field is to what extent production via graphic symbols differs from verbal production. Therefore, this study provides a direct comparison between children's verbal and graphic symbol production after they watch short, silent videos depicting a boy engaging in different actions.

Picture Communication Symbols graphic symbols (signifiers) were selected for this study because of their easy learnability. However, it is important to note that this set of graphic symbols include symbols representing both concrete and abstract referents, thus ranging from transparent to translucent. For example, concrete referents (e.g., a symbol of a boy representing a boy) are considered to have high transparency and are more iconic, as often one can "look through" the symbol (signifier) and easily extract its meaning (Loncke, 2014). In contrast, a graphic symbol is considered translucent when the relationship to the meaning becomes clear only after revealing or learning its meaning. Consequently, representing certain linguistic features (e.g., prepositions, connectors) via graphic symbols may be challenging and not reflect one's actual linguistic mental representation.

Characteristics of the Graphic Symbol Modality

Smith (2006) describes the characteristics of the graphic symbol modality in comparison to natural spoken language. One noted difference, as described above, is the connection between the symbol and its referent; while a spoken word consists of arbitrary sounds that represent a specific referent and therefore the connection between the spoken word and its referent is arbitrary, a graphic symbol is designed to be iconic.

Another difference between graphic symbols and spoken language is related to segmental features. Words are composed of a limited set of meaningless segments (phonemes) that can represent infinite meaningful morphemes and that can be combined to create new meanings, resulting in a simultaneously economical and productive language system. In contrast, graphic symbols (e.g., PCS, SymbolStix) represent a finite set of symbols that cannot be divided into subcomponents and create new meanings. Therefore, the set of graphic symbols is not characterized by the same productivity as oral language. A final difference is that oral language is produced by the human body via a process of sorting words and linguistic structures from the mental lexicon, while graphic symbols are represented externally and visually by a finite set of symbols designed and organized by someone else.

One significant challenge for those who use graphic symbols is the mismatch between spoken language input and graphic symbol output (Trudeau et al., 2007). Since the goal of graphic symbol communication is the recording of verbal-based messages into graphic symbols, the output is expected to mirror spoken language structural properties. Therefore, the individual using graphic symbols needs to make a connection between

[1]SymbolStix© Available online at: https://www.cricksoft.com/uk/products/symbol-sets

[2]Widgit Available online at: https://www.cricksoft.com/uk/products/symbolsets/widgit-symbols.2010.529619

the two modalities in a task of "translation" in which they must switch from one modality to the other by recruiting metalinguistic abilities (Sutton et al., 2002, 2020; Smith, 2006; Trudeau et al., 2007).

Moreover, representing linguistic structures in graphic symbols is also challenging because grammatical category boundaries of spoken utterances may be unclear when transmitting the message into graphic symbols. For instance, when representing verbs via graphic symbols, information regarding the predicate and its arguments may all appear in a single graphic symbol: the action (verb) and its agent (pronoun) may be displayed simultaneously in a single static graphic symbol. An example of this is the verb SIT which is represented by PCS as a line drawing of a person sitting on a chair, viewed in profile. This graphic symbol includes the agent and the object sat upon, and simultaneously represents the action of sitting, the agent, and the object (Smith, 2015). Therefore, rather than selecting three different graphic symbols that represent the three content words (square brackets in the example below) of its eight morphemes, the *single* symbol, in essence, represents the following full sentence:

The person	is	sitting	on a	chair.
Art[NOUN]	Copula	[VERB + ING]	PrepArt	[NOUN]

Similarly, graphic symbols for verbs THROW and PUSH also depict the agent who performs the action, the action itself, and the object of the verb (Sutton et al., 2002).

In recent decades, studies in the field of AAC have attempted to explore the characteristics and constraints of the graphic symbol modality. One related question is to what extent the patterns observed in the word order, syntactic complexity, and lexicon of individuals who use graphic symbols differ from those of typically developing individuals using spoken language, and whether the observed differences are due to individual differences or due to the modality itself (Trudeau et al., 2007, 2010a,b; Savaldi-Harussi et al., 2019; Sutton et al., 2020).

Graphic Symbol and Language Outcomes

Analyzing linguistic structures produced via graphic symbols requires defining the *unit of analysis* based on the characteristics of aided communication (Müller and Soto, 2002; Kovacs and Hill, 2017). In psycholinguistics, an utterance is defined as behavioral stretches of oral output and a clause is defined as "any unit that contains a unified predicate... (that is) a predicate that expresses a single situation" (Berman and Slobin, 1994). For the purposes of this study, the term *clause construction* refers to construction via graphic symbols. Due to the co-constructed interaction feature of aided conversation, in which the message is co-constructed with the adult's scaffolding; researchers often study utterances produced during conversation turns rather than utterances, *per se*; in such a case, the unit of analysis would be the message constructed via graphic symbols without the adult intervening (Savaldi-Harussi and Soto, 2016).

Studies conducted in English language environments have found that children who use AAC have difficulties in tasks that evaluate morpho-syntactic knowledge. Such tasks require judgment about whether a target sentence sounds correct or not (e.g., "Tomorrow they walked"). Children who use AAC have demonstrated difficulties identifying and marking mandatory inflections, manifested in nouns or verbs used without following grammatical standards of verbal and nominal inflections (e.g., suffixes "-ed," "-s," and "-ing" for verbs, and plural "-s" for nouns) and the nominative case (e.g., possessive "'s"), resulting in short construction (e.g., "girl eat banana" instead of "THE girl IS eatING A banana") (Sutton and Gallagher, 1993; Redmond and Johnston, 2001; Blockberger and Johnston, 2003; Savaldi-Harussi and Soto, 2018, Savaldi-Harussi et al., 2019).

Four main patterns have been identified in the expressive language of individuals who use graphic symbols: (1) dominance of utterances (messages) with a single symbol; (2) perseverance of simple structures; (3) changes in word order; and (4) grammatical errors (Smith, 2015). Although output of a single symbol is widely reported (Sutton et al., 2002; Savaldi-Harussi and Soto, 2018), children who use AAC can generate multi-symbol utterances (messages) in which simple constructions of subject–verb–object (SVO) are the common structure (Sutton et al., 2002; Savaldi-Harussi et al., 2019). However, these simple constructions were not found to follow the typical word order of the common clause structure in English consisting of SVO (e.g., MAN DRIVE CAR). Instead, these children were found to use the following structures instead: Subject-Object-Verb (MAN CAR DRIVE), Verb-Subject-Object (DRIVE MAN CAR), or Object-Verb-Subject (CAR DRIVE MAN). When forming complex construction, graphic symbol users tend to change word order in multiple positions: GIRL TREE HELP NEST CLIMB BOY (instead of "the girls help the boy climb a tree to get a nest"; Soto, 1997). Lastly, constructions via graphic symbols have been reported to include key symbols but lack grammatical markers such as auxiliaries, articles, prepositions, and suffixes-even though grammatical markers are available in the communication devices-resulting in ungrammatical structures (Binger and Light, 2008; Sutton et al., 2010).

Indeed, morpho-syntactic differences between the graphic symbol modality and spoken language have been observed in children (Blockberger and Johnston, 2003), adolescents (Redmond and Johnston, 2001), and adults (Sutton and Gallagher, 1993) with typical and atypical language production. Sutton et al. (2010) observed how 30 preschool children transferred SVO structures to graphic symbols, reporting that at least one core element (subject, verb, or object) was missing in more than 50% of the expressions produced, with verbs accounting for 78% of the omissions. One possible explanation of this finding is a relatively low level of iconicity of target verbs in graphic symbols, making it developmentally difficult for young children to represent them (Von Tetzchner and Grove, 2003).

Another possible explanation for atypical linguistic patterns in messages constructed in English via graphic symbols is a lack of attention to linguistic markers that are perceptually less salient to AAC users. Less attention is paid to aspects of language that have little semantic value due to insufficient learning and practicing of morphological rules among this population (Blockberger and Johnston, 2003). For instance, parts of speech in English (e.g.,

nouns, verbs, and adjectives) often appear as bare stems or as free morphemes, and inflections only play a minor role in the relationship between parts of a sentence, while word order provides the critical information (Dromi et al., 1993). Moreover, adding morphological markers to lexical stems via graphic symbols requires cognitive-linguistic effort, memory, and physical effort (Loncke, 2014); thus, short sentences are an effective strategy to enhance the communication pace. This explanation strengthens the notion of particular challenges when mapping spoken language structures onto graphic symbols and puts the modality as the source of the atypical structure, beyond the communication difficulties of those who use it.

To explore whether the level of attention to grammatical markers impacts the morpho-syntactic differences between graphic symbol expression and spoken language, researchers suggested conducting cross-linguistic studies in languages that include grammatical morphology with greater perceptual salience than English (Blockberger and Johnston, 2003; Smith, 2015). Such cross-linguistic studies can shed light on the impact of the graphic symbol modality on message construction beyond individual differences, language knowledge, and language-specific patterns. This is the purpose of the current study.

Contrasting English and Hebrew

Hebrew is a Semitic language with rich morphology. In contrast to English, in which nouns, verbs, and adjectives are often used as bare stems, and are formed by affixation (e.g., dance + er → dancer), zero-conversion (e.g., work–to work), and compounds (e.g., high-school, daycare) (Clark and Berman, 1987; Dromi et al., 1993; Berman et al., 2009), formation of verbs and certain adjective and nouns in Hebrew occur through integrating a consonantal root (e.g., R-Q-D) into a pattern (e.g., CaCCan) to form the word (RaQDan = dancer). The root conveys the core meaning of a word (R-Q-D represents "dance") and often consists of three consonants. Words are also inflectionally marked for number and gender (in Hebrew, animate and inanimate nouns are also marked for number and gender). Verbs are also inflected for tense and need to agree with their subject noun in number, gender, and person: present tense forms are marked for number and gender, whereas past tense forms are marked for person (first, second, and third) as well as number and gender. Moreover, verbs have a special form for the imperative and infinitive (Berman, 1985; Dromi et al., 1993). The least inflected form of Hebrew verb is the masculine singular in present tense (e.g., "moxer" = sell) and the third-person masculine singular in past tense (e.g., "maxar" = sold) that have no prefixes or suffixes. These forms in Hebrew are treated as basic although they are inflected (Dromi et al., 1993). English and Hebrew also differ in their functional categories. English has a definite as well as an indefinite article (*the* and *a(n)*, respectively) while Hebrew marks only the definite article and has an overt accusative marker *et* before a definite object, which is not marked in English.

Subject–verb–object structure is the canonical form in English in which subject-first forms predominate. This especially occurs in utterances involving two nouns and a verb in which the agent is animate and the patient inanimate (Slobin and Bever, 1982). In Hebrew, the word order of a sentence including a verb resembles the English word order SVO (Glinert, 2017). Slobin and Bever (1982) found that the average age for children to use word order strategy is around 3:6, and children are attuned to these canonical sentences.

While graphic symbol use has been researched in English, little research has been done in Hebrew (Vinder, 2016; Mano-Lerman, 2017). Consequently, the aim of current study is twofold: (1) to compare the constructions produced via graphic symbols to those produced verbally across different syntactic structure: subject verb (SV), SVO, and two coordinated clause SV[and]SV, and (2) to compare these constructions by focusing on differences in lexicon, syntactic complexity, and word order across different syntactic levels. This study focused on typically developed (TD) Hebrew-speaking children aged 4–5 years, as at this age children are expected to be at the late linguistic stage in which they acquire coordination structures (Dromi et al., 1993) and develop good internal visual representation. Early literacy skills, such as letter knowledge and print concept, also emerge at this age (Treiman et al., 2007).

The goal of this paper, therefore, is to answer the following questions:

(1) Are there production differences in semantic-syntactic representation of clause structure (lexicon, syntactic complexity, and word order) in Hebrew when using graphic symbols vs. speech?
(2) Are there production differences in clause structure in Hebrew when using different syntactic structures: SV, SVO, [SV] and [SV] across modalities (verbal vs. graphic symbol)?

MATERIALS AND METHODS

Participants and Setting

Nineteen TD preschoolers between age 3:8 and 5:01 (years:months) participated in the final cohort (9 girls/10 boys; M_{age} = 4:03, SD_{age} = 0.45). The original sample included 20 participants, but one was excluded from the study once it was determined that they were receiving speech and language therapy. To qualify for inclusion, children had to meet the following criteria:

(a) be a native speaker of Hebrew;
(b) attained a Sentence Repetition score within 1.5 standard deviations [using a subtest from *The Goralnik Screening Test for Hebrew* (Goralnik, 1995). This test, also known as "sentence recall and sentence imitation," includes different morpho-syntactic structures and serves as a reliable screening task to identify specific language impairments (Theodorou et al., 2017)]; and
(c) have hearing, visual, neurological, linguistic, and communicative development with the normal range based on parental report. No record of speech and language impairment.

The study was carried out at each child's home in a quiet room. A familiar adult was permitted to join the session and instructed

to observe without participating. Third-year speech-language pathology students administered tests under the supervision of the first author. The study was approved by the IRB University (AU-HEA-GH-20190130-1) and was conducted in accordance with appropriate ethical standards.

Materials

Videos Displayed on Interactive Board Game

Sixteen short videos (M = 5.05 s, SD = 3.2 s) of a young boy performing different actions were displayed on a laptop screen within a fun, interactive path board game using Power Point slides. These videos provided only visual representation with no verbal input. The path board included 15 interactive steps organized on a screen with numbers from 1 to 15. When a participant clicked on a step, a video appeared; after watching the video, the participant was asked to describe verbally, and then via graphic symbols, what had just been seen. The videos were designed to elicit target utterances of various syntactic structures described below. To elicit two coordinated clauses, two videos were displayed next to each other with a plus sign (+) between them. Each slide had a button on the right that navigated back to the home board game. Two different orders of videos were used to control the effect of fatigue on the last sentences elicited; the differently ordered videos were randomly assigned to each participant.

Verbs and Syntactic Structures Probe

The target structures were utterances with one clause or two clauses. For utterances with one clause, the target structures were SV and SVO; for utterances with two clauses, the target structure was SV[and]SV. A total of 15 utterances were targeted which included five utterances for each syntactic structure: SV, SVO, and two coordinated clauses (using two SV clauses and the Hebrew coordinator VE [and]). The targets SV, SVO, and SV[and] SV are depicted in **Table 1**. For the SV structure, five non-transitive verbs were selected: jump (*kofets*), laugh (*tzoxek*), dance (*roked*), sleep (*yashen*), and shower (*mitkaleax*). For the SVO structure, five transitive verbs were selected as follows: open (*poteax*); throw (*zorek*), wear (*lovesh*), hold (*maxzik*), and hug (*mexabeck*). These verbs were selected because they emerge early in children's lexicons. Subject-Verb agreement was singular (SG) and masculine (MS) in grammatical number and gender (*kofets* = jumpSG.MS.Present), which is the basic form in Hebrew, and was targeted by presenting one boy (agent) performing different actions in the videos.

Grid Symbols Display

Twenty-three graphic symbols were displayed on a communication board using the AAC application GRID© on an iPad. Twenty-two colored PCS symbols (Mayer-Johnson, 1981–2011) and one letter symbol were displayed on the board as follows: 16 verbs (jump, dance, sleep, showering, laugh, open, throw, wear, hug, slide, swing, hold, play, cry, walk, and dry), six nouns, and one letter for the word *and*. The Hebrew conjunction word VE (*and*) is a bound morpheme attached as a prefix to the words it connects, for example: "John and Mary" would be "John ve Mary." The parts of speech were organized on the board from right to left, the direction of writing in Hebrew, with the pronoun BOY, the agent of all the action, in the right column and the verbs in the left columns. The background of the symbols follows the Fitzburg Key (Fitzgerald, 1969) color codes for distinguishing different parts of speech: green for verbs and yellow for nouns. The board included a message window that visually presented the constructed symbols and voice output in the form of digitized speech ("Matan"). The user can activate the message window and receive auditory feedback on the constructed message. Each symbol/button also has a voice output that serves as auditory feedback. The children could modify the message by deleting a single symbol or the whole message and indicate when the constructed message was done.

Procedure

Prior to the start of the study, a verbal explanation was provided to the guardian of each of the participants about the purpose and procedures of the research, and they were also given a detailed written explanation via the consent form. Those who wished to take part in the research provided signed informed consent on behalf of their children to participate in the study and filled out a questionnaire about their child's personal, developmental, and demographic details. As a part of the study, a screening test was conducted for each participant using the sentence repetition subtest from the Goralnik assessment tool. Next, children were trained to use the AAC (see section "Familiarization and Training" for further details). During the experimental phase, the children were given the following instruction about the interactive game board: "Watch a video and say aloud what you see. Then, say it with the symbols on the board." All participants' productions, both verbal and graphically symbolic, were documented. All meetings took place in a quiet room and the presence of an adult familiar to the participants was allowed in order to achieve maximum cooperation on the part of the participants. Each session lasted about 45 min and was conducted by university students from the Communication Disorders department. Participants' results were assessed by at least two evaluators. All experimental sessions were carried out by research assistants who were speech-language pathology students under the supervision of the first author.

Familiarization and Training

Before the experiment was conducted, the children were trained to use the AAC board. Research assistants presented the symbols that appear on the AAC board. The familiarization phase included two steps, as follows: first, the participant was asked to name each symbol and then to click on the symbol to receive the auditory feedback. The children's naming of the symbols was documented. Then, the child was asked to construct six structures (two SV, two SVO, and two coordinated clauses) with the graphic symbols that were different from the target sentences. Following each structure, a research assistant modeled how to construct it correctly with the displayed graphic symbols. The graphic symbols used during training were the same as those used for the experimental phase, but the combinations of the verbs and the syntactic structure were different. For example, the verb "walk" was modeled with the syntactic structure SV in "The boy

is walking," by selecting the target symbols "BOY, WALK," but during the trial it appeared within two coordinated clauses ([SV] AND [SV]) of "The boy is walking, and the boy is sleeping."

The naming of the graphic symbols in the training phase was coded to identify symbol transparency. The coding for naming was as follows: three points for a correct response, two points for a semantically close response, and one point for saying an unrelated word or phrase. The transparency level was calculated as the average score of each symbol. Nouns were fully transparent while the verbs varied in their transparency level. We also documented if the participants were familiar with the letter VAV (yes/no). Only 4 (21%) out of the 19 participants identified the letter symbol VAV when it was first introduced.

Scoring and Reliability

All children's production of verbal and graphic symbols (570 clauses) were coded and analyzed in three aspects: lexicon, syntactic complexity, and word order, as described in **Table 2**.

For evaluating the children's lexicon, the maximum score was two points when all target content words were used in the production, one point was given when one content word was omitted or replaced, and zero points given for more than one error.

Syntactic complexity refers to the difference between [SV] and [SVO] (having a complement to the verb) or between [SV] and [SV and SV] (a simple clause or a coordinate clause). For syntactic complexity, based on the metric adapted from Savaldi-Harussi et al. (2019) and modified for the current study, each content word (verb or noun) received one point while the connection word "*and*" (indicating more complex production) gained additional two points. Therefore, a maximum syntax score of 6 was allocated for the two coordinated clauses [SV] AND [SV].

Word order refers to the order of the constituents within the different structures. For example, for the structure SVO, children might answer SV, VSO, or OSV. Only the last two are counted for word order errors, while all three will be counted as errors when comparing the different structures. The maximum score for the word order component was one point when the content words followed the canonical order of Hebrew sentence structure. For example, for the target sentence "The boy (NOUN1) is hugging (VERB) the bear (NOUN2)," a participant may have produced the following responses (a) verbally and (b) in graphic symbols:

(a) *Hugging* (VERB) *bear* (NOUN2)
(b) *BOY* (NOUN1) *BEAR* (NOUN2)

Scoring of the above example would have been done as follows. In (a), for the verbal production consisting of [VERB, NOUN2], only one point would have been given for the lexicon component as the participant omitted one content word (NOUN1), two points would be allocated for the syntactic complexity as one verb and one noun were used, and one point would have been allocated for the word order as the verb-object order had been maintained. For the constructions produced via graphic symbols in (b), [NOUN1, NOUN2], one point would have been allocated for the lexicon component as one content word [VERB] was omitted, zero points given for syntactic complexity as a verb was not used, and one point given for the word order as the argument order of the SVO structure was maintained.

Two communication disorders students coded the 570 constructions produced by the children. Although the analysis was straightforward, unclear cases were discussed and resolved with the first author. For example, the verb *wear* can take two forms in Hebrew to express the meaning of "the boy is wearing": the first, a transitive verb "lovesh," requires an object, while the other form, an intransitive verb *hitlabesh* (literally "dressed himself"), does not require an object. Therefore, a syntactic score of two points was allocated to both forms.

Analysis

For statistical analysis purposes, the scores of lexicon, syntactic complexity, and word order were converted to percentages (raw scores divided by maximum score for each sematic-syntactic representation) as presented in **Table 2**. A two-way repeated measures ANOVA was performed, with symbol modality (verbal, graphic), syntactic structure (SV, SVO, and SV + SV), and

TABLE 1 | Target constructions (SV, SVO, and Coordination Clauses).

Subject verb	Subject verb object	Coordination sentence
The *boy* is *jumping*	The *boy* is *opening* the *door*	The *boy* is *laughing*, and the *boy* is *crying*
The *boy* is *dancing*	The *boy* is *throwing* the *ball*	The *boy* is *showering*, and the *boy* is *drying*
The *boy* is *sleeping*	The *boy* is *wearing* a *shirt*	The *boy* is *swinging*, and the *boy* is *sliding*
The *boy* is *showering*	The *boy* is *hugging* the *bear*	The *boy* is *playing*, and the *boy* is *jumping*
The *boy* is *laughing*	The *boy* is *holding/reading* a *book*	The *boy* is *walking*, and the *boy* is *sleeping*

TABLE 2 | Metric score for word order, syntactic complexity, and lexicon.

Scoring	Word order	Syntactic complexity	Lexicon
0	Did not maintain proper word order	Arguments only (noun)	Replaced or omitted more than one content word from the target sentence
1	Maintained proper word order	Verb only without arguments	Replaced or omitted one content word from the target sentence
2		Verb + argument [SV]	Retained all target content words
3		Verb + 2 arguments [SVO]	
6		[SV] AND [SV]	

semantic-syntactic representation (lexicon, syntactic complexity, word order) as within-subjects variables. *Post hoc* analyses were performed using least significant differences (LSD) and *t*-tests.

RESULTS

Table 3 presents the mean score of the 19 participants for each semantic-syntactic component (lexicon, syntactic complexity, word order) across syntactic structure (SV, SVO, [SV] + [SV]) and modality (verbal, graphic symbol). The maximum score for each semantic-syntactic component, the mean score (*M*), and the percentage (%) out of the maximum score, are all presented.

Significant main effects were found for symbol modality, $F(1,17) = 17.695$, $p = 0.001$, partial $\eta^2 = 0.510$, syntactic structure, $F(2,34) = 9.935$, $p = 0.000$, partial $\eta^2 = 0.369$, and semantic-syntactic representation, $F(2,34) = 16.193$, $p = 0.000$, partial $\eta^2 = 0.488$. No effect was found for gender, $F(1,17) = 0.358$, $p = 0.558$, partial $\eta^2 = 0.021$. In general, verbal scores were higher ($M = 69.992\%$, $SD = 23.865$) than when graphic symbols were used ($M = 43.704\%$, $SD = 30.443$). Higher scores were obtained for SV ($M = 56.611\%$, $SD = 24.336$) and SV + SV ($M = 64.160\%$, $SD = 23.569$) structures than SVO structures ($M = 49.772\%$, $SD = 27.173$, $LSD = 6.840$, $SE = 2.720$, $p = 0.022$ and $LSD = 14.389$, $SE = 3.246$, $p < 0.001$, respectively). Finally, higher scores were observed for lexicon ($M = 59.602$, $SD = 23.002$) and syntactic complexity ($M = 61.719$, $SD = 19.271$) than for word order ($M = 49.222$, $SD = 29.872$, $LSD = 10.380$, $SE = 2.494$, $p = 0.001$ and $LSD = 12.497$, $SE = 2.911$, $p < 0.001$, respectively).

Interactions of Modality × Syntactic Structure, $F(2,34) = 10.305$, $p = 0.000$, partial $\eta^2 = 0.377$, Modality × Semantic-Syntactic Representation, $F(2,34) = 24.699$, $p < 0.001$, partial $\eta^2 = 0.592$, and Syntactic Structure (Type of Clause) × Semantic-Syntactic Representation, $F(4,68) = 29.914$, $p < 0.001$, partial $\eta^2 = 0.638$, were found, as well as a Modality × Type of Clause × Complexity interaction, $F(4,68) = 4.904$, $p = 0.002$, partial $\eta^2 = 0.224$. **Figure 1** presents scores for each type of syntactic structure when produced verbally and via graphic symbols; **Figure 2** presents scores for each semantic-syntactic component when produced verbally and via symbols. As can be observed from both figures, although the interactions were significant, verbal production gained higher scores than graphic production for all syntactic structures (SV: $t(18) = 2.059$, $p = 0.054$; SVO: $t(18) = 4.440$, $p < 0.001$; SV + SV: $t(18) = 4.701$, $p < 0.001$) and all semantic-syntactic representations (lexicon: $t(18) = 2.294$, $p = 0.034$; syntactic complexity: $t(18) = 4.432$, $p < 0.001$; word order: $t(18) = 4.759$, $p < 0.001$).

Figure 3 presents scores for each syntactic structure (SV, SVO, SV + SV) when produced in verbal and graphic symbols, separately for: (**Figure 3A**) lexicon, (**Figure 3B**) syntactic complexity, and (**Figure 3C**) word order. As can be observed, when divided between semantic-syntactic representation, verbal production earned higher scores than graphic symbols in almost all, but not all, conditions. Verbal production elicited higher lexicon scores in SVO clauses ($t(18) = 3.200$, $p = 0.005$), but not in SV ($t(18) = 1.189$, $p = 0.250$) or SV + SV clauses ($t(18) = 1.455$, $p = 0.163$). Verbal production also resulted in higher syntactic scores in SVO ($t(18) = 4.720$, $p < 0.001$) and SV + SV ($t(18) = 4.443$, $p < 0.001$) clauses, but not in SV ($t(18) = 1.994$, $p = 0.062$) clauses. Finally, verbal production elicited higher word order scores in all types of clauses (SV: $t(18) = 2.516$, $p = 0.022$; SVO: $t(18) = 4.194$, $p = 0.001$; SV + SV: $t(18) = 5.463$, $p < 0.001$) compared to graphic symbols.

DISCUSSION

The goal of this study was to shed light on the impact of the graphic symbol modality on clause construction in Hebrew, a Semitic language with rich morphology, and to explore the relationship between verbal production and the graphic symbol modality. Specifically, this study was designed to examine the effect of the graphic symbol modality on the semantic-syntactic representation (lexicon, syntactic complexity, word order) of different syntactic structures presented in SV, SVO, and coordinated clauses ([SV] AND [SV]) among young typically developing children aged 4–5 years who speak Hebrew.

In general, the young, typically developing children who participated in this study earned higher verbal production scores than graphic production scores for all syntactic structures (SV, SVO, and SV + SV) and for all semantic-syntactic representations (lexicon, syntactic complexity, and word order). The results of this study are consistent with the existing literature in the English language regarding word order and syntactic structure. For example, Smith (1996) found that, among five typically developing preschoolers aged 3:5–4:7, even after 10 weeks of learning and practicing the production of sentences using a board that included 53 PCS, the differences between verbal and AAC production were significant; production using graphic symbols was mostly single-image expression. In another study (Sutton and Morford, 1998) 32 typically developing

TABLE 3 | Sematic-syntactic scores across syntactic structure and modalities.

Syntactic structure		**SV**			**SVO**			**[SV] AND [SV]**		
Sematic-syntactic		**Lexicon**	**Syntactic complexity**	**Word order**	**Lexicon**	**Syntactic complexity**	**Word order**	**Lexicon**	**Syntactic complexity**	**Word order**
Max score		2	2	1	2	3	1	2	6	1
Modality	Verbal *M* (%)	1.38 (68.95)	1.44 (72.11)	0.48 (48.42)	1.11 (55.26)	2.18 (72.63)	0.72 (71.58)	1.45 (72.63)	4.61 (76.84)	0.88 (88.42)
	GS *M* (%)	1.25 (62.63)	1.24 (62.63)	0.24 (24.21)	0.66 (33.16)	1.11 (36.84)	0.28 (28.42)	1.28 (64.21)	2.95 (49.12)	0.33 (32.63)

GS = Graphic symbol.

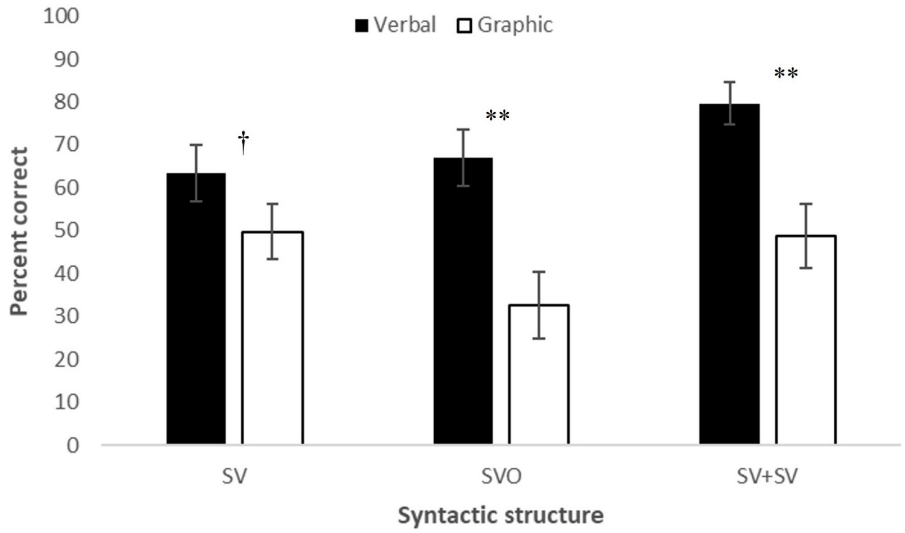

FIGURE 1 | Verbal and graphic scores for different syntactic structures. $^{\dagger}p = 0.05$; $^{**}p < 0.001$.

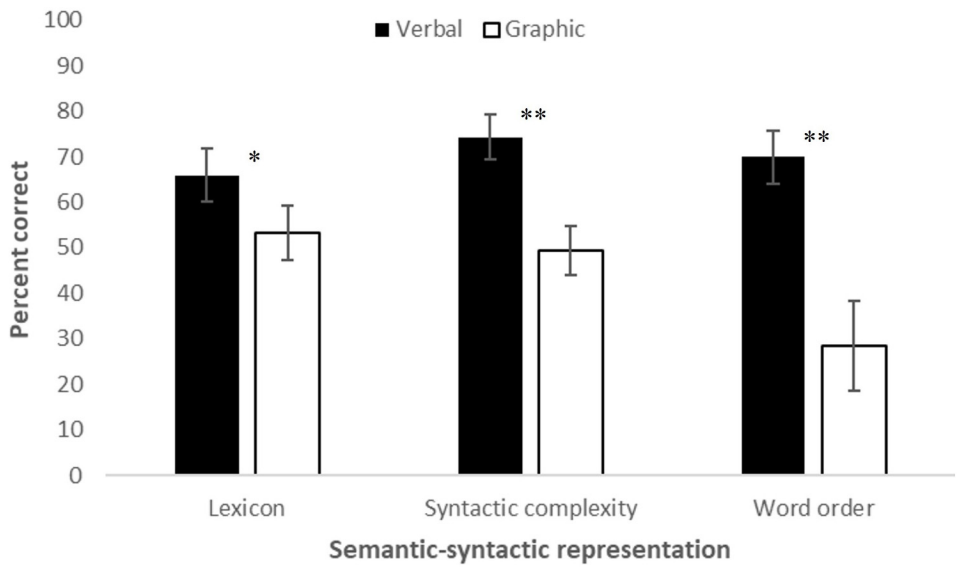

FIGURE 2 | Verbal and graphic scores for the different sematic-syntactic representation. $^{*}p < 0.05$; $^{**}p < 0.001$.

children aged 5:9–12:7 were tasked with producing 24 SVO structures (both verbally and through graphic symbols) by watching videos of an agent performing a transitive action on a patient. More than half of the responses from the kindergarten-age group did not follow English constituent order; moreover, although the older group performed better than the young group, their results still showed significant differences in English constituent order between verbal and graphic symbol responses. However, it is important to note that recent studies focusing on improving the language outcome of children who use AAC have shown that children with severe speech disorders who received adequate training based on appropriate intervention techniques can easily learn to produced SVO structures and rule-based messages via graphic symbols (Binger et al., 2017, 2020; Soto et al., 2020).

These findings strengthen the notion that atypical structures produced via graphic symbols are related to the graphic symbol modality and not to the child's linguistic knowledge. Moreover, the superior results produced by using the verbal modality over the graphic symbol modality can be also explained by the general notion that the ability to process external (visual) representation is unlike processing spoken language, a universal ability for all typically developing children (Lee and Karmiloff-Smith, 1996).

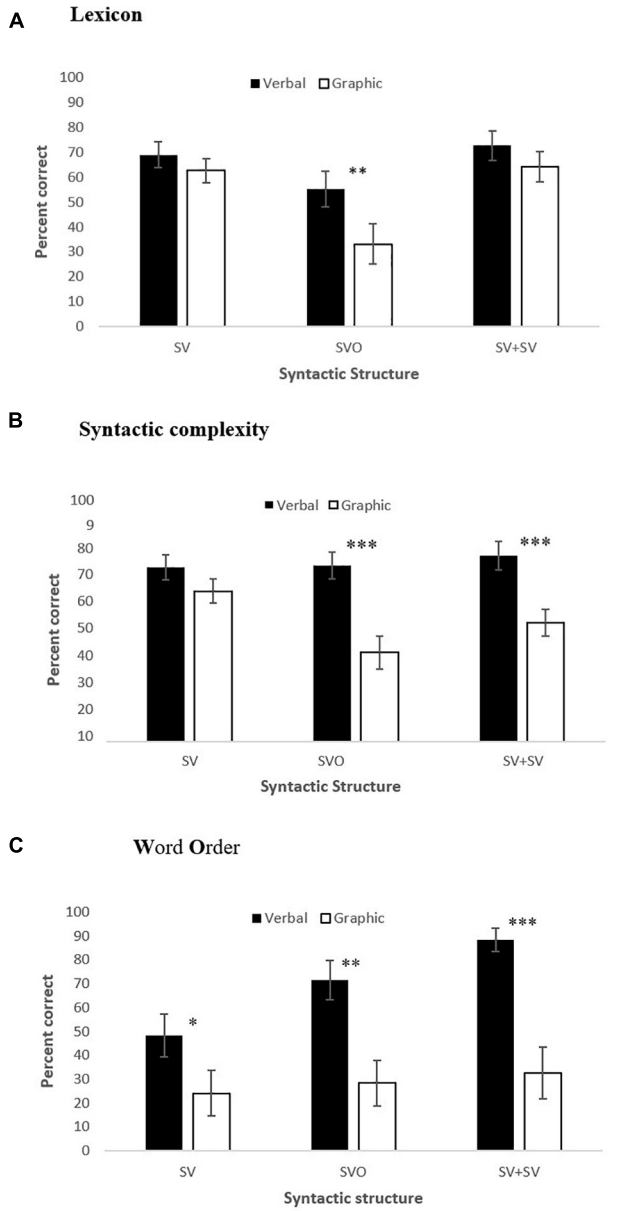

FIGURE 3 | Scores for SV, SVO, and SV + SV clauses when produced in verbal and graphic symbols, separately for **(A)** lexicon, **(B)** syntactic complexity, and **(C)** word order. **(A)** $^{**}p < 0.01$. **(B)** $^{***}p < 0.001$. **(C)** $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The set of graphic symbols used in this study for representing content words (nouns and verbs) included PCS graphic symbols. While this set of symbols is designed to be as transparent as possible, a letter was used for representing the functional word "*and*." Accordingly, the set of visual forms used in this study vary in their level of iconicity, ranging from transparent to translucent. Previous psycholinguistic studies found that preliterate young children differentiate between drawing and writing, as demonstrated in sorting tasks in which they were asked to decide which combination of elements belongs to a specific notational system. However, in production tasks when they were asked to "write a letter to a friend or "leave a message" their focus remained on the content they wanted to convey, and drawings were therefore used to express that content (Landsmann and Karmiloff-Smith, 1992). This discrepancy is explained by the distinction between notation as a domain of knowledge and notation as a domain referential-communicative tool in which the focus is on the content and usage of drawing. Moreover, Landsmann and Karmiloff-Smith (1992) also explain that one of the distinctions between drawing and writing is the relative closure constraint that is similar to the distinction made in linguistics between open class categories (e.g., nouns and verbs) and closed class categories (e.g., articles and conjunctions); within the open class category, it is always possible to add new elements, while within the closed class the set of elements is finite.

Examination of the sematic-syntactic representation of different syntactic structures in this study revealed that verbal production earned higher scores than graphic production in almost all, but not all, structures. The results revealed an effect of structural complexity and lexicon on graphic symbol production in SVO and coordinated clauses ([SV] AND [SV]) but not in SV structure. This might be explained by SV structure being more directly transmitted onto the graphic symbol modality due to the: (1) Domain of referential-communicative tool: preschoolers prefer to focus on content words (nouns and verbs) represented by iconic symbols to convey their message, and (2) Domain of knowledge: avoidance of the lexicon and syntactic modification of longer and complex structures requires a verb complement and use of a functional word (and) represented by a non-iconic symbol.

Lexicon and Modality

Children in this study earned higher lexicon scores during verbal production, compared to graphic symbol modality, only in the SVO structure, but not in the SV or SV + SV structures. One explanation for this outcome may be the nature of the task demands. The children were asked to watch a video and verbally express the semantic relation of the verb to its arguments and then to express it via graphic symbols. In the SVO structure, the children needed to identify the relation of the agent (Subject) and the person affected by the action (Object), as well as express the target verb with its two arguments; in the SV and SV + SV structures, they needed to identify the relation of the verb with one argument. As such, SVO structures may lead to less accuracy than SV structures when selecting the target content words (verbs and nouns) via graphic symbols due to lexical voids—missing words in the communication board.

Another explanation may be a strategy of enhancing the communication pace by expressing only specific content words because the video content is known to the child and the examiner and is therefore shared knowledge and common ground. Moreover, some transitive verbs may be less transparent than others, and their meaning dependent on context. For example, the transitive verb HOLD, used in this study as a target verb, was notably a non-transparent graphic symbol, as noted during the familiarization phase. The video presenting the verb HOLD with a boy holding a book was also not clear: seven

out of 19 participants verbally indicated the verb READ instead of HOLD ("the boy is reading a book" instead of "the boy is holding a book"). As this answer was not expected, the verb "read" was missing from the communication board, and thus resulted in a lexical void.

Syntactic Complexity and Modality Use

Children in this study gained higher syntactic scores in SVO and SV + SV, but not in SV, structures. These findings may also be explained by the task demands, as the short constructions of the SV structure may be more easily transmitted onto the graphic symbol modality. These results are similar to the results found English and French, in which structural complexity may play an important role in graphic symbol construction. Short constructions require less modification of spoken constituent order and a lower level of linguistic analysis to complete the task (Trudeau et al., 2007; Sutton et al., 2020).

Furthermore, constructing coordinated clauses may require metalinguistic skills and exposure to formal writing instruction. Representing functional words, such as the conjunction word "*and*," is challenging as its level of iconicity is very low. Such words are often represented as "sight words" on the communication board; in this study, the word "*and*" was represented by its Hebrew written form. Only 4 (21%) participants identified the letter that represented the word "*and*" when it was first introduced during the familiarization phase. The lack of formal writing instruction and literacy skills among the preschool children affected their ability to transpose the word "*and*" into the graphic symbol modality, resulting in atypical structures of two coordinated clauses, even though they possessed this structure in their mental representation.

Word Order and Modality

Children in this study earned higher word order scores in all types of clauses (SV, SVO, and SV + SV) using verbal production compared to graphic symbols. As stated before, metalinguistic knowledge and literacy skills may be required to create graphic symbol constructions that maintain the verbal production order. Therefore, it is not surprising that young, typically developing 4- to 5-year-old children did not maintain the word order in graphic symbols. This observation is consistent with graphic symbol findings in other languages that demonstrated metalinguistic skills develop gradually in the early school years and ultimately affect children's abilities to transmit complex sentences into graphic symbols (Trudeau et al., 2007; Sutton et al., 2020). Moreover, this finding is consistent with those found in English in which typically developing children often produced graphic symbol constructions that did not follow the canonical word order of spoken sentences (Smith, 1996; Trudeau et al., 2010a,b).

Implications for Clinical Practice

Cross-linguistic studies on graphic symbols are necessary to shed light on the characteristics and constraints of the graphic symbol modality beyond individual differences and linguistic knowledge. Across various languages, including Hebrew, individuals who cannot use their natural speech and who are not yet literate use the same set of graphic symbols (e.g., PCS) to transmit their thoughts and express ambient language. In recent years, there has been impressive progress in designing evidence-based language interventions that enhance the linguistic outcome of individuals who use AAC (Binger et al., 2017; Soto and Clarke, 2017). Moreover, the advanced technology provides access to a wide range of grammatical markers via graphic symbols that support morpho-syntactic representation. As language is developed through language use (Tomasello, 2009), it is essential to provide explicit instruction that supports language growth for individuals who use graphic symbols.

However, constructing the structures of the spoken language via graphic symbols is still a challenge. The question remains to what extent atypical clause structures observed via the graphic symbol modality relate to intrinsic factors of individuals who utilize AAC or the modality itself.

Exploring how typical-developing children who are not yet literate use graphic symbols without adequate training is necessary to understand the relationship between verbal production and the graphic symbol modality. Based on the current findings, professionals working with children aided by AAC as their main modality of communication should take into consideration that atypical construction may not reflect linguistic knowledge. During formal AAC interventions aimed at transmitting spoken language utterances onto the graphic symbol modality, the following should be considered:

(1) Subject verb structures with non-transitive verbs require the least metalinguistic demands and modifications of spoken utterances, specifically when using iconic symbols (e.g., sleep, slide, swing, walk, play).
(2) Subject–verb–object structures impose additional challenges, both in the selection of the graphic symbols that represent the agent who initiates the action and the person affected by the action, and in the ordering of verb arguments in the canonical order. Semantically transitive verbs may be less iconic (such as the verb HOLD) and require further instruction to learn the symbol meaning; and
(3) Constructing two coordinated clauses using the coordination word "*and*" via graphic symbols requires the additional literacy skill of identifying sight words, non-iconic symbols, as well as receiving formal instruction. This is due to the coordination word "*and*" not being iconic and being represented by a sight word.

Limitations and Future Research Needs

The primary goal of this study was to extend previous findings that graphic symbol construction may differ from verbal utterances regardless of the level of linguistic knowledge of the spoken language and regardless of the specific language. Therefore, the experimental tasks for this study were designed for typically developing Hebrew-speaking children who were not literate but had mastered the syntactic structures used in this study. Various factors may affect the translation of spoken language onto graphic symbols including metalinguistic demands

presented by the complexity of the structures, exposure to formal writing, level of symbol iconicity, and symbol availability on the communication board. The children in study received minimal exposure and training in graphic symbol use, and the tasks presented did not examine different metalinguistic knowledge and literacy skills. To generalize these findings and further explore the relationship between verbal and graphic symbol production in Hebrew, additional research is needed- among school-aged children and adults, and with larger samples, longer training periods; and constructions including the use of various functional words. Future research is needed to explore the patterns of non-canonical word order and types of content words that were omitted in the graphic modality.

CONCLUSION

The results of this study indicate that typically developing 4–5 year old Hebrew-speaking children display semantic-syntactic representation via graphic symbols that differs from spoken language. Verbal production was superior in lexicality, syntactic complexity, and word order across different task demands presented by different syntactic structures. Differences were notable in structures with transitive verbs (SVO) and two coordinated clauses ([SV] AND [SV]), but not in SV structures. Although preschoolers have the mental representation of these structures, when utilizing the graphic symbols as a referential-communicative tool their focus appears to be on the content they want to transmit, resulting in a focus on content word and iconic symbols. Explicit instruction appears needed to use functional words within clause construction via graphic symbols as these words are represented by non-iconic symbols and relate to the domain of knowledge and do not serve as referential communicative tools.

Similarities between our findings and those in English point to the notion that atypical structures produced via graphic symbols are related to the modality itself and the task demands, not to a child's linguistic knowledge and specific language. All our conclusions pertain to young children who speak Hebrew; it remains to be seen whether the same relationships hold up across various ages with different clause structures, functional words, literacy and metalinguistic skills.

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 study was approved by the IRB of Ariel University (AU-HEA-GH-20190130-1) and was conducted in accordance with appropriate ethical standards. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

GS-H worked on the study design, literature review, data scoring, and writing. LF worked on the study design, data analysis, and writing. All authors listed made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We are especially grateful to Sharon Armon-Lotem for comments that greatly improved the manuscript. Gila Hermon and Liron Lipnick Traivitz also played an important role in collecting and coding the data. Warm thanks are extended to all of them.

REFERENCES

American Speech-Language-Hearing Association (2020). *Augmentative and Alternative Communication (Practice Portal)*. Available online at: www.asha.org/Practice-Portal/Professional-Issues/Augmentative-and-Alternative-Communication/ (accessed November 28, 2020).

Atkin, A. (2014). "Peirce's theory of signs," in *The Stanford Encyclopedia of Philosophy (Summer*, 2013 Edn, ed. E. N. Zalta (Amsterdam: Elsevier).

Berman, R. (1985). "The acquisition of Hebrew," in *The Crosslinguistic Study of Language Acquisition, Volume 1: The Data*, ed. D. Slobin (Hillsdale, NJ: Lawrence Erlbaum), 255–371.

Berman, R. A., Lieber, R., and Štekauer, P. (2009). "Acquisition of compound constructions," in *Handbook of Compounding*, eds R. Lieber and P. Štekauer (Oxford: Oxford University Press), 298–322.

Berman, R. A., and Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Erlbaum.

Binger, C., Kent-Walsh, J., Harrington, N., and Hollerbach, Q. C. (2020). Tracking early sentence-building progress in graphic symbol communication. *Lang. Speech Hear. Serv. Sch.* 51, 317–328. doi: 10.1044/2019_LSHSS-19-00065

Binger, C., Kent-Walsh, J., King, M., and Mansfield, L. (2017). Early sentence productions of 3-and 4-year-old children who use augmentative and alternative communication. *J. Speech Lang. Hear Res.* 60, 1930–1945. doi: 10.1044/2017_jslhr-l-15-0408

Binger, C., and Light, J. (2008). The morphology and syntax of individuals who use AAC: research review and implications for effective practice. *Augment. Altern. Commun.* 24, 123–138. doi: 10.1080/07434610701830587

Blockberger, S., and Johnston, J. R. (2003). Grammatical morphology acquisition by children with complex communication needs. *Augment. Altern. Commun.* 19, 207–221. doi: 10.1080/07434610310001598233

Clark, E. V., and Berman, R. A. (1987). Types of linguistic knowledge: interpreting and producing compound nouns. *J. Child Lang.* 14, 547–567. doi: 10.1017/s030500090001028x

Clark, J. (1997). *Symbolstix.* News 2 You. Available online at: https://www.n2y.com/symbolstix-prime

Dromi, E., Leonard, L. B., and Shteiman, M. (1993). The grammatical morphology of Hebrew-speaking children with specific language impairment: some competing hypotheses. *J. Speech Lang. Hear Res.* 36, 760–771. doi: 10.1044/jshr.3604.760

Fitzgerald, E. (1969). *Straight Language for the Deaf*. Washington, DC: Volta Bureau.

Fuller, D. R., Lloyd, L. L., and Schlosser, R. W. (1992). Further development of an augmentative and alternative communication symbol taxonomy.

Augment. Altern. Commun. 8, 67–74. doi: 10.1080/0743461921233127 6053 doi: 10.3109/07434618.2011.592217

Glinert, L. (2017). *Modern Hebrew: An Essential Grammar*, 3rd Edn. London: Routledge.

Goralnik, E. (1995). *Goralnik Screening Test for Hebrew*. Even Yehuda: Matan.

Kennedy, G. (2004). *First Keys 2,* Vol. 9, Leamington Spa: Widgit Software Ltd. 28–30. doi: 10.1080/19404150409546753

Kovacs, T., and Hill, K. (2017). Language samples from children who use speech-generating devices: making sense of small samples and utterance length. *Am. J. Speech Lang. Pathol.* 26, 939–950. doi: 10.1044/2017_AJSLP-16-0114

Landsmann, L. T., and Karmiloff-Smith, A. (1992). Children's understanding of notations as domains of knowledge versus referential-communicative tools. *Cogn. Dev.* 7, 287–300. doi: 10.1016/0885-2014(92)90017-l

Lee, K., and Karmiloff-Smith, A. (1996). "The development of external symbol systems," in *Perceptual and Cognitive Development*, eds R. Gelman and T. Kit-Fong (Cambridge, MA: Academic Press).

Loncke, F. (ed.) (2014). "AAC: a general introduction," in *Augmentative and Alternative Communication Models and Applications For Educators, Speech-Language Pathologist, Psychologist, Caregivers, and Users*, (San Diego, CA: Plural Publishing), 1–17. doi: 10.14818/aac.2015.6.3.1.1

Mano-Lerman, L. (2017). *Linguistic and Communicative Characteristics of Experienced AAC Users in Different Production Tasks*. [Master's thesis]. Haifa: University of Haifa.

Mavrou, K., Charalampous, E., and Michaelides, M. (2013). Graphic symbols for all: using symbols in developing the ability of questioning in young children. *J. Assist. Technol.* 7, 22–33. doi: 10.1108/17549451311313192

Mayer-Johnson (1981–2011). *Boardmaker [Software]*. Solana Beach, CA: Mayer-Johnson LLC.

Müller, E., and Soto, G. (2002). Conversation patterns of three adults using aided speech: variations across partners. *Augment. Altern. Commun.* 18, 77–90.

Pampoulou, E., and Fuller, D. R. (2020). Exploring AAC graphic symbol choices: a preliminary study. *J. Enabling Technol.* 14, 171–185. doi: 10.1108/JET-03-2020-0013

Redmond, S. M., and Johnston, S. S. (2001). Evaluating the morphological competence of children with severe speech and physical impairments. *J. Speech Lang. Hear. Res.* 44, 1362–1375. doi: 10.1044/1092-4388(2001/106)

Savaldi-Harussi, G., Lustigman, L., and Soto, G. (2019). The emergence of clause construction in children who use speech generating devices. *Augment. Altern. Commun.* 35, 109–119. doi: 10.1080/07434618.2019.1584642

Savaldi-Harussi, G., and Soto, G. (2016). Using SALT: considerations for the analysis of language transcripts of children who use SGDs. *Perspect. ASHA Spec. Interest Groups* 1, 110–124. doi: 10.1044/persp1.sig12.110

Savaldi-Harussi, G., and Soto, G. (2018). Early verbal categories and inflections in children who use speech-generating devices. *Augment. Altern. Commun.* 34, 194–205. doi: 10.1080/07434618.2018.1490925 doi: 10.1080/07434618.2018.1490925

Slobin, D. I., and Bever, T. G. (1982). Children use canonical sentence schemas: a crosslinguistic study of word order and inflections. *Cognition* 12, 229–265. doi: 10.1016/0010-0277(82)90033-6

Smith, M. (1996). "The medium or the message: a study of speaking children using communication boards," in *Augmentative and alternative communication: European perspectives*, eds S. von Tetzchner and M. H. Jenson (San Diego, CA: Singular Publishing Group), 119–136.

Smith, M. (2006). Speech, language and aided communication: connections and questions in a developmental context. *Disabil. Rehabil.* 28, 151–157. doi: 10.1080/09638280500077747

Smith, M. (2015). Language development of individuals who require aided communication: reflections on state of the science and future research directions. *Augment. Altern. Commun.* 31, 215–233. doi: 10.3109/07434618.2015.1062553

Soto, G. (1997). "Multi-unit utterances and syntax in graphic communication," in *Communication…naturally. Theoretical and Methodological Issues in Research in Augmentative and Alternative Communication: Proceedings of the Fourth ISAAC Research Symposium*, eds B. Åkesson and P. Linsay (Toronto: Malardalen University Press), 26–32.

Soto, G., Clarke, M., Nelson, K., Starowicz, R., and Savaldi-Harussi, G. (2020). Recast type, repair, and acquisition in AAC mediated interaction. *J. Child Lang.* 47, 250–264. doi: 10.1017/S0305000919000436

Soto, G., and Clarke, M. T. (2017). Effects of a conversation-based intervention on the linguistic skills of children with motor speech disorders who use augmentative and alternative communication. *J. Speech Lang. Hear Res.* 60, 1980–1998. doi: 10.1044/2016_JSLHR-L-15-0246

Sutton, A., and Gallagher, T. (1993). Verb class distinctions and AAC language Encoding implications. *J. Speech Hear. Res.* 36, 1216–1226. doi: 10.1044/jshr.3606.1216

Sutton, A., Soto, G., and Blockberger, S. (2002). Grammatical Issues in graphic symbol communication. *Augment. Altern. Commun.* 18, 192–204. doi: 10.1080/07434610212331281271

Sutton, A., Trudeau, N., Morford, J., Rios, M., and Poirier, M. (2010). Preschool-aged children have difficulty constructing and interpreting simple utterances composed of graphic symbols. *J. Child Lang.* 37, 1–26. doi: 10.1017/S0305000909009477

Sutton, A., Trudeau, N., Morford, J. P., and Smith, M. M. (2020). Expressive and receptive use of speech and graphic symbols by typically developing children: what skills contribute to performance on structured sentence-level tasks? *Int. J. Speech Lang. Pathol.* 23, 155–167. doi: 10.1080/17549507.2020.1756406

Sutton, A. E., and Morford, J. P. (1998). Constituent order in picture pointing sequences produced by speaking children using AAC. *Appl. Psycholinguist.* 19, 525–536. doi: 10.1017/S014271640001034

Theodorou, E., Kambanaros, M., and Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilectal children with SLI. *Front. Psychol.* 8:2104. doi: 10.3389/fpsyg.2017.02104

Tomasello, M. (2009). "The usage-based theory of language acquisition," in *The Cambridge Handbook of Child Language*, ed. E. L. Bavin (Cambridge, MA: Cambridge Univ. Press), 69–87. doi: 10.1017/cbo9780511576164.005

Treiman, R., Cohen, J., Mulqueeny, K., Kessler, B., and Schechtman, S. (2007). Young children's knowledge about printed names. *Child Dev.* 78, 1458–1471.

Trudeau, N., Morford, J. P., and Sutton, A. (2010a). The role of word order in the interpretation of canonical and non-canonical graphic symbol utterances: a developmental study. *Augment. Altern. Commun.* 26, 108–121. doi: 10.3109/07434618

Trudeau, N., Sutton, A., Morford, J., Cote-Giroux, P., Pauze, A.-M., and Vallee, V. (2010b). Strategies in construction and interpretation of graphic-symbol sequences by individuals who use AAC systems. *Augment. Altern. Commun.* 26, 299–312. doi: 10.3109/07434618.2010.529619

Trudeau, N., Sutton, A., Dagenais, E., de Broeck, S., and Morford, J. (2007). Construction of graphic symbol utterances by children, teenagers, and adults: the effect of structure and task demands. *J. Speech Lang. Hear Res.* 50, 1314–1329. doi: 10.1044/1092-4388(2007/092)

Vinder, H. (2016). *How Do Augmentative and Alternative Communication Users Deal With Lexical Voids?* [Master's thesis]. Haifa: University of Haifa.

Von Tetzchner, S., and Grove, N. (2003). "The development of alternative language forms," in *Augmentative and Alternative Communication Developmental Issues*, eds S. von Tetzchner and N. Grove (London: Whurr), 1–27. doi: 10.1002/9781118784235.eelt0387

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Savaldi-Harussi and Fostick. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fcomm.2021.689009

A Computer-Assisted Writing Tool for an Extended Variety of *Leichte Sprache* (Easy-to-Read German)

Karin Harbusch* and Ina Steinmetz

Institute for Computational Visualistics, Computer Science Faculty, Universität Koblenz-Landau, Koblenz, Germany

Leichte Sprache (LS; easy-to-read German) defines a variety of German characterized by simplified syntactic constructions and a small vocabulary. It provides barrier-free information for a wide spectrum of people with cognitive impairments, learning difficulties, and/or a low level of literacy in the German language. The levels of difficulty of a range of syntactic constructions were systematically evaluated with LS readers as part of the recent LeiSA project (Bock, 2019). That study identified a number of constructions that were evaluated as being easy to comprehend but which fell beyond the definition of LS. We therefore want to broaden the scope of LS to include further constructions that LS readers can easily manage and that they might find useful for putting their thoughts into words. For constructions not considered in the LeiSA study, we performed a comparative treebank study of constructions attested to in a collection of 245 LS documents from a variety of sources. Employing the treebanks TüBa-D/S (also called VERBMOBIL) and TüBa-D/Z, we compared the frequency of such constructions in those texts with their incidence in spoken and written German sources produced without the explicit goal of facilitating comprehensibility. The resulting extension is called *Extended Leichte Sprache* (ELS). To date, text in LS has generally been produced by authors proficient in standard German. In order to enable text production by LS readers themselves, we developed a computational linguistic system, dubbed *ExtendedEasyTalk*. This system supports LS readers in formulating grammatically correct and semantically coherent texts covering constructions in ELS. This paper outlines the principal components: (1) a natural-language paraphrase generator that supports fast and correct text production while taking readership-design aspects into account, and (2) explicit coherence specifications based on *Rhetorical Structure Theory* (RST) to express the communicative function of sentences. The system's writing-workshop mode controls the options in (1) and (2). Mandatory questions generated by the system aim to teach the user when and how to consider audience-design concepts. Accordingly, users are trained in text production in a similar way to elementary school students, who also tend to omit audience-design cues. Importantly, we illustrate in this paper how to make the dialogues of these components intuitive and easy to use to avoid overtaxing the user. We also report the results of our evaluation of the software with different user groups.

Keywords: Augmentative and Alternative Communication (AAC), Controlled Languages (CL), Plain Language, Natural Language Generation (NLG), paraphrase generation, writing workshop/Schreibwerkstatt, Leichte Sprache corpus, treebank study

OPEN ACCESS

Edited by:
Scott A Hale,
University of Oxford, United Kingdom

Reviewed by:
Anouschka Foltz,
University of Graz, Austria
Hassan Mohebbi,
European Knowledge Development Institute (EUROKD), Turkey

****Correspondence:***
Karin Harbusch
harbusch@uni-koblenz.de

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

Received: *31 March 2021*
Accepted: *30 September 2021*
Published: *03 January 2022*

Citation:
Harbusch K and Steinmetz I (2022) A Computer-Assisted Writing Tool for an Extended Variety of Leichte Sprache *(Easy-to-Read German). Front. Commun. 6:689009. doi: 10.3389/fcomm.2021.689009*

1 INTRODUCTION

Leichte Sprache (LS) is a simplified variety of German. It was developed as part of the plain language movement of the 2000s (cf. easy-to-read English), which aimed to produce easy-to-understand texts for people with intellectual disabilities or learning difficulties (Bredel and Maaß, 2016, p. 60), who often have low literacy skills (Light et al., 2019). In Germany, LS is enshrined in law as the means of choice for providing accessible information in text form (BITV2.0, 2011).

The LeiSA project[1] identified a range of easily comprehensible syntactic constructions that are nonetheless beyond the scope of the core LS rules. It can be assumed that these constructions are used in language production, i.e., for putting thoughts into words. This leads to a research question concerning the *target grammar of our system*: What constructions might LS readers like to use in a writing tool? In order to obtain quantitative estimates of the incidence of the constructions evaluated in the LeiSA study, we built a parsed corpus of 245 published LS documents (a *Leichte Sprache treebank* we call LST). Constructions were found to have reasonable incidence, and no more than medium difficulty was included. The frequencies in LST of syntactic structures that had not yet been evaluated were compared to their frequencies in two treebanks of spoken and written German (VERBMOBIL and TüBa-D/Z, respectively), i.e., texts that had not been produced with the explicit goal of facilitating comprehensibility. The frequency of a construction's occurrence in the spoken corpus was compared with its frequency in the written one to provide the basis for an estimate of its ease of production; however, in order to keep the number of additional constructions to a minimum, we also judged whether or not a pure LS construction could easily replace one that is not included. The resulting extension of LS is called *Extended Leichte Sprache* (ELS).

To date, it has been usual for texts in LS to be produced by authors proficient in standard German and then evaluated for ease of comprehension by people with intellectual disabilities or learning difficulties (BITV2.0, 2011; Netzwerk Leichte Sprache, 2013). One factor preventing LS readers from producing texts themselves may be the lack of technical support during the process from *message conceptualization* (in the mind of the speaker/writer) to *sentence realization* (in a computer-assisted writing tool that remedies reading/writing deficits). Here, we consult terms used in *natural-language generation* (NLG) (cf. **Section 3.1.2** for more details) to illustrate the complexity of the language production process involved in producing a text, compared to that of producing oral utterances in a face-to-face conversation.

To the best of our knowledge, there is no easy-to-use LS writing system that offers linguistic support beyond the phrasing of simple, partly personal sentences, let alone a system capable of teaching the concepts of written text production. The writing of coherent, understandable text requires an emphasis on audience-design concepts (Bell, 1984) because its target (i.e., the reader) cannot seek clarification—unlike the listener in face-to-face communication. German elementary school children learn written text production by the widely applied method of the *Schreibwerkstatt/Schreibkonferenz* 'writing workshop' (see, e.g., Reichardt et al. (2014) for a broad survey). This technique is comparable to sentence-combining exercises in the Anglo-Saxon language area that teaches students to integrate sets of short, disconnected sentences to form longer, more effective ones (see Nordquist (2018) for an online introduction, Ney (1980) for the history, and Saddler and Preschern (2007) for the school context). This leads to two research questions concerning *assisted writing*: What individual support can help a range of users with intellectual disabilities, learning difficulties, and/or low literacy skills to write understandable, coherent text in ELS? Can we transform concepts from all stages in NLG into intuitive dialogues at the individual LS-reader level?

We present *ExtendedEasyTalk*, a writing tool with its main emphases on the extensive use of linguistic processing and on interactive user guidance aimed at compensating for a lack of grammatical knowledge and ensuring syntactic correctness and understandability. In order to produce a coherent text, *ExtendedEasyTalk* actively stimulates the user to add text-understandability and text-coherence elements, at both the constituent structure and the sentence-combining levels. For example, the sentences in (1-a) express the train of thought much better than the staccato phrases in (1-b), thanks to the use of coherence cues (*therefore/so, tomorrow*, and the colon).

(1) a. Du bist krank. Darum: Ich will Dir helfen: Ich koche morgen das Mittagessen.
You are ill. Therefore: I want you to_help: I cook tomorrow the lunch.
'You are ill, so I want to help you; I will cook lunch tomorrow.'

b. Du bist krank. Ich helfe Dir. Ich koche Mittagessen.
You are ill. I help you. I cook lunch.
'You are ill. I will help you. I will cook lunch.'

This paper is organized as follows. In **Section 2**, we define ELS. After briefly introducing the core LS constructions, we list those evaluated in the LeiSA project (see **Section 2.2**). Then, we outline the comparative corpus study into constructions LS readers are likely to use in communication. (Readers wishing to skip the detailed linguistic argumentations can go directly to **Table 8**, which provides a list of the constructions included in ELS.) In **Section 3**, we present *ExtendedEasyTalk* as follows: First, we summarize the state-of-the-art technical writing support in the research area of *Augmentative and Alternative Communication* (AAC); a particular highlight of the descriptions is automatic NLG. Then, we give an intuitive overview of how the system works (see **Section 3.2**), before going into the technical details of sentence-constituent and sentence-coherence production (see **Sections 3.3** and **3.4**, respectively). We also outline the active control mechanism of *ExtendedEasyTalk* for teaching text-production concepts (see **Section 3.5**). Finally, we give the results of our evaluation with different user groups. **Section 4** draws some conclusions, discusses open issues, and suggests directions for future work.

2 EXTENDED LEICHTE SPRACHE

In this section, we define ELS, the target language of *ExtendedEasyTalk*. First, the syntactic constructions included in LS are outlined; this is followed by the evaluation results of

[1]https://research.uni-leipzig.de/leisa/ (2014–2018).

the LeiSA study. In **Section 2.3**, the treebank study is presented as follows: First, the creation of an LS treebank, LST, is described (see **Section 2.3.1**). Then, we quantify syntactic constructions in LST and compare them to standard German to classify their ease of comprehension. Finally, we give a summary of all additional constructions included in ELS, i.e., those that extend LS.

2.1 Leichte Sprache

Controlled/simplified natural languages, like Basic English (Ogden, 1930), have long been a topic of great interest (see Kuhn (2014) for a broad survey). The rules for Leichte Sprache were originally derived from practical experience (Bredel and Maaß, 2016, p. 60). The three main rule books (Netzwerk Leichte Sprache, 2013; Inclusion Europe, 2009; BITV2.0, 2011) have been the subject of previous scientific investigation (Lieske and Siegel, 2014; Maaß et al., 2014; Löffler, 2015; Zurstrassen, 2015; Bredel and Maaß, 2016; Bock, 2019; Nüssli, 2019; Pottmann, 2019; Hansen-Schirra and Maaß, 2020). Many rules concern the vocabulary (e.g., "Use easy words" or "No abbreviations") or the avoidance of complex structures, for example, the use of:

- metaphors;
- more than one statement per sentence;
- punctuation other than: ".", "?", "!", ":";
- complex clauses;
- inversions;
- the genitive case;
- the passive voice;
- the subjunctive mood; and
- the simple past tense.

In other words, only main clauses are included in LS. In main declarative clauses, the canonical word order is subject–verb–object (SVO). All sentences should be phrased in the active voice, indicative mood, and present or present perfect tense.

The primary LS audience of people with cognitive impairments or learning difficulties is very heterogenous, and the available authentic text data by which to identify the range of constructions, LS readers naturally use in the process of formulating an idea are sparse. Usually, LS texts are written by authors proficient in standard German. Contrary to the recommendation in Netzwerk Leichte Sprache and Inclusion Europe, ease of comprehension is not always tested by members of the target readership.

Inspired by the finding of the LeiSA study that the majority of easily understandable LS texts do not strictly adhere to LS rules, we have explored possibilities for extending those rules to include the syntactic constructions that LS readers are likely to use when putting their thoughts into words.

2.2 Observations From the LeiSA Study

Part of the LeiSA study (Bock, 2019) was concerned with estimating the comprehension difficulty of individual syntactic constructions (see **Table 1**). Through well-established comprehension tests conducted using a five-point scale, constructions—not only within but also beyond the scope of LS—were classified according to the error rates measured in experiments with participants with intellectual disabilities and functional illiteracy. We refer to the error ranges by the following difficulty levels:

- 0% $\leq$ error rate $\leqq$ 5.9%: **easy**;
- 5.9% $<$ error rate $\leqq$ 12.5%: **low**;
- 12.5% $<$ error rate $\leqq$ 37.5%: **medium**;
- 37.5% $<$ error rate $\leqq$ 47.3%: **high**; and
- 47.3% $<$ error rate $\leqq$ 75.9%: **extreme.**

Constructions that show low error rates in the comprehensibility tests can be expected to be included in ELS. However, how can the full range of such constructions be obtained? We sought to identify the syntactic structures that LS readers are likely to use when putting their thoughts into words. In the following, we describe our search strategy, which leads to a broader set of rules, supported by the *ExtendedEasyTalk* system (see **Table 8** for a summary of the additionally included constructions).

2.3 A Comparative Treebank Study With Three German Corpora

Syntactic constructions that do not adhere to the LS rules are not hard to find in published LS documents. This suggests the advisability of inspecting a broad collection of LS texts and analyzing the constructions found therein. This requires the use of a Leichte Sprache treebank. As, to the best of our knowledge, no previous syntactically annotated corpus of LS texts exists, we have created one: LST (see **Section 2.3.1**).

Not all of the constructions we found in LS texts were evaluated with LS readers in the LeiSA project. In order to identify easy-to-understand constructions, we employed a treebank study as an alternative to the evaluation of example sentences by LS readers and compared the frequencies of constructions in LST to those in spoken and written standard German. We argue that constructions with high frequencies in spoken German are easy to produce due to the time-pressured nature of speech production. In a written text, the author is able to embellish the text in revision cycles, replacing simple constructions with more complex ones. Thus, the written corpus serves as a baseline. The appearance of a given construction in spoken language with a frequency higher than or equal to its appearance in written language is indicative of an easy-to-understand or unavoidable construction. Conversely, higher frequencies in written text imply difficult constructions employed under the non-time-critical conditions of revision and editing.

For the quantification of syntactic constructions in standard German, we used TüBa-D/S (also called VERBMOBIL), a treebank of spoken German, and TüBa-D/Z, a treebank of written German. To avoid confusion, we will use the name VERBMOBIL for TüBa-D/S. In the VERBMOBIL project (see, e.g., Stegmann et al., 2000 or Wahlster, 2000), more than 400 spontaneously produced spoken dialogues (concerning

TABLE 1 | Overview of the syntactic constructions evaluated in the LeiSA study (Column 1: Difficulty level; Column 2: Is the concept included in the core rules of LS? (Y/N); Column 3: Characterization of the syntactic construction; Column 4: Examples from the LeiSA study; and Column 5: English translations).

Difficulty	LS	Linguistic phenomenon	LeiSA-study example	English translation
easy	N	Negations using *nicht* 'not'	*Der Bus fährt nicht*.	'The bus isn't running.'
easy	Y	Phrases and sentences with two or three elements	*ein kleiner Kürbis/das Handtuch falten*	'a small pumpkin/to fold the towel'
easy	Y	Prepositional phrases with the prepositions *in/auf* 'in/on'	*Die Schrauben sind in der Kiste./Alle Werkzeuge liegen auf dem Schrank*.	'The screws are in the box./All tools are on the cupboard.'
low	N	Temporal subordinate clauses with the conjunction *während/nachdem* 'while/after'	*Während Ihr Kind spielt, ist es zufrieden./Nachdem Sie alles markiert haben, sägen Sie die Form aus*.	'While your child is playing (s)he is happy./After you have marked everything, saw out the form.'
low	N	Subject relative clauses	*Mitarbeiter, die zufrieden sind, kommen gern zur Arbeit*.	'Employees, who are happy, enjoy coming to work.'
low	Y	Word forms in plural (they are hard to avoid)	no example provided	cf. *die Schrauben* 'the screws' mentioned above
low	Y	Spatial prepositions: *unter/über* 'under/above'	*den Brief unter das Buch legen/Das Handtuch hängt über dem Abfall-Eimer*.	'place the letter under the book/The towel hangs above the trash-can.'
medium	N	Passive forms with *werden* 'to be'—even reversible ones	*Der Antrag wird von Ihrem Chef gestellt./Dirk wird von Kristin gesucht*.	'The application is filed by your boss./Dirk is searched by Kristin.'
medium	Y	Present perfect tense	*Die Angestellten haben den Bus verpasst*.	'The employees have missed the bus.'
medium	N	Negations using *weder–noch* 'neither–nor'	*Weder der Tisch noch der Stuhl waren schön./Das Paket ist weder groß noch schwer*.	'Neither the table nor the chair were beautiful./The package is neither big nor heavy.'
medium	Y	Double object constructions	*Die Küchenhilfe zeigt dem Kollegen den Speiseplan./Das Amt schickt Ihnen einen Brief*.	'The kitchen aid shows the colleague the menu./The government agency sends you a letter.'
medium	N	Coordination with 'and' and elliptical structures	*Der Schuh ist auf dem Stift und ist blau*.	'The shoe is on the pencil and is blue.'
high	N	Topicalization	*Tische und Stühle müssen Sie zuerst wegräumen. Bei Verbrennungen helfen diese Maßnahmen*.	'Tables and chairs, you have to remove first./In case of burns, these measures help.'
high	N	*dass* 'that' clauses with varying reference of the accusative object	*Der Junge sieht, dass die Frau sich/sie sieht*.	'The boy notices that the woman sees herself/(her/them).'
extreme	N	Non-subject relative clauses	*Das Buch, auf dem der Stift ist, ist rot*.	'The book, on which the pencil is, is red.'

TABLE 2 | Overview of the three treebanks.

	LST	VERBMOBIL	TüBa-D/Z
Number of corpus graphs/syntactic structures	29,170	38,328	104,785
Number of tokens/leaves of structures	255,714	360,084	1,959,038
Average number of tokens per corpus graph	8.8	9.4	18.7
Number of inner nodes	457,324	496,466	2,402,421
Number of edges	683,868	818,222	4,353,888

appointment scheduling) were transliterated and syntactically annotated. The Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z; see, e.g., Telljohann et al., 2009) is a syntactically annotated corpus based on the German newspaper *die Tageszeitung* (*taz*).

Table 2 shows the overall sizes of the three investigated corpora. In all three corpora, corpus graphs (i.e., depictions of the syntactic structures) do not necessarily encode complete sentences in the linguistic sense—they also include, for example, headlines, terms in brackets, incomplete turns, and self-repairs. Tokens (i.e., the leaves of corpus graphs) cover not only word forms but also punctuation. As expected, in LST, the average number of tokens per corpus graph (roughly speaking, the sentence length) is shorter than in spoken utterances, although not to a great degree. This surprising circumstance results from long item lists (cf. example (8) below) that do not occur in spoken utterances, and the fact that we added missing punctuation symbols to improve the automatic syntactic analysis (*parsing*).

In the next section, we introduce the Leichte Sprache treebank LST used in our study.

2.3.1 Building the Leichte Sprache Treebank LST, a Syntactically Annotated Leichte Sprache Corpus

From a variety of sources freely available on the internet, spanning the years 2018–2021, we assembled a corpus of 245 LS texts with more than 300,000 word forms. To build a representative data set of LS texts of sufficient variety, we selected a broad spectrum of institutions, authors, and validators: according to the credits, at least 153 authors, 116

validators, and 53 institutions were involved in the creation of these texts. Each corpus graph in LST provides a feature with detailed source information of the original text. (For this purpose, we adopt the practice followed in historical corpora; see, e.g., the treebank of Old High German *Tatian*; Petrova et al., 2009.) We had originally planned to use this information to distinguish the following two subcorpora:

- LST-WithP, comprising only those texts that were proofread by LS readers, and
- LST-NoP, including texts without explicitly mentioned LS-reader participation.

To our surprise, the two subcorpora do not differ with respect to the number of violations of LS constraints; we therefore omitted the planned step of investigating differences between their construction frequencies. Nevertheless, we identified that a majority of the texts follow the LS rules. These texts not only deal with simple topics (e.g., fairytales) but also concern many spheres of life, including patient decrees, voters' rights, financial matters, and laws of succession. The implication is that the conformity of a text to LS rules does not depend on the complexity (or simplicity) of the topic but on whether or not its authors are aware of *best practices* (e.g., rephrasing *if* sentences as questions). We plan to investigate this observation in more detail in the future.

In the following, we sketch the process of obtaining the syntactic structures by parsing.

Preprocessing was done to improve the parsing results. From all PDF files, the plain text was extracted. In the extracted text, meta-text (e.g., running titles, page numbers, tables of contents, address lists, or links) was removed. Moreover, mediopoints (a specific functional LS symbol to segment compound nouns) and dashes without capitalized trailing word forms were removed to make full use of the compound analysis during parsing. In a series of pretests, we noticed that line breaks cause underspecification (cf. subscript "Dat./–" in (2-a), i.e., dative-case assignment in the sentential context vs. morphological underspecification if parsed in isolation). More seriously, it is not unusual that sentence fragments in separate lines are parsed incorrectly (cf. (2-b) where the isolated second line denotes a finite main declarative clause due to matching subject–verb agreement). We therefore removed line breaks (represented by the symbol "//" in the following) and colons within clauses (cf. (2-c)) in order to obtain correct case or grammatical function assignments.

(2) a. Ein Impfstoff schützt vor einer // Krankheit.
A vaccine protects against an illness$_{DAT/-}$

b. Sie dürfen zum Beispiel nicht in den Urlaub fahren // oder Freunde
You may for example not into the holiday drive or friends$_{ACC/NOM}$
besuchen.
visit$_{INF/FIN}$
'For example, you may not go on vacation or visit friends.'

c. Die Person heißt: // der Bevollmächtigte.
The person is_called the authorized_representative$_{PRED_{NOM}/-}$

Although subordinate clauses are not included in LS, they occur in the corpus material, with or without the correct German punctuation but usually in a separate line. We reconstructed the overall sentence, including punctuation, according to conjunction and the verb position. In German declarative clauses with at least three constituents, main and subordinate clauses differ in word ordering. The finite verb form fills the second constituent position in main clauses (V2), whereas it is final in subordinate clauses (VF). For example, relative clauses, especially those where the relativizer differs from *der/die/das*$_{inflected}$ 'who/which,' are not recognized as such when parsed in isolation (PRELS refers to a relative pronoun; PWS to a substituting interrogative pronoun):

(3) Dann berichtet sie in ihrer Fernsehsendung genau das, // (was$_{PRELS/PWS}$ die Menschen
Then reports she in her tv_show exactly that what the people
erzählt haben)$_{relative-clause/root-clause}$.
told have
'Then she reports in her tv show exactly what the people have said.'

However, we did not change the typical LS construction where a conjunction/causative adverb stands in a separate line—possibly followed by a colon—when the next sentence has V2 word order, as in example (4):

(4) Das ist nicht gut. // Denn: // Die Bauern müssen$_{VMFIN}$ deshalb sehr viele Tiere
This is not good because the farmers have_to therefore very many animals
halten$_{VVINF}$.
keep
'This is not good because the farmers have therefore to keep a great many animals.'

In example (5), the word order of the second line would be parsed in isolation as a yes/no question:

(5) Damit Sie das Heft gut lesen können: // Haben wir immer die männliche Form geschrieben.
So_that you the leaflet good read can have we always the male form written
'So that you can read the leaflet easily we use the male gender form exclusively.'

In example (6), an obligatory/complement clause, unrelated to the main clause, would remain:

(6) Familien-Hebammen achten darauf: // Dass es Ihrem Kind immer gut geht. // Und dass es
Family-midwives watch_out for_it that it your baby always well goes and that it
Ihnen immer gut geht.
you always well goes
'Family midwives make sure that your baby is always fine and that you are always fine.'

Lists are very frequent constructions in LS texts. For correct parsing, it is necessary to revise the punctuation throughout the sentence. The list in (8) cuts into the main clause, although the clause in (7) ends with a colon. The list starts with a prepositional phrase modifying the direct object of the main clause. Then, it continues with a long list of subordinate clauses. The reconstructed complex sentence consists of more than 75 words. This example explains the unexpectedly high average length of the corpus graphs in LST.

(7) Das Robert-Koch-Institut sammelt Informationen:
The Robert Koch Institute collects information

(8) a. - über das Corona-Virus
about the Corona virus

b. - Wie viele Menschen krank sind
how many people ill are
'how many people are ill'

c. - Wo Menschen krank sind
where people ill are
'where people are ill'

d. - Wie viele Menschen wieder gesund sind
how many people again healthy are
'how many people have recovered'

e. - Ob es besser oder schlechter wird
whether it better or worse gets
'whether it gets better or worse'

...

In a further change, we systematically added question marks to make questions more easily recognizable in LST. Although they are included in LS, a large majority of questions in the corpus are printed with a full stop or without any punctuation.

Quality of the syntactic structures. A small LS sample text was evaluated with different parsers (e.g., from the wide spectrum of deep-learning approaches; see Linzen and Baroni (2021) for a recent survey). We finally decided to employ PARZU (Sennrich et al., 2013). PARZU is a dependency parser for German with a rich part-of-speech (POS) and morphological tag inventory (see Tritscher (2016) for an evaluation with German prose text). In a random sample of 100 sentences from the overall LS corpus, we identified four, partially very minor, errors (provided in the **Supplementary Material**). Consequently, the overall quality of LST is shown to be less perfect than the manually inspected standard German treebanks; however, given the overall treebank size of LST, the accuracy is deemed sufficient for the identification of clear trends. Thus, using the power of a treebank search that exceeds the scope of a manual inspection of a small sample or a pure word-form-based text search, LST gives rise to valuable insights.

In our study, the dependency trees produced by PARZU were transformed into the TIGER-XML format (König and Lezius, 2003), in which VERBMOBIL and TüBa-D/Z are also available. All three treebanks were inspected with TIGERSearch[2] (König and Lezius, 2003). Note that VERBMOBIL is not morphologically annotated. Therefore, some queries cannot be answered in this corpus (they will be referred to as "n.a.", i.e., not applicable, in the tables).

2.3.2 Typical Syntactic Constructions in LST Compared to VERBMOBIL and T Ba-D/Z

In the following section, we study the frequencies of a wide range of syntactic constructions in LST. Constructions that fulfill one of the following conditions are added to ELS:

- Constructions that have at the most medium-level difficulty for LS readers (according to the LeiSA study) and that occur frequently in LST; and
- Constructions not covered by the LeiSA study, whose LST frequencies compare favorably to their frequencies in spoken German, but which cannot be easily transformed into pure LS constructions.

We present the syntactic phenomena, ordered according to their level of construction complexity: (1) word-related, (2) phrase-related, and (3) clause type-related constructions. Within each level, we first refer to the phenomena mentioned in **Table 1**; then, we discuss typical simple constructions that are beyond the scope of pure LS and were not evaluated in the LeiSA study. The systematic nature of this search demonstrates that we assessed the whole range of simple syntactic constructions. For reasons of space, we omit many details here, especially when no constructions are added to ELS.

Negations. **Table 3** provides the frequencies of several negation words. The absolute numbers are provided in brackets. (This format is preserved in **Tables 4**–**7**.) Negation is forbidden in LS; however, it is difficult to avoid completely (Bock, 2017). As *nicht* 'not' is easy for LS readers to understand (according to the LeiSA study), we opt to add it to ELS. According to the frequency of its occurrence in LST (comparable to that in VERBMOBIL), $kein_{inflected}$ should also be added to prevent forcing a reformulation with *nicht*. All other constructions including negation are very infrequent in both LST and VERBMOBIL.

Prepositions. The POS of a preposition is distinguished in pre-positioned prepositions (APPRs) (POS starting with APPR = APPR.*), post-positioned prepositions (APPOs), and partly fronted and partly trailing prepositions (circum-positioned POS = APZR [according to the German term *Zirkumposition*]). APPRART specifies prepositions with an agglutinated definite article (APPR.* refers to APPR+APPRART).

APPRART should not occur in LS; nevertheless, this construction is frequently used. Half of the cases occur in the idiomatic prepositional phrase (PP) *zum Beispiel* 'for example,' which should not be abbreviated in LS. In practice, the use of APPRART makes sentences shorter. Moreover, using *zu dem Beispiel* for *z.B.* 'e.g.' sounds odd in German. We therefore add APPRART to ELS. **Table 4** provides frequencies for the explicitly mentioned prepositions in **Table 1**. Note that the treebanks do not distinguish spatial/temporal use. Thus, the frequencies presented here for the prepositions *unter/über* are overall figures.

The two partially or completely post-positioned preposition types, APPO and APRZ, occur in LST. All 47 cases of APPOs specify a temporal duration with *lang* 'long', as in *10 Tage* $lang_{APPO}$ 'for 10 days'. There are 16 APZRs occurring (e.g., *von zuhause* aus_{APZR} 'from home' and *von Anfang* an_{APZR} 'since the beginning'). Under LS rules, all but genitive-taking prepositions (absent in LST) are included; therefore, no further extension is suggested for prepositions.

Nouns and pronouns. In noun phrases (NPs), the following POS tags occur as heads: NE (proper noun); NN (lexical noun); and PRO (pronoun), which summarizes PDS (substituting demonstrative pronoun), PIS (substituting indefinite pronoun), PPOSS (substituting possessive pronoun), PRELS (relative pronoun), PRF (reflexive personal pronoun), PPER (irreflexive personal pronoun), and PWS (substituting interrogative pronoun). In **Table 5**, N.*+PRO refers to any NP head filler, where POS = N.* refers to NE+NN (suppressed in the table because it can be reconstructed by subtracting the frequency of PRO from N.*+PRO; e.g., in LST, 75.51 = 1–25.49%, and 58,832 = 78,961–20,129 cases). We first investigate the morphological features *number* = *plural* and *case* = *genitive*, which were studied in the LeiSA project, in LST, and TüBa-D/Z. Then, we present the frequencies of the individual PRO types in all three treebanks. The frequencies of lexical nouns are given in the bottom row. The LeiSA study excluded pronouns and proper nouns from its investigations.

Plural forms are of low difficulty. They occur slightly more frequently in LS than in TüBa-D/Z (no data are available for VERBMOBIL). This construction is included in LS. The frequency of genitives, forbidden in LS, is very low; therefore, we do not add genitives to ELS, although we note that such

[2]To replicate the data, a TÜNDRA-based search (see https://weblicht.sfs.uni-tuebingen.de/Tundra/) for all three specified corpora in the dependency-tree format offers another option. The query format in TÜNDRA is based on TIGERSearch. All queries used in the following are provided in the **Supplementary Material**.

TABLE 3 | Frequencies of negations in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Freq. of *nicht* 'not' in **a**ll **t**oken**s** (ats)	6.31 %	(1,840)	6.29 %	(2,411)	12.84 %	(13,457)
Freq. of *nichts* 'nothing' in ats	0.27 %	(78)	0.55 %	(209)	1.14 %	(1,190)
Freq. of *nie*(*mals*) 'never' in ats	0.17 %	(50)	0.08 %	(32)	0.64 %	(674)
Freq. of $niemand_{inflected}$ 'nobody' in ats	0.39 %	(114)	0.01 %	(3)	0.40 %	(417)
Freq. of $kein_{inflected}$ 'no' in ats	2.19 %	(640)	1.59 %	(611)	3.80 %	(3,982)
Freq. of *weder – noch* 'neither-nor' in ats	0.00 %	(0)	0.01 %	(4)	0.28 %	(289)

TABLE 4 | Frequencies of APPR.* in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Freq. of APPR.* in all tokens	8.09 %	(20,690)	6.55 %	(23,588)	8.86 %	(173,574)
Freq. of APPRART in APPR.*	1.72 %	(4,401)	1.79 %	(6,446)	1.53 %	(29,944)
Freq. of *in/auf* in APPR.*	2.49 %	(6,359)	1.60 %	(5,755)	2.78 %	(54,464)
Freq. of *unter/über* in APPR.*	0.02 %	(54)	0.01 %	(22)	0.34 %	(6,752)

TABLE 5 | Frequencies of (pro)nouns in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Total number of N.*+PRO (n+p)		78,961		94,725		579,511
Freq. of N.*+PRO in plural in n+p	27.71 %	(21,883)		n.a	24.17 %	(140,089)
Freq. of genive case-n+ps in n+p	0.68 %	(533)		n.a	9.14 %	(52,953)
Freq. of PRO in n+p	25.49 %	(20,129)	50.79 %	(48,108)	17.93 %	(103,897)
Freq. of PDS in PRO	10.08 %	(2,030)	18.59 %	(8,944)	8.13 %	(8,445)
Freq. of PIS in PRO	15.25 %	(3,070)	5.79 %	(2,786)	15.01 %	(15,598)
Freq. of PPOSS in PRO	0.00 %	(0)	0.04 %	(18)	0.03 %	(35)
Freq. of PRELS in PRO	3.64 %	(733)	0.75 %	(361)	14.78 %	(15,359)
Freq. of PRF in PRO	7.94 %	(1,599)	5.75 %	(2,764)	14.15 %	(14,701)
Freq. of PPER in PRO	59.47 %	(11,971)	67.03 %	(32,247)	44.94 %	(46,695)
Freq. of PWS in PRO	3.61 %	(726)	2.05 %	(988)	2.95 %	(3,064)
Freq. of NN in NN+NE	91.85 %	(54,036)	87.11 %	(40,610)	78.41 %	(372,949)

TABLE 6 | Frequencies of verb forms in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Total number of verb forms (V.*)		41,996		50,679		243,705
Freq. of VA.* in V.*	22.92 %	(9,625)	35.78 %	(18,134)	28.60 %	(69,706)
Freq. of VM.* in V.*	19.33 %	(8,117)	13.68 %	(6,934)	22.60 %	(19,983)
Freq. of V.FIN or V.IMP in V.*	71.88 %	(30,186)	77.26 %	(39,154)	66.90 %	(163,033)
Freq. of V.PP in V.*	6.28 %	(2,639)	2.76 %	(1,400)	16.35 %	(39,852)
Freq. of V.INF/V.IZU in V.*	21.84 %	(9,171)	19.98 %	(10,135)	16.75 %	(40,820)
Freq. of preterite tense in V.*	2.16 %	(907)		n.a	20.13 %	(49,049)
Freq. of passives in V.*	1.51 %	(633)	0.11 %	(53)	5.43 %	(13,238)
Freq. of subjunctive mood in V.*	0.73 %	(308)		n.a	6.06 %	(14,744)
Freq. of double object verbs in V.*	1.08 %	(454)	2.09 %	(1,057)	1.74 %	(4,240)

constructions are often used in idiomatic expressions, e.g., for the names of institutions. A comparison of pronoun frequencies places LST somewhere between the written and spoken corpora. However, we did not expect high numbers of pronouns, for the following reasons: The three major LS rule sets Netzwerk Leichte Sprache, 2013, Inclusion Europe, 2009, and BITV2.0 (2011) insist on consistent naming, i.e., using exactly the same word for the same thing/person throughout a text; additionally, the LS rule set of Inclusion Europe forbids pronominal resumption, favoring nominal resumption. In

TABLE 7 | Sentence complexity in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Total of sent.-onset frontfieds (sofs)		29,170		11,600		72,586
Freq. of NP-SB in sofs	47.48 %	(13,851)	48.85 %	(5,667)	50.48 %	(36,645)
Freq. of short NP-SB in SB sofs	91.31 %	(12,648)	98.61 %	(5,588)	34.62 %	(25,128)
Total of relative clauses (rels)		657		427		17,017
Freq. of short rels in rels	86.61 %	(569)	90.16 %	(385)	56.47 %	(9,609)
Freq. of SB-rels in rels	64.08 %	(421)	n.a		64.51 %	(10,978)
Freq. of short SB-rels in SB rels	86.94 %	(366)	n.a		57.03 %	(6,261)
Total of subordinate clauses (subs)		2,811		3,294		23,302
Freq. of short subs in subs	76.91 %	(2,161)	80.90 %	(2,665)	47.94 %	(11,171)
Freq. of '**w**hile/after' (wcls) in subs	0.18 %	(5)	0.18 %	(6)	9.99 %	(2,327)
Freq. of short wcls in wcls	40.00 %	(2)	83.33 %	(5)	35.45 %	(825)
Freq. of '**if**'cls. (icls) in wcls	53.90 %	(1,515)	37.74 %	(1,243)	15.52 %	(3,616)
Freq. of short icls in icls	79.87 %	(1,210)	88.74 %	(1,103)	61.37 %	(2,219)
Freq. of '**t**hat' cls. (tcls) in wcls	16.22 %	(456)	26.29 %	(866)	34.46 %	(8,031)
Freq. of short tcls in tcls	72.81 %	(332)	71.25 %	(617)	37.42 %	(3,005)

particular, many occurrences of PIS that characterize abstract referents (e.g., *man* 'one', $jemand_{inflected}$ 'somebody', *etwas* 'something', and $alle_{inflected}$ 'all') resemble those in written text. Due to their frequency, to shorten the resulting sentences, and to enable the use of abstract referents to circumvent passive constructions, we include all pronouns in ELS. The frequency of proper nouns in LST is similar to that in VERBMOBIL. Because this construction is already included in LS, no extension is required.

Verb forms. Before examining phrases and clauses, we study verb forms (see **Table 6**). Part-of-speech tags starting with "VA" (POS = VA.*) refer to auxiliary verbs, including the copula use of 'to be.' The prefix "VM" (POS = VM.*) characterizes modal verbs, and "VV" (POS = VV.*) lexical verbs. The total VV.* frequencies are omitted in favor of reconstructing VV.* = V.* – VA.* – VM.* for each corpus. The frequencies, resembling those of TüBa-D/Z, qualify LST as a variety of written text. Not distinguishing between the verb types (V. = VA + VM + VV), the proportions of finite verb forms (V.FIN/V.IMP, i.e., any POS type ending with a finite [FIN] or imperative [IMP]) in declarative, interrogative, and imperative clauses give a rough estimation of clause simplicity. A clause with a finite lexical verb form that is not one of the few non-finite verb complement-taking verbs, like *to try/hate/forbid to do (something)*, cannot contain other verb forms. (Present participles are encoded as adjectives in all German corpora.) Auxiliaries and modals, which are included in LS, can dominate non-finite verb forms to build the present perfect tense and specify the modality of other verbs, respectively. Contrary to expectation (that finite verb forms would be most frequent in LST), the frequency of finite verb forms in LST is between the frequencies in VERBMOBIL and TüBa-D/Z. This finding can be attributed to the presence of a higher number of present perfect tense constructions in LST than in the non-LS written corpus (cf. the V.PP frequencies). The frequency of infinitives (V.INF and V.IZU, i.e., infinite with *zu* 'to', as in *Man hat so versucht*$_{FIN}$ *Corona aufzuhalten*$_{IZU}$ "One has tried to stop Corona in this manner") is similar in LST and VERBMOBIL. However, the number of modals that are likely to entail a lexical verb is higher in LST (as a variety of written text) than in VERBMOBIL. As mentioned above, specific lexical verbs can dominate non-finite verbs with POS = V.IZU. The verb lemma *versuchen* 'to try' occurs 16 times in LST, 26 in VERBMOBIL, and 472 in TüBa-D/Z. The frequency in LST resembles that in VERBMOBIL, and the infinitive construction is similar to that of modals. We therefore add complement-taking verbs that belong to the restricted LS vocabulary to ELS. However, constructions with *um zu* 'for the purpose of/in order that' (KOUI), which occur 46 times in LST, 115 times in VERBMOBIL, and 2,426 times in TüBa-D/Z), are not added to ELS. The frequency of KOUI is 50% lower in LST than in VERBMOBIL. Moreover, the construction can straightforwardly be segmented into: 'for the (following) purpose/thereby/:/. . .//' and a main-clause construction, without obstructing the train of thought. For example, the sentence *Es braucht*$_{FIN}$ *Zeit um*$_{KOUI}$ *sich zu erholen*$_{INF}$ 'It takes time to recover' is divided into the following three lines: *Es braucht*$_{FIN}$ *Zeit. // Damit/Bis(:) // Sie erholen*$_{FIN}$ *sich (wieder).*

The preterite tense, passive voice, and subjunctive mood are forbidden in LS. The preterite occurs very infrequently in LST, and most of the 800 cases pertain to auxiliaries and modals. The few lexical verb cases can be replaced by present perfect tense forms (included in LS) without the meaning being changed. To support this argument, we searched VERBMOBIL for preterite forms of the three most frequent lexical verb lemmas according to Kempen and Harbusch (2019). For *sehen* 'to *see*' and *machen* 'to make,' no incidences were found. Forms matching *ging/-st/-t/-et/-en* of *gehen* 'to go' occur 31 times; however, only half of the matches are related to preterite forms. These all occur in the idiomatic phrase *das ging schnell/gut* 'that went quickly/well'. The other cases match subjunctive mood forms referring to potential time slots/connections/etc. As all verb forms of auxiliaries and modals appear with high frequency, and sentences are shortened by the use of the preterite (instead of present

perfect) tense, we include the preterite for VA and VM lemmas in ELS. The frequency of passive constructions is low in all three corpora. Given that such constructions are of medium difficulty according to the LeiSA study and that it is often hard to find a simple reformulation in the active voice that conveys the same nuance of meaning, we include passive constructions in ELS.

We noticed that nearly all subjunctive mood cases in LST are forms of auxiliaries or modals (e.g., *wären* 'would be,' and *möchten* 'would like'). These are frequent word forms in German: there are 7,836 occurrences in VERBMOBIL for the rough search pattern POS = VM.* or VA.* with the word form matching the prefix = *möcht.*/könnt.*/würd.*/wär.*/hätt.**. We therefore include the subjunctive forms of VA and VM in ELS. Constructions with double objects are of medium difficulty, according to the LeiSA study. Such constructions are equally rare in all three corpora examined in this study. However, common verbs like *geben* 'to give (somebody something)' qualify for this construction. As these verbs are included in LS, no extension is required.

Phrase and sentence complexity. A treebank search allows for very detailed syntactic specifications. However, it is necessary to keep in mind that the three inspected treebanks are differently encoded in this respect. For example, a noun phrase (NP), i.e., cat = NX in VERBMOBIL and TüBa-D/Z, covers constituents that differ from the nodes at the ends of edges labeled subj, objа, pn, etc., in PARZU. In an NX, adverbs can be seized. In PARZU, adverbs—if not in the frontfield—belong to the sentential level. For example, the phrase *auch schon viele barriere-freie Gebäude* 'also already many barrier-free buildings', occurring in the midfield of a clause, is assigned to three constituents on the sentence level in PARZU, whereas in the two other treebanks, the phrase is assigned to one NX node. There is no simple solution to this problem without manually inspecting all adverbs in LST. Hence, not entirely accurately, but in line with the characterization in the LeiSA study, we define phrase complexity by a simplistic dichotomy with respect to a length: We distinguish *short* (up to three words) from *long* (more than four words) constituents. This concept translates to sentence-complexity calculations: We define *short* sentences as containing up to nine words (assuming that these occur in no more than three constituents, each comprising no more than three words) and *long* sentences as containing more than nine words. (Note that punctuation is excluded from consideration here because any phrase-level search refers to an inner node of a corpus graph. In each of the three treebanks, punctuation is governed by the root node—cf. the example trees provided in the **Supplementary Material**. The discontinuous positions of punctuation symbols in the surface word order can be accessed by the list of word forms in the TIGER-XML format.) For example, *Bis zu diesem Gehalt zahlt man Beiträge* 'Up to this income, you pay contributions' is a short sentence, but *Die Krankenkasse oder die Agentur für Arbeit zahlt die Beträge für sie.* 'The health insurance or the employment agency pays the contributions for you.' is a long one. On average, this simple distinction identifies complicated constructions: in our examples, a sentential modifier PP and an NP-modifying PP in an NP coordination, respectively.

In LS, no explicit length restriction for phrase complexity is stated. However, the LeiSA study qualifies as easy only those phrases with no more than three elements; therefore, no new constructions are added to ELS, and we omit detailed numbers here (see the **Supplementary Material** for the frequencies of the three argument NPs: subject [SB], indirect object [IO], and direct object [DO]—important elements for sentence understandability) and move on to investigations of sentence complexity.

We first quantify the frequency of the canonical SVO word order in main declarative clauses (see the upper panel of **Table 7**). We restrict the search to frontfields at the onset of a sentence to abstract away from elided constituents. (*Forward Conjunction Reduction* and/or *Gapping* are the only ellipsis phenomena that can elide the left periphery and only in the second conjunct of a coordinated sentence; see Ross, 1967). Moreover, according to Temperley (2019), the most complex constructions occur at the onset of a sentence because more mental capacity is available here.

In line with LS, the LeiSA study allows only SVO word order; even mild forms of topicalization were judged to be very difficult for LS readers. Unexpectedly, the frequency of the canonical word order is found to be very similar in all three treebanks; constructions with the SVO word order comprise only half of all constructions in the LS corpus, i.e., the other half are very complicated for the target readers. Clearly, the standard German writers of LS texts adhere to the standard rules of German discourse structure. In mild cases, a one-word constituent occupies the frontfield (cf. example (3) above). However, we also found complex frontfield fillers, such as conditional clauses, sentential subjects, and objects. For example, in a sentence (9), the fronted object is interpreted as the argument of the finite verb form *wollen* until *barriere-frei* occurs (i.e., this is a garden-path sentence). Given the difficulties arising from simple deviations from the canonical word order, we do not report the frequencies of individual constructions here, and no deviations are included in ELS.

(9) Das$_{ACC}$ wollen$_{FIN}$ wir in NRW barriere-frei machen$_{INF}$.
This want we in NRW barrier-free make
'These things we want to make barrier-free in NRW.'

For the canonical word ordering, the average length of the frontfield is longer in LST and TüBa-D/Z, the two written varieties—as expected. The subjects in VERBMOBIL are extremely short due to the use of personal pronouns in dialogue. (Note that VERBMOBIL often has discourse markers, self-repairs, etc., at the onset of a sentence. Therefore, the total numbers for LST and VERBMOBIL diverge more here than in other tables.)

For subordinate clauses, we distinguish between relative clauses and subordinate clauses starting with a subordinating conjunction (KOUS) in the lower panels of **Table 7**. Both constructions are forbidden in LS. According to the LeiSA study, subject-relative clauses are of low difficulty for LS readers, whereas any other type of relative clause (i.e., a relativizer in the dative or accusative case due to its grammatical function or dominating preposition, respectively) is extremely difficult. Surprisingly, both types occur with approximately equal frequency in LST and TüBa-D/Z. As

expected, the relative clauses in LST are considerably shorter than those in TüBa-D/Z. Given their frequency and the LeiSA evaluation, we opt to include subject-relative clauses in ELS (although this adds a VF construction to the included word order patterns); however, we suggest that such clauses are not discontinuous and that they should be short.

All types of subordinate clauses are forbidden in LS. In the LeiSA study, subordination with the temporal conjunctions 'while/after' is considered of low difficulty. Although 'while' does not occur in LST, *nachdem* 'after' is used five times. The conditional conjunction *wenn/falls* 'if' is by far the most frequent (see **Table 7**). Other conjunctions used in LST (with their respective frequencies) include: *als* 'when' (8); *bevor* 'before' (18); *bis* 'until' (37); *damit* 'so that' (5); *indem* 'by' (2); *nachdem* 'after' (5); *ob/obwohl* 'whether' (204); *seit* 'since' (1); *solange* 'as long as' (3); and *weil* 'because' (180).

Rather than including specific subordinate clause types, we suggest adding all subordinating conjunctions to ELS. However, the conjunction and the sentence should be presented in two separate consecutive lines, and the trailing sentence should have main clause word order (cf. example (4) for the paratactic conjunction *denn*, which always entails main clause word order). The same construction works with the synonymous subordinating conjunction *weil*. In VERBMOBIL, the subordinating conjunction *weil* 'because' occurs in half of the cases with SVO order (Kempen and Harbusch, 2016). (This phenomenon is widely studied as the *weil*-V2 phenomenon in spoken standard German; see Reis (2013) for a broad overview.) This strategy also covers the highly difficult construction of dependent *that* clauses, which occur in LST slightly less often than in VERBMOBIL. This construction can straightforwardly be avoided by replacing *that* by a colon. The content of the *that* clause is presented as a main clause in the canonical SVO order. Thus, we do not add this construction to ELS.

Coordination and ellipsis. Coordination and ellipsis are of medium difficulty, according to the LeiSA study. However, the tested examples are very simple. Coordinations in LST consist of very long lists, as illustrated above in example (8). Often, formal definitions are replaced by long lists of examples, probably to avoid the use of overcomplicated technical terms. Therefore, although we do not add any new constructions, we recommend using only short lists of coordinated constituents.

The same holds for ellipsis. The use of ellipsis in spoken and written text (see corpus studies into VERBMOBIL (Harbusch and Kempen, 2009) and the TIGER treebank, another syntactically annotated German newspaper corpus (Harbusch and Kempen, 2007)) goes beyond the scope of very limited Forward Conjunction Reduction restricted to the subject, which prevails in LST, and which was the only type of ellipsis evaluated in the LeiSA study. As this construction is judged to be of medium difficulty, and it can be circumvented by explicitly repeating or pronominalizing the subject, we choose not to add it to ELS.

2.3.3 Summary of the Constructions in Extended Leichte Sprache

Table 8 sums up all extensions included in ELS that we proposed in the previous section. Now that the range of syntactic constructions of ELS has been defined, we detail how our system supports the writing of ELS text.

TABLE 8 | Summary of constructions in Extended Leichte Sprache that extend Leichte Sprache.

Phenomenon	Usage
Negation	Restricted to *nicht* as verb modifier and $kein_{inflected}$ as determiner or substituting pronoun
POS = APPRART	Agglutination of the definite article following the preposition
POS = PRO	NP filled with a substituting pronoun
Preterite tense	Restricted to auxiliaries and modals
Subjunctive mood	Restricted to auxiliaries and modals
Passive voice	Restricted to forms of *werden*
Complement-taking verb	Restricted to verb lemmas in the vocabulary of LS that dominate a nonfinite verb form with the word ordering SV_{finite} $OV_{nonfinite}$
Subordination	*dass* 'that' is replaced by a colon; any other conjunction obtains a separate line; the constituents of the subordinate clause are presented with V2-word order in a new line
Relative clause	Restricted to subject-relative clauses, however, not discontinuous

3 EXTENDEDEASYTALK

First, we present the state of the art in writing support tools. In **Section 3.2**, we give an intuitive impression of how *ExtendedEasyTalk* works. **Sections 3.3** and **3.4** go into the details of the computational linguistic mechanisms used to support the writing of a sentence and the production of a sentence-coherence element, respectively. In **Section 3.5**, we sketch the active mode of text-production teaching. Finally, we present the results of our evaluation with different user groups.

3.1 The State of the Art in Writing Support Tools

First, we describe the state of the art in technical writing support in the overall area of AAC. Then, we focus on NLG-based approaches. This latter section provides the blueprint for a knowledge-based automatic natural-language generator, enabling us to refer back to concepts used in our system.

3.1.1 Augmentative and Alternative Communication

For people with congenital or acquired communication impairments, the use of *Augmentative and Alternative Communication*[3] (*AAC*) is often an essential means of inclusion, i.e., for self-determined participation and self-expression. AAC offers a wide range of communication techniques, including gestures,

[3]https://isaac-online.org (information provided by the *International Society for Augmentative and Alternative Communication*).

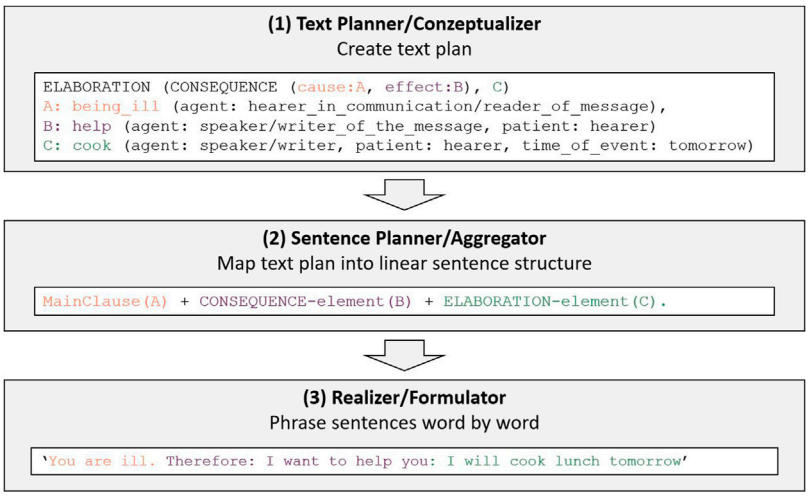

FIGURE 1 | Basic elements of a natural-language generator (bold), illustrated with our example (in text boxes).

signs, and graphic symbols, as well as technical communication aids. Technical AAC solutions range from simple concatenations of symbols for needs-based, functional communication using limited vocabulary to complex customizable systems (see Lancioni et al. (2019) for a detailed survey).

Here, we focus on AAC systems that aid users with at least low literacy skills both to express basic necessities and to write about topics that create social closeness, for example, to share personal information and experiences (see, e.g., Light et al., 2019). The rich morphology and the relatively complex word-order rules of German complicate the generation of useful and grammatically correct suggestions. Commercial systems that go beyond functional communication include MindExpress, Gateway, and Snap Core First[4]. These systems essentially concatenate words, word groups, and symbols into sentences, thereby providing basic linguistic support, such as adaptive word prediction and automatic inflection for simple constituents. Technical AAC solutions are currently evolving rapidly and are increasingly available on mainstream devices (e.g., smartphones and tablets; Light and McNaughton (2012)). All popular free apps for German allow users to access large customizable vocabularies of (visual) symbols. However, they do not provide well-founded linguistic support for sentence construction and/or text production (cf. LetMeTalk and SymboTalk[5]). Importantly, these systems are mainly intended for direct (face-to-face) communication.

3.1.2 Natural Language Generation

There is an increasing demand for language support through linguistic processing by computer. However, the currently available AAC systems do not exploit the full potential of computerized linguistic processing (Waller, 2019). The technical authoring support available for LS includes tools for automatic text simplification based on parsing (for German see, e.g., Suter et al., 2016) and text validation tools (see, e.g., LanguageTool[6], a system that flags violations of the LS rules).

To the best of our knowledge, there is no recent NLG-based text-production system customized to AAC-user needs (cf. the pioneering approach by Demasco and McCoy (1992); Gatt and Krahmer (2018), who illustrate the potential of NLG systems in general; and G2.com (2021), which subcategorizes systems as "Highest rated/Easiest to use/Free" and provides links to writing support based on NLG).

In the following, we outline the steps from a speaker's intention to a context-sensitive utterance to allow us to highlight the NLG concepts we employ in *ExtendedEasyTalk*. **Figure 1** shows the typical three-stage pipeline architecture of a declarative text generation system (see, e.g., Figure 3 on page 13 by Reiter and Dale, 2000), illustrated in terms of example (1-a) above. The overall input to the Text Planner/Conceptualizer (component 1) is encoded as the speaker's *goal*:

> INFORM(*S*, *H*, KNOW(*H*, cooking(agent_of_action_cooking: *S*, object: lunch, time: tomorrow))).

S refers to the speaker, and *H* to the hearer. The goal is an INFORM-speech act, i.e., *S* wants *H* to know something that *S* assumes *H* does not yet know. (In a REQUEST-speech act, *S* wants to obtain new information that *S* assumes *H* knows, often resulting in a question like "*Can you please tell me what you would like to eat for lunch tomorrow?*".) Roughly speaking, the goal highlights the discrepancy between the knowledge bases of *S* and

[4]www.jabbla.com (MindExpress); www.gatewaytolanguageandlearning.com (Gateway); www.tobiidynavox.com/pages/snap-core-first (Snap Core First).
[5]www.letmetalk.info (LetMeTalk); www.symbotalk.com (SymboTalk).
[6]www.languagetool.org/de/leichte-sprache (LanguageTool).

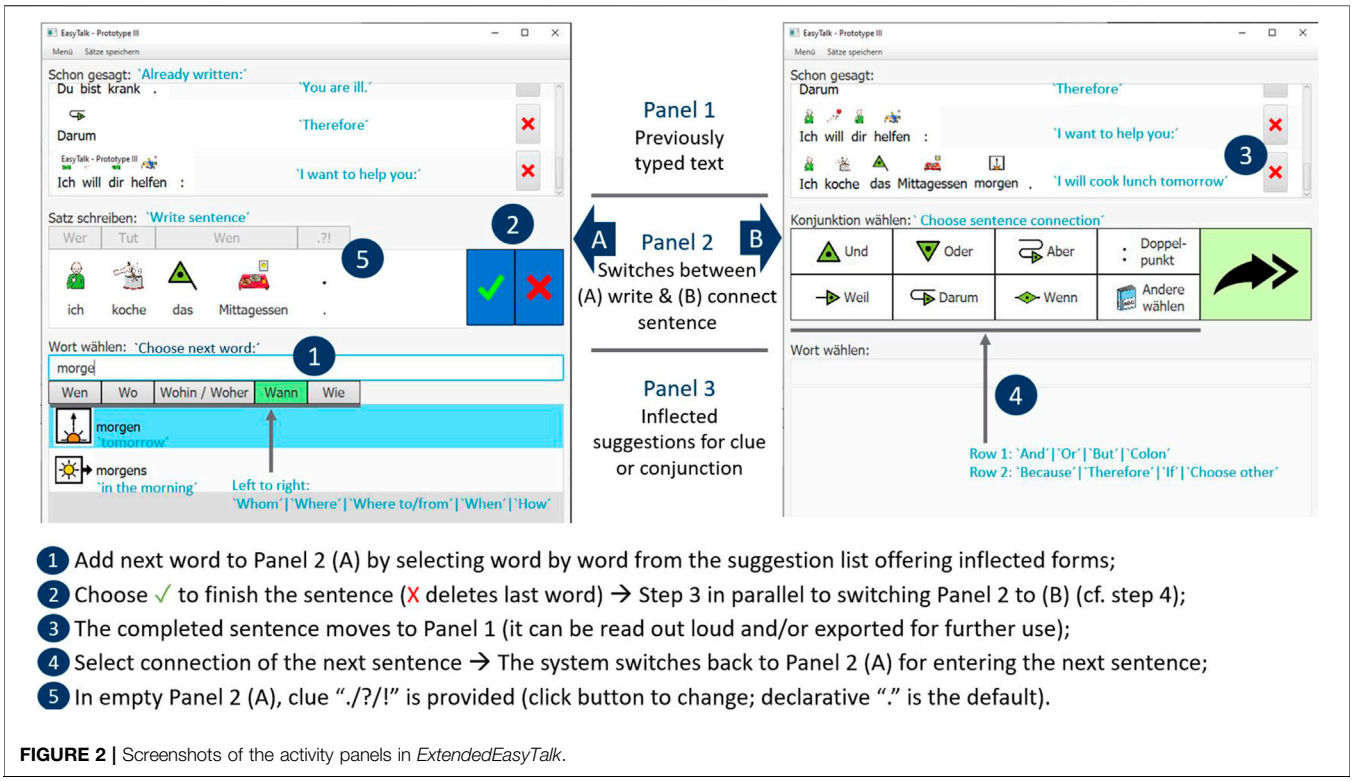

FIGURE 2 | Screenshots of the activity panels in *ExtendedEasyTalk*.

H, which should be removed through a series of speech acts, i.e., the not-yet-verbalized conceptual messages (propositions dealt with the Conceptualizer). Ideally, after the delivery of the message by *S*, *H* knows all communicated facts (and facts that can be inferred by *H*). This task requires separate representations of the speaker's and the hearer's knowledge about the current situation and about their presupposed *world knowledge*, respectively (including the implications of/inferences from all the facts).

In the example, we assume that *S* notices that *H* is ill. *S* wants to indicate to *H* that *S* is aware of this fact. By explicitly informing *H* that *S* knows that *H* is ill, the speaker enriches the utterance with a known fact to make the context/intention of the utterance clear, thereby creating an overall *discourse structure*. Because of the close personal relationship between *S* and *H*, *S* decides to help by preparing lunch for *H* the next day. This plan results from the world-knowledge fact that relieving an ill person of a task helps that person to rest and recover.

The Conceptualizer decides which information should be communicated. In our example, the facts *A*, *B*, and *C* are selected (the propositions are rendered here in the form of sentences, abstracting away from the logic representation, and detailed argumentation; instead, each proposition is supplemented by the intended interpretation of the hearer):

A: *Du bist krank* 'You are ill': mutual agreement about the context of the utterance
B: *Ich will Dir helfen* 'I want to help you': reason for a proposed action
C: *Ich koche morgen das Mittagessen* 'I will cook lunch tomorrow': communication of the planned action

Importantly, propositions do not stand in isolation, but in the relationship, in order to express the discourse structure/the speaker's intention. A widely used technique for this purpose is that of *Rhetorical Structure Theory* (RST; see Hovy, 1988 and Mann and Thompson, 1988). Two important examples of relations between propositions are ELABORATION and CONSEQUENCE. The resulting hierarchical structure of interrelated propositions is called the *text plan*. Text plans are handed over to the Sentence Planner/Aggregator module (component 2), which has the task of linearizing the hierarchical structure. The linearization process involves, among other things, the insertion of coordinating and subordinating conjunctions and other lexical items that instantiate RST relations (although not all RST relations need to surface explicitly in the final text). In terms of our example:

- *A* is realized as the main clause;
- CONSEQUENCE(*B*) is realized by the causative adverb *therefore* preceding the main-clause realization of *B*, and
- ELABORATION(*C*) is realized by a colon preceding the main-clause realization of *C*.

Component 3, the Realizer/Formulator, provides the subsequently generated text—one of the many realization options of the overall generation system.

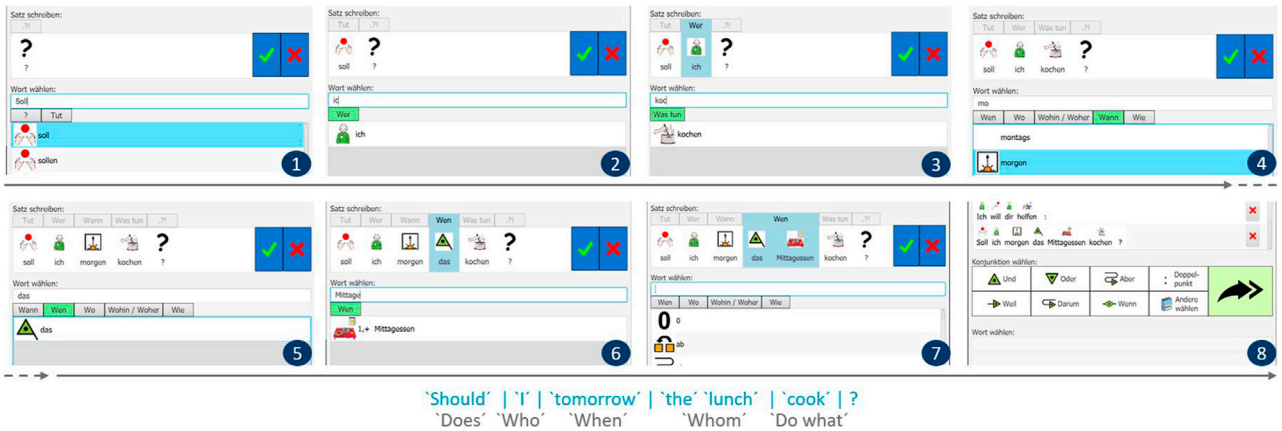

FIGURE 3 | Screenshots of the process for writing the question *Soll ich morgen das Mittagessen kochen?* 'Should I cook lunch tomorrow?' in eight stages (cf. the numbers in blue in the lower-right corners). For reasons of space, all screenshots but the last one (screenshot 8) depict only Panels 2 (A) and 3.

In the NLG nomenclature, all AAC systems mentioned in **Section 3.1.1** are restricted to formulator problems. In the following, we illustrate the potential of using concepts from all three generation steps in our system.

3.2 A Writing Session With *ExtendedEasyTalk*

Let us familiarize ourselves with the assisted text-production process of *ExtendedEasyTalk* through an outline of its five essential steps (cf. the numbers in blue circles in **Figure 2**).

The system permanently displays three panels. Panel 1, at the top, contains the previously written text. In the middle panel, Panel 2 (A) accumulates the word forms of a sentence chosen word by word from Panel 3, and Panel 2 (B) provides predefined sentence connectors. Panel 3 offers the list of suggestions for the currently typed string. The example depicts the fact that the user has already produced a sequence of sentences: *Du bist krank. // Deshalb // Ich will dir helfen:* 'You are ill. // Therefore // I want to help you:' in the left-hand side of Panel 1. The sentence under construction, *Ich koche Mittagessen.* 'I will cook the lunch', is displayed in Panel 2 (A). The user has two options for how to proceed: the sentence is either continued or finished.

In step 1, a new word is selected to be added to the sentence using Panel 3 (cf. lower-left corner). *ExtendedEasyTalk* offers inflected word forms according to the wh-cue header in green (here, **Wann** 'when' refers to the grammatical function provided in simplistic wording) matching any (possibly empty) input string the user types in Panel 3. In the figure, the user has typed *morge* 'tomorro'. As predictions, the system presents inflected completions from the lexicon, according to the grammatical function referred to by the active cue word. In the example, two alternatives are retrieved. The top-most element (shown in blue) has been selected here. Consequently, the element moves to Panel 2 (A) (the result is not depicted in the static figure). In step 2, the user finishes the process of sentence production by selecting the green checkmark in Panel 2 (A). In response, the completed sentence moves to Panel 1. The result is depicted on the right-hand side of Panel 1 (marked as step 3), where the newly completed sentence *Ich koche das Mittagessen morgen.* 'I will cook lunch tomorrow.' is appended to the previously written text. The user can scroll through Panel 1 to look back within the flow of thoughts. A read-aloud function serves to remedy reading deficits. To support writers with low literacy skills, AAC symbols[7] supplement each word form. The produced text can be exported from the system for further use, with or without symbols. When the completed sentence is added to Panel 1 in step 3, Panel 2 simultaneously switches to the menu (B), offering sentence connectors (step 4). After an element that meets the user's communicative intention is chosen (or skipped *via* the large green arrow button), it is also moved to Panel 1; the system then switches Panel 2 back to (A) for the next sentence to be entered (step 5). At this point, only the punctuation cue is provided in order for the sentence type of the next sentence to be selected (cf. **Figure 3** for an illustration of how a question is typed).

In the following two sections, we elaborate on the computer-based linguistic support in the two phases of the text production process, i.e., within a sentence and between consecutive sentences. A particular highlight of the descriptions comprises arguments for an adequate user interface (UI) that, in addition to supporting low literacy skills, must compensate for factors such as working memory deficits within the target user group of people with a wide spectrum of cognitive impairments and/or learning difficulties.

3.3 Fast and Correct Extended Leichte Sprache Sentence Production

With respect to the goal of fast and correct typing, using Panels 2 (A) and 3 (cf. the left-hand side of **Figure 2**), we employ a variant of an NLG formulator. Its goal is to build up a derivation tree

[7]Here, we use the ARASAAC symbol set: www.arasaac.org. The symbol set can be adapted to the preferences of the user. Moreover, advanced users can switch this mode off.

based on the rules of a syntactic grammar (here, ELS constructions; if desired, the declarative grammar can easily be restricted to pure LS rules) so that syntactic correctness is automatically maintained. Based on this representation, the system produces correctly inflected word forms.

In NLG, the formulator usually administers only the best sentence representation; there is no UI enabling the selection of another option (*paraphrase*). We adopt a slightly more flexible formulator approach, developed for L2 learners of German. *COMPASS* (Harbusch et al., 2007, 2014) is a *natural-language paraphrase generator* that constructs the sentence the user has in mind in a step-by-step dialogue. The process is called *scaffolded writing*, in reference to the fact that the system is able to maintain the syntactic correctness of the construction the user is typing after the user has specified its grammatical function (Harbusch and Kempen, 2011). COMPASS is based on the grammar rules in *Performance Grammar*, a psycholinguistically motivated grammar formalism (Harbusch and Kempen, 2002; Kempen and Harbusch, 2002). The separation into distinct dominance and word order rules enable a flexible sentence-production process to suit the user's preferences. For example, the user can enter all arguments first to empty the short-term memory and then fully concentrate on arranging the constituents according to the intended discourse structure. Revisions made throughout the sentence at any point in time are retained. Upon request, the system reports whether or not a given construction is authorized by the grammar.

In essence, the overall lexicon covers the German CELEX (Gulikers et al., 1995). To obtain a reasonable suggestion list, this lexicon is restricted to L2-learner level A1/A2 in *ExtendedEasyTalk*. The system can also be adapted to the user's personal vocabulary (e.g., to include proper names of protagonists or places) or specific contexts (e.g., for school purposes).

The set of declarative rules applied by *ExtendedEasyTalk* is restricted to ELS constructions (cf. **Table 8**). For example, the range of verb forms is restricted to the active voice, indicative mood, and present and present perfect tenses. For auxiliaries and modals only, the preterite and the subjunctive mood are also offered. The system favors non-inversion word order. See **Table 9** for the order in which the constituents are presented in main declarative sentences. To provide an intuitive UI, it is crucial to avoid linguistic terms; therefore, we use cues in the forms of interrogative pronouns, as outlined in Column 1, to communicate with the user about grammatical functions and maintain scaffolded writing. (This technique resembles elementary school exercises for identifying grammatical function fillers in a sentence.) In return, the system is enabled to propose correctly inflected forms.

For the cues in Panel 2 (A), we illustrate how much information the system controls for word forms rather than showing the overall derivation tree here:

Wer 'who$_{nom}$' = **Subject**: $Ich_{PPER:1stPers,Sing,Nom}$
Tut 'does' = **Verb**: $koche_{VVFIN:1stPers,Sing,Pres,Ind,ActiveVoice}$
Wen[a] 'whom$_{acc}$' = **Direct object**: $das_{ART:Def,Neut,Sing,Acc}$ $Mittagessen_{NN:Neut,Sing,Acc}$

Now, we illustrate how this information is collected in an easy and intuitive step-by-step manner. Initially, the system presents the cues for all sentence components according to the canonical word order of the chosen sentence type. In a declarative main clause presupposing SVO word order, the subject is entered first. Based on the subject's number and person features, the system provides only correctly inflected verb forms for any typed word prefix managed by the cue **Tut** 'does' in Panel 3. In the example, the first two cues are filled with *ich koche* 'I cook.' In the list of choices, all forms with a separable verb prefix (SVP) (e.g., *koche ab* 'to boil off') that are covered by the currently selected lexicon are presented to the user. In the example, the verb lemma *kochen* (without any SVP) is chosen. If a verb with an SVP is selected, *ExtendedEasyTalk* assigns its word order position automatically (cf. *ich koche das Wasser ab* 'I boil off the water'). If the finite verb is an auxiliary, modal, or complement-taking verb included in ELS (e.g., *to try*), the sentence can continue with either a direct object or another verb with its own valency frame to be filled (e.g., *ich will ein Eis* 'I want an ice cream' vs. *ich will Ball spielen gehen* 'I want to go play ball'). This decision is presented to the user in a simple manner as a choice between the cues **Wen** 'whom$_{acc}$' and **Tut was** 'does what', respectively (see Steinmetz and Harbusch (2020) for details of how the user is supported in filling the valency frame of recursively added verbs).

After the verb is entered, the system keeps track of the overall valency restrictions/arguments provided in the lexicalized grammar, and every word form is supplemented with the appropriate syntactic structure. Incompletely filled valency frames cannot be moved to Panel 1; i.e., only correct sentences can be typed in *ExtendedEasyTalk*. In the example, a direct object cued by **Wen** 'whom$_{acc}$' has been filled with *das Mittagessen* 'the lunch', i.e., the cue **Wen** overarches the whole direct object. In Panel 3, the grammatical function currently active in Panel 2 (A) can be expanded (i.e., so that the same cue is active in Panels 2 (A) and 3) until the user selects another cue in Panel 3 or finishes the sentence. Modifier/adjunct cues are facultative. Displaying them should prompt the user to

TABLE 9 | List of constituents in a main declarative sentence (in the top panel, subject and finite verb forms are obligatory; the second panel enumerates further arguments/valency-frame fillers of the finite verb; and in the lower panel, adjuncts/modifiers are enumerated). Column 1 provides the cue words to be displayed as headers in Panels 2 (A) and 3.

Cue	Automatically inflected filler
Wer 'who$_{nom}$'	Elements of the **SUBJ**ect in nominative case
Tut 'does'	**FIN**ite verb form in active voice, present tense, coinciding in person and number with the subject
Wem 'whom$_{dat}$'	Elements of the **I**ndirect **O**bject in dative case
Wen 'whom$_{acc}$'	Elements of the **D**irect **O**bject in accusative case
P$_f$ was 'P$_f$ what'	Elements of the **P**repositional **O**bject in the case **P$_f$**, the instantiated preposition requires
Was tun 'what to do'	**P**ast **P**articiple in case the finite verb form is an auxiliary or **INF**initive in case the finite form is a modal or **I**nfinitive_with_**ZU** in case the finite form is a complement-taking verb
Wann 'when'	Elements of **MOD**ifier_time
Wo 'where$_{loc}$'	Elements of **MOD**ifier_location
Woher/-hin 'where$_{dir}$'	Elements of **MOD**ifier_direction from/to
Wie 'with what'	Elements of **MOD**ifier_instrument

supplement the sentences properly with audience-design information, such as the time and place of an event (cf. **Section 3.5**). In tests with L2 learners, beginners completely ignored this offer without feeling disturbed.

In our example, by selecting the modifier cue **Wann** 'when' (displayed in green), the user decided to add *tomorrow* as a temporal specification that the reader should know. At the point shown, the user has typed "*morge*" in the text-input field of Panel 3. Accordingly, a choice list presenting only temporal expressions is retrieved from the lexicon matching the current input string. In the figure, two items differing in inflected endings qualify as matches for "morge". The user navigates the completion list by scrolling to the intended form.

The typing speeds of all users, not only LS writers, are supported by prediction/completion lists (cf. typing on reduced keyboards on cell phones). The structure of Panel 3 borrows this concept. For any string prefix—even an empty one[8]—a suggestion list is displayed according to the active cue. *ExtendedEasyTalk*'s inflected suggestions speed up typing by unifying the two-stage process of selection and manual morphological adaptation. Hence, not only is syntactic correctness maintained but also typographical errors are also avoided. The weakness of this method became apparent to us during the evaluation of the system (cf. **Section 3.6.2**). Spelling deficits lead to empty lists and lengthy trial-and-error attempts. This issue will therefore have high priority in our future work.

In order to sum up the supportive features of *ExtendedEasyTalk*, we describe the typing of the six-word question *Soll ich morgen das Mittagessen kochen?* 'Should I cook lunch tomorrow?' in eight steps in **Figure 3**. The punctuation cue in Panel 2 (A) provides a declarative main clause by default. The user can switch to any other sentence type by *scanning*, i.e., repeatedly pressing the punctuation cue button until the correct choice appears. The word order, i.e., the order of the cue words in Panel 2 (A), is adapted according to the selected sentence type. The chosen punctuation mark automatically remains sentence-final at the end of the sentence during the process of typing the sentence.

In the example, the user has selected a yes/no question. For the typed string prefix 'soll', the system ranks the forms $soll_{1st/3rdPers,Sing}$ and $sollen_{1st/3rdPers,Plur}$ 'should' in the topmost positions. When the verb is typed in a sentence-initial element, i.e., lacking the features of the subject, the system cannot do any better. Thus, all possible (ELS-approved) verb forms have to be enumerated. Next, the user is required to fill in the obligatory subject. Now, subject–verb agreement can be used to filter the subject forms according to the chosen inflected verb form $soll_{1st/3rdPers,Sing.}$ (cf. step 2). In line with the typed string prefix 'ic,' the personal pronoun *ich* 'I' is the only option in the completion list. In step 3, the user has to follow the obligatory cue **Tut was** 'does what' elicited by the modal finite verb form. (The lexicon used here does not contain any lemma where *kochen* holds a separable verb prefix.) In step 4, the cues in Panel 3 present a list of the next constituents. The user can omit the facultative indirect object (**Wem**). In our example, the user decides to add the time of the event, *morgen* 'tomorrow,' before the direct object, *das Mittagessen* 'the lunch.' Note that advanced users can deviate from the default order by jumping directly to a certain cue in the list; the correct overall German word order is maintained by the system regardless. From step 5 onward, the filled cue **Wann** 'when' is displayed in Panel 2 (A). In steps 5 and 6, the user enters the direct object. In step 7, the user operates the checkmark button, and the sentence from Panel 2 (A) moves to Panel 1. In parallel, Panel 2 (B) appears (this is discussed in more detail in the next section).

3.4 Production of Elements for Sentence Coherence

Writing support is not restricted to intra-sentential items. Text consisting of a series of simple sentences with canonical SVO order lacks flow, and the writer's thoughts are only partially communicated. As in the conceptualizer of an NLG system, RST-like cues relating to the individual sentences should verbalize the user's communicative goal. As mentioned in **Section 1**, techniques for exemplifying RST relations are learned in exercises for complex clause construction in school. However, complex clauses with informative conjunctions are not available in either LS or ELS.

Having noticed in our corpus study that constructions of the form (conjunction/adverb [possibly followed by a colon] // main clause) improve text understandability, we decided to add this concept to *ExtendedEasyTalk*. We assume that our users are familiar with the use/meaning of most conjunctions in German (cf. the frequencies of subordinating conjunctions, i.e., KOUS, in LST, provided in **Section 2.3.2**). Moreover, the LeiSA study evaluated subordinate clauses with *während*/*nachdem* 'while/after' as easy for LS readers, even with subordinate VF word order. Thus, instead of using technical terms like ELABORATION to refer to RST relations, we ask the user to select an appropriate conjunction/adverb. Whenever the user finishes a sentence (by pressing the green checkmark button; cf. step 2 in **Figure 2** and step 8 in **Figure 3**), Panel 2 switches to menu (B). This menu consists of nine buttons (pressing the green arrow button on the right side of the menu omits the addition of a connector). In accordance with suggestions made by AAC experts (cf. **Section 3.6.1**), we restrict the choice to those forms widely used under LS rules Netzwerk Leichte Sprache to avoid overtaxing the user.

We group the elements in the menu according to conjunction type. In the upper row, the coordinating conjunctions *und* 'and' *oder* 'or' and *aber* 'but', and the colon are provided. We realize that the colon is highly ambiguous in LS texts; however, its use is widespread (Bredel and Maaß, 2016, p. 254). We therefore offer this choice to prevent users from having to search for this option. In the second row, the user is presented with the subordinating conjunctions *weil* 'because' and *wenn* 'if', the adverb *darum* 'therefore', and a button *Andere wählen* 'Choose other'. In our corpus study into LST, we observed further variation for POS = KOUS; hence, more advanced users can

[8]Cf. the active completion list of an empty word prefix in screenshot 7 in **Figure 3**. The list reflects the currently very limited context.

Initial story writing checklist

1	Lieber Peter!	Dear Peter!
2	Denk dir eine Geschichte aus!	Think up a story!
3	Bist Du fertig?	Are you ready?
4	Beantworte zuerst die wichtigen Fragen von	First answer important question with respect
5	der Geschichte!	to the story.
6	Weil:	Because:
7	Die Leser*innen wissen viele Dinge nicht.	The readers does not know about many things.
8	Deshalb:	Therefore:
9	Wir schreiben die wichtigen Dinge!	We write the important things.
10	*Wer soll deine Geschichte lesen?*	*Who is the reader of your story?*
11	*Mache ein Kreuz bei den Passenden:*	Mark the fitting ones:
12	• Mama/Papa	• Mum/Dad
13	• Frank, mein Betreuer	• Frank, my caregiver
14	• Susi, meine Schulfreundin	• Susi, my schoolmate
15	*Wer fehlt in der Liste?*	*Who is missing in the list?*
16	Wir müssen jetzt herausfinden:	We have to find out now:
17	Was wissen die Leser*innen schon?	What do the readers already know?
18	• Wann findet die Geschichte statt?	• When does the story take place?
19	• Wo findet die Geschichte statt?	• Where does the story take place?
20	• Wer spielt mit?	• Who takes part?

FIGURE 4 | Initial text writing checklist presented at the beginning of a writing workshop session (for details, see the text; to aid readability, the formatting in *ExtendedEasyTalk* is omitted here).

browse through all conjunctions. For consistency and overall ease of use of the system, Panel 3 provides a list of conjunction choices with the same selection options as for word forms in sentence typing. In case the option selected—either by button or in Panel 3—is a word, it is added as a separate line at the end of Panel 1 (cf. *darum* 'therefore' in the second line on the left-hand side of Panel 1 in **Figure 2**); the colon is appended to the last sentence in Panel 1 and replaces the previously written punctuation symbol.

So far, we have illustrated how users can use *ExtendedEasyTalk* to freely type ELS constructions. In the next section, we focus on the teaching of text production concepts by wrapping an active control structure around the key components for typing.

3.5 Control Mechanisms of the Writing-Workshop Mode in *ExtendedEasyTalk*

Prolific writers know that coherent, understandable text has to emphasize audience/reader-design concepts. *ExtendedEasyTalk* can teach basic writing workshop concepts; to do this, the system takes the initiative by asking questions (stated in ELS) at different stages of text production. For convenience, this mode can be easily ended or reactivated at any point in time.

At the beginning of a text, the user has to answer questions from a checklist (cf. **Figure 4** for an excerpt; the questionnaire presented to the user can be adapted to specific text genres). Depending on the user's reading fluency, the questions can be read aloud to them (e.g., by a caregiver), or the read-aloud function of the system can be used to speed up the dialogue. As far as possible, the individual questions of the checklist offer a range of alternatives. Where this is not possible, the user types the answer using *ExtendedEasyTalk*. The answer lists can be pre-adapted to the current user, e.g., the names of the user (in our example, *Peter*, a male user) and caregivers, friends, teachers, etc. The dialogue starts with an introductory text (cf. lines 1–9). Lines 10–15 collect background information on the hearer/reader in an intuitive manner. Lines 16–20 show part of the collection of background information for the text the user would like to write.

To characterize all the protagonists in the list of actors (line 20) so that the reader can identify them clearly, a sequence of questions is asked. Different options are tested. Does the reader already know the name of the actor(s)? Can they be introduced by name? Can a characterization of the person(s) be added to enable the reader to become familiar with them (e.g., *Frank is my caregiver; Susi is my schoolmate*)? Such a session avoids the need for relative clauses (although these are included in ELS).

Similarly, the background of every sentence is established through questions referring to the adjuncts (cf. the modifier cues in **Table 9**). In the writing-support mode, the system will infer that the place and time have changed. Instead of simply displaying the temporal modifier cue (cf. step 4 in **Figure 3**), it will ask the user an explicit question to keep track of all changes or details unknown to the reader.

This looks cumbersome; however, the effort pays off when protagonists are referred to during sentence production. Not only can suggestions of personal pronouns be made by the system but also, in addition, the temporal and spatial modifiers are prefilled with the initial/most recent filler, ruling out wrong assumptions by the reader. In **Figure 5**, the first sentence, S1, refers to *tomorrow* due to the intervention of the system to initialize time and place. Without active intervention, the second sentence, S2, would also display the **Wann** cue filled with *tomorrow*. This leads the user to notice the clash with their intended content; in the example, the cue is revised to *already*. The

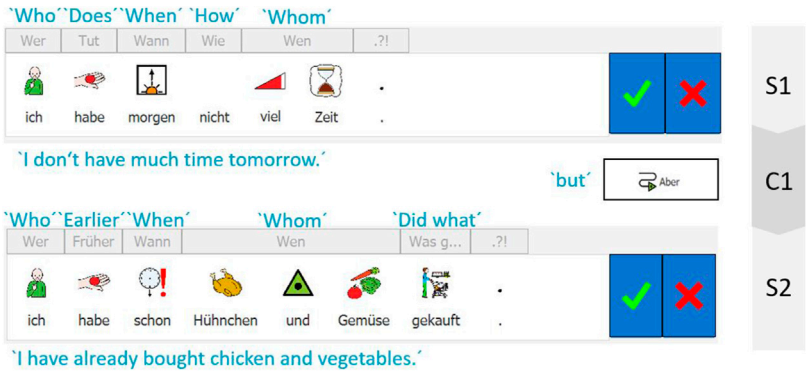

FIGURE 5 | Temporal modifier updating in the writing-workshop mode of *ExtendedEasyTalk*.

user types S2 in the present perfect tense. The question of whether or not the system should actively provide tense suggestions remains open. We hesitate to make our system overly adaptive. Many users—irrespective of their specific user group—do not appreciate non-static UIs (Lee and Yoon, 2004).

Finally, we report our evaluation results.

3.6 Evaluation

As testing software with people with disabilities presents special challenges and organizational overheads (see, e.g., Henry, 2007 or Lazar et al., 2017), we decided to test as many facets of our system as possible with substitute users. For our target group, the initial impression is crucial; many AAC solutions are abandoned due to avoidable interface flaws (see, e.g., Dawe 2006; Fager et al., 2006; Waller 2019). We therefore tried to make clever use of two substitute groups:

(1) Experts in the field of accessible learning and barrier-free communication, who are able to judge the simplicity/adequacy of the individual steps to be performed; and
(2) L2 learners, ranging from those with limited computer skills, who reviewed early versions of the interface in order to make the UI as simple as possible, to those with high computer skills, who reviewed the most recent interface to identify the situations in which users expect more proactive linguistic support.

The testing of the newest version of the system with participants with intellectual/learning disabilities had to be suspended for more than a year due to the COVID-19 pandemic. At the time of writing, we have just begun a series of tests (see **Section 4**).

In the following sections, we summarize the main findings[9].

3.6.1 System Evaluation via Interviews With AAC Experts

The group of LS readers is very heterogenous (Bredel and Maaß, 2016, p. 139). We talked to experts in AAC and LS familiar with its diverse needs. The expert group consisted of:

(1) A male LS reader with learning disabilities, who regularly reviews LS texts and does some writing himself; accordingly, he can be qualified as an (advanced) real user from our target group;
(2) A male LS writer, who leads an LS writing workshop and regularly writes and proofreads LS texts;
(3) A female AAC expert, who implements AAC solutions within a facility for people with multiple and/or severe cognitive or sensory disabilities; and
(4) A male domain expert, who has worked with people with severe cognitive and physical disabilities for over 35 years and is familiar with many communication methods and their evolution.

We used the *case study* method[10], employing exploratory think-aloud probes followed by semi-structured interviews. As the first probe, the participants were asked to write a series of prepared sentences of varying complexity, then, to freely formulate a text with *ExtendedEasyTalk*. In the interview, we focused on the participants' assessment regarding potential reasons for the (non-)acceptance of our system. Important insights are presented below.

All experts acknowledged that *ExtendedEasyTalk* meets the requirements of users, who know alphabetic characters and have basic spelling skills but have difficulties writing whole words or complex sentences and coherent texts.

All experts gave positive feedback on the use of AAC symbols in combination with word forms in *ExtendedEasyTalk*. They

[9]Additional details on the test setups can be found in the **Supplementary Material** of this paper.

[10]The qualification of this study as a case study results from its matching the four key criteria given by Lazar et al. (2017): (1) In-depth examination of a small number of cases, (2) examination in context, (3) multiple data sources, and (4) emphasis on qualitative data and analysis.

appreciated the fact that the range of both symbols and vocabulary is easily customizable and expandable in *ExtendedEasyTalk* as it is crucial for users to be able to use familiar vocabulary and proper names to describe their everyday lives and individual interests. One of the AAC experts phrased it like this: "People need to be able to describe their world in their own words." This includes users being able to use familiar symbols (e.g., *PCS* or *METACOM*[11]), given that AAC symbols and words are commonly learned in combination. In the broader area of evaluation for AAC needs, they focused on the accessibility of *ExtendedEasyTalk*. The experts appreciated the read-aloud function for the produced text in *ExtendedEasyTalk* and the option for users to use their own input devices (e.g., accessible keyboards). They suggested adding the ability to operate the system by *scanning*[12], i.e., the system iterates sequentially through all options until the user instructs the system to stop and make a selection.

The concept of writing a sentence by answering sequences of wh-questions was appreciated by all experts. They related the answering of wh-questions to parent–child dialogues (see, e.g., Brandt et al., 2016). The experts liked the predictive force of the completion list, which reduces the need for typing and supports correct inflection. They anticipate that this will simultaneously give users a feeling of security and speed up typing. Regarding the mechanism for using complement-taking verbs, we received positive feedback from our LS reader and the AAC expert; they described it as a "reasonable way" to access these constructions. However, all experts recommended thorough testing with varied groups of LS users. The RST-related aspects of the system were recognized by the experts as a good way to practice connecting sentences.

We were pleased with the largely positive feedback and the high level of interest in the system shown by the experts; however, we are aware that the participants knew they were talking to the developer. Moreover, we keep in mind the general warning that one should not over-generalize case study results.

3.6.2 Testing With Two L2-learner Groups

Testing with L2 learners with low computer and low literacy skills

Here, we again chose the case study method as an appropriate way to gain insights. We tested the system with three male L2 learners with predominantly oral German language skills at Common European Frame of Reference (CEFR) L2-level A1–A2[13]. All three are literate in their native languages (Amharic, Tigrinya, and French/Cotocoli). Their computer skills were rudimentary. They were able to write only very short messages in German, e.g., to make appointments *via* messenger apps.

We conducted semi-structured interviews (in ELS wording) assessing the supportive features for sentence formulation—supplemented with situational follow-up questions to evaluate possible workarounds when deficiencies were discovered. Here, we sum up findings of interest that have led to revisions in the interface.

During all tests, the same barrier to selecting word forms from the completion list hindered fast typing: Spelling errors or mistakes in selecting the gender of an article (*der/die/das*$_{inflected}$ 'the') resulted in unexpected completion lists. The support while entering sentences with complement-taking verbs was highly appreciated. (*ExtendedEasyTalk* automatically moves the infinitive (with *zu*) to the clause-final position in German—a different position from the participants' mother tongues.) Moreover, all participants liked the support for correctly conjugating verbs and choosing correctly inflected word forms.

Without over-generalizing, we observed that beginner-level users neither recognize nor use the full extent of the system's linguistic scope and support. This observation led to the design of an active teaching strategy in form of the "writing workshop" (cf. **Section 3.5**). At the same time, it demonstrated that users are able to write according to their personal preferences and skill levels. We plan to document in a longitudinal study whether users improve their personal writing skills over time with the support of *ExtendedEasyTalk*. Finally, we received positive feedback regarding the combination of words and visual symbols. Participants emphasized that this helped them recognize and remember words more easily. Without being asked in the probes, the users actively resorted to the read-aloud function.

Based on these observations, we developed the current interface of *ExtendedEasyTalk*, presented in this study.

Testing with L2 learners with high computer and high literacy skills

In recent test sessions via remote desktop control, conducted with ten L2 learners with different native languages (Arabic, Romanian, Swedish, Mandarin, and Spanish), we tried to find indications of the linguistic support our target group would expect from an advanced writing tool. We recruited this group from among IT experts with German skills between CEFR L2-level A1–B2; additionally, all participants were fluent in English. This was therefore used as the common language between participants and the interviewer during test sessions.

Here, we apply *discount testing* (Nielsen, 1989), a well-established method in *Human–Computer Interaction* (HCI), which was essentially born of necessity. Usability projects should not fall at the hurdle of small budgets. To avoid this, Nielsen proposed a methodology for cheap usability testing: With a handful of participants, a focus on qualitative studies, and the use of the think-aloud method, the majority of usability flaws can be identified. We have chosen the term "discount testing" rather than "case study" here to emphasize that the tasks to be performed were fixed, in contrast to those in the two earlier evaluations (reported above), which allowed for digression during probing (cf. condition (2) in Footnote 11; we are aware that the line between case studies and discount testing is blurred.)

[11]www.goboardmaker.com/pages/picture-communication-symbols (PCS symbols) and www.metacom-symbole.de (METACOM symbols).

[12]https://praacticalaac.org/tag/scanning/ (broad overview of different scanning techniques).

[13]www.coe.int/en/web/common-european-framework-reference-languages (CEFR).

TABLE 10 | Table of all abbreviations.

Systems/Theories/ Corpora	Expansion
AAC	Augmentative and Alternative Communication
ELS	Extended Leichte Sprache
LS	Leichte Sprache
LST	Leichte Sprache treebank
NLG	Natural-language generation
RST	Rhetorical Structure Theory
TüBa-D/S	treebank of spoken German (systematically called VERBMOBIL)
TüBa-D/Z	treebank of written German
VERBMOBIL	synonym for TüBa-D/S
UI	user interface
VUI	voice user interface
Linguistic terms	**Expanded term**
DO	direct object
IO	indirect object
NP	noun phrase
POS	part-of-speech (see next panel for POS tags used here)
PP	prepositional phrase
PRED	predicate
SB	subject
V2	verb second word order in German main declarative clauses
VF	verb final word order in German subordinate clauses
SVO	the canonical word order in German main declarative clauses
Used POS tags	**Part of Speech**
APPO	post-postitioned preposition
APPRART	pre-postitioned prepositions with agglutinated definite determiner
APPR	pre-postitioned preposition
APZR	circum-positioned preposition
KOUI	subordinating conjunction followed by *um zu* + INFinitive
KOUS	subordinating conjunction followed by a sentence
NN	lexical noun
PDS	substituting demonstrative pronoun
PIS	substituting indefinite pronoun
PPER	irreflexive personal pronoun
PPOSS	substituting possessive pronoun
PRELS	relative pronoun
PRF	reflexive personal pronoun
PWS	substituting interrogative pronoun
SVP	separable verb prefix
VA(FIN/IMP/PP/INF/IZU)	auxiliary as FINite form/IMPerative/PastParticiple/ INFinitive/InfwithZU
VM(FIN/IMP/PP/INF/IZU)	modal as FINite form/IMPerative/PastParticiple/INFinitive/ InfwithZU
VV(FIN/IMP/PP/INF/IZU)	lexical verb as FINite form/IMPerative/PastParticiple/ INFinitive/InfwithZU
POS wild cards	**Expanded term**
APPR.*	APPR and APPRART (".*" refers to zero to *n* arbitrary characters)
N.*	NE and NN
PRO	PDS and PIS and PPOSS and PRELS and PRF and PPER and PWS
V.FIN	VAFIN/VMFIN/VVFIN (". " refers to one arbitrary character)

To obtain comparable results, we presented each participant with a picture supplemented by five sentences that provided the necessary German vocabulary, allowing the participants to focus on typing rather than on finding German words. We asked them to think aloud while typing sentences in *ExtendedEasyTalk* and using the connectors provided in Panel 2 (B) to relate the sentences in a short story—supposedly to be emailed to a solely German-speaking child. At the end of each session, we conducted a short semi-structured interview and asked the users to fill in a user experience questionnaire[14], in English for their convenience. We focused mainly on the question: What support do IT specialists with few (written) German skills expect from an advanced writing system? We also took notice of any flaws in the menus.

The results for within-sentence support were mainly positive; all participants felt supported by the system. Moreover, they reported that they found the system easy to understand and intuitive to operate. This feedback is reflected in the fact that all learners managed to use the system autonomously, following a demonstration in which the interviewer produced an example sentence using *ExtendedEasyTalk*. As in the other L2-learner group, the IT experts appreciated the combination of symbols and words, the automatic inflection of words, and the word-ordering support.

However, this group asked for active support in spelling and/ or finding a German word—an issue we also noticed in the earlier test sessions. This problem needs to be addressed in the next *ExtendedEasyTalk* prototype, particularly as many LS readers presumably have spelling problems as well.

We want to take up the idea of three participants who proposed adding suggestions based on words entered via a microphone to the already available vocabulary suggestions based on spelling. We are planning a sub-series of tests with users with functional speech, employing a speech recognition device as a voice user interface (VUI). In line with suggestions by our participants, we want to implement a mode offering lemmas for a preselected topic or domain—similar to common AAC grid layouts where the vocabularies of customizable categories or contexts, such as 'food', 'hygiene', 'in school', or 'at home', are proposed (cf. the AAC systems presented in **Section 3.1.1**).

The menu producing sentence connectors outlined in **Section 3.4** was judged to be intuitive and meaningful by all participants. However, all participants expected the system to provide feedback on the quality of their choices. This expectation probably originates from the fact that most participants have experience with fill-the-gap German language-learning software. While they certainly appreciated the freedom that *ExtendedEasyTalk* offers compared to such exercises, they would welcome help in this area similar to the system's word order expertise. This problem is difficult to solve in *ExtendedEasyTalk* as we have no overall content representation to determine what is a reasonable/ necessary relation to use.

The system was perceived by all participants as uncluttered, organized, and clearly structured. However, the IT experts criticized the "outdated" look and feel of its graphical UI design. We assume that this feedback is connected to the fact that people working in IT are used to modern UI aesthetics[15], whereas the UI of *ExtendedEasyTalk* is oriented towards the design patterns of AAC applications and focuses on accessible

[14]https://www.ueq-online.org/ (user experience questionnaire UEQ).

[15]https://developer.apple.com/design/ ("Flat Design"); https://material.io/design ("Material Design").

design and good readability for the target group (cf. the UI design of the AAC systems cited in **Section 3.1** and the style guide for LS texts in Netzwerk Leichte Sprache).

Nevertheless, according to the user-experience questionnaire, eight of the 10 participants would recommend or use the system for language learning. Six of them would prefer to run *ExtendedEasyTalk* as a smartphone app instead of a desktop app. In line with the recent developments in AAC technology on common commercial handheld devices (which we have cited above), a new version of *ExtendedEasyTalk* for smartphones and tablets is on our to-do list.

4 CONCLUSIONS AND FUTURE WORK

We have defined Extended Leichte Sprache (ELS), a native extension of LS based on observations from a corpus study of LST, VERBMOBIL, and TüBa-D/Z. *ExtendedEasyTalk* facilitates fast and correct typing for the target language, ELS, by employing linguistic processing to a large extent. The system strives for correct and understandable text during both sentence production and sentence combination. Interactive user guidance is tailored to the personal level of grammatical support needed to produce correct and coherent complex content. Most importantly, linguistic decisions are formulated in an easy and intuitive manner. In order to promote LS text production by the LS community themselves, *ExtendedEasyTalk* teaches text-writing skills to more advanced users. In the writing-workshop mode, the writer is trained to foresee/resolve underspecifications that result from matching the presented facts with the readers' presupposed knowledge.

As mentioned above, usability tests with the target group in the presence of the researchers have been impossible for more than a year now due to COVID-19. Unfortunately, people with disabilities are particularly affected by the social isolation caused by COVID-19 (Rödler, 2020; Portal et al., 2021). This highlights the importance of the further development of systems like *ExtendedEasyTalk* that support the target group of LS readers to communicate their thoughts (remotely) in the form of text (messages). We are currently conducting a broad-scope system-evaluation study with members of the target population, aiming to gain new insights to guide the redesigning of the UI for the next version of *ExtendedEasyTalk*. As mentioned in the evaluation in **Section 3.6.2**, this prototype should include a better strategy in Panel 3 to avoid empty suggestion lists when choosing words. A series of prototypes for a cell phone version is also under evaluation.

In addition, we would like to extend the system's writing-workshop mode. For example, in a revision phase across the text provided in Panel 1, referents could become pronominalized or elided in consecutive SVO sentences. The use of other types of substituting pronouns (e.g., demonstratives [PDS]) could be trained. We realize that these constructions are not authorized by all LS definitions; however, they are frequent in LS texts, and readers can therefore be expected to be familiar with them. As pronouns shorten a text, writers should be trained to use them.

Given the fact that our natural-language paraphrase generator administers the syntactic structures of all typed sentences, we plan to implement an additional export function that transforms the typed LS text into syntactic constructions in standard German. So, documents for non-LS readers can automatically be produced. For this purpose, a component similar to the Aggregator (component (2) in the NLG blue print in **Figure 1**) has to be added. As mentioned, the difficulty with this task is the problem to map appropriate portions of sentences and their connectors onto the overall discourse structure of the text. The latter topic is related to the question we brought up in **Section 2.3**. What is a good way of replacing standard German constructions by (E)LS conform ones? For finding easy-to-perform rules of thumb, we plan more thorough studies into LST.

SORTED LIST OF ABBREVIATIONS USED IN THE PAPER

Due to the interdisciplinary nature of our study, in **Table 10**, we provide a list of all acronyms used in the text, grouped in four panels.

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 were 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.

AUTHOR CONTRIBUTIONS

This is joint work; the substantive focus of each author was as follows: KH: ELS definition based on corpus study; and IS: ExtendedEasyTalk prototype and evaluations.

ACKNOWLEDGMENTS

The authors are extremely grateful to the two reviewers and the editor Scott Hale for their constructive, very detailed, and insightful suggestions and comments. All remaining errors are the authors' own responsibility.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.689009/full#supplementary-material

REFERENCES

Bell, A. (1984). Language Style as Audience Design. *Lang. Soc.* 13, 145–204. doi:10.1017/S004740450001037X

BITV2.0 (2011). Verordnung zur Schaffung Barrierefreier Informationstechnik Nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0). Barrierefreie-Informationstechnik-Verordnung vom 12. September 2011 (BGBl. I S. 1843), die zuletzt durch Artikel 1 der Verordnung vom 21. Available at: http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html.

Bock, B. M. (2017). Das Passiv- und Negationsverbot 'Leichter Sprache' auf dem Prüfstand – Empirische Ergebnisse aus Verstehenstest und Korpusuntersuchung. *Sprachreport.* 33 (1), 20–28. doi:10.1007/978-3-476-04852-3_39

Bock, B. M. (2019). *Leichte Sprache" – Kein Regelwerk: Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeiSA-Projekt*. Berlin, Germany: Frank & Timme.

Brandt, S., Lieven, E., and Tomasello, M. (2016). German Children's Use of Word Order and Case Marking to Interpret Simple and Complex Sentences: Testing Differences Between Constructions and Lexical Items. *Lang. Learn. Development.* 12, 156–182. doi:10.1080/15475441.2015.1052448

Bredel, U., and Maaß, C. (2016). *Leichte Sprache: Theoretische Grundlagen. Orientierung für die Praxis*. Berlin, Germany: Bibliographisches Institut.

Dawe, M. (2006). "Desperately Seeking Simplicity," in Proceedings of the 2006 Conference on Human Factors in Computing Systems (SIGCHI 2006), Montréal, QC, Canada, April 22–27, 2006, 1143–1152. doi:10.1145/1124772.1124943

Demasco, P. W., and McCoy, K. F. (1992). Generating Text From Compressed Input. *Commun. ACM.* 35, 68–78. doi:10.1145/129875.129881

Fager, S., Hux, K., Beukelman, D. R., and Karantounis, R. (2006). Augmentative and Alternative Communication Use and Acceptance by Adults With Traumatic Brain Injury. *Augmentative Altern. Commun.* 22, 37–47. doi:10.1080/07434610500243990

G2.com (2021). Best Natural Language Generation (NLG) Software. Available at: https://www.g2.com/categories/natural-language-generation-nlg (Accessed January 16, 2021).

Gatt, A., and Krahmer, E. (2018). Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation. *J. Artif. Intelligence Res.* 61, 65–170. doi:10.1613/jair.5477

Gulikers, L., Rattnik, G., and Piepenbrock, R. (1995). *German Linguistic Guide of the CELEX Lexical Database Tech. Rep*. Philadelphia, MA: Linguistic Data Consortium.

Hansen-Schirra, S., and Maaß, C. (2020). *Easy Language Research: Text and User Perspectives*. Berlin, Germany: Frank & Timme. doi:10.26530/20.500.12657/42088

Harbusch, K., Härtel, J., and Cameran, C. J. (2014). "Automatic Feedback Generation for Grammatical Errors in German as Second Language Learners Focused on the Learner's Personal Acquisition Level," in Proceedings of EUROCALL 2014 - Call Design: Principles and Practice (Groningen, Netherlands), 140–145.

Harbusch, K., and Kempen, G. (2002). "A Quantitative Model of Word Order and Movement in English, Dutch and German Complement Constructions," in Proceedings of the 19th International Conference on Computational Linguistics – Volume 1 (COLING '02), Taipei, Taiwan, 1–7. doi:10.3115/1072228.1072262

Harbusch, K., and Kempen, G. (2011). "Automatic Online Writing Support for L2 Learners of German Through Output Monitoring by a Natural-Language Paraphrase Generator," in WorldCALL - International Perspectives on Computer-Assisted Language Learning, New York, NY: Routledge/Taylor&Francis Group, 128–143.

Harbusch, K., and Kempen, G. (2007). "Clausal Coordinate Ellipsis in German: The TIGER Treebank as a Source of Evidence," in Proceedings of the Sixteenth Nordic Conference of Computational Linguistics (NODALIDA 2007), Tartu, Estonia, 81–88.

Harbusch, K., and Kempen, G. (2009). "Coordinate Ellipsis and its Varieties in Spoken German: A Study With the TüBa-D/S Treebank of the VERBMOBIL Corpus," in Proceedings of the 8th International Workshop on Treebanks and Linguistic Theories (TLT 8), Milano, Italy, 83–94.

Harbusch, K., van Breugel, C., Koch, U., and Kempen, G. (2007). "Interactive Sentence Combining and Paraphrasing in Support of Integrated Writing and Grammar Instruction: A New Application Area for Natural Sentence Generators," in Proceedings of the 11th European Workshop on Natural Language Generation (ENLG 2007), Dagstuhl, Germany, 65–68.

Henry, S. L. (2007). *Just Ask: Integrating Accessibility Throughout Design*. Madison, WI: Shawn Lawton Henry. Available at: www.uiAccess.com/JustAsk/.

Hovy, E. H. (1988). "Planning Coherent Multisentential Text," in Proceedings of the The 26th Annual Meeting of the Association for Computational Linguistics (ACL 88), New York, NY, 163–169. doi:10.3115/982023.982043

Inclusion Europe (2009). Informationen für alle – Europäische Regeln, wie man Informationen leicht lesbar und leicht verständlich macht. Available at: https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE_Information_for_all.pdf (Accessed March 29, 2021).

Kempen, G., and Harbusch, K. (2019). Mutual Attraction Between High-Frequency Verbs and Clause Types With Finite Verbs in Early Positions: Corpus Evidence From Spoken English, Dutch, and German. *Lang. Cogn. Neurosci.* 34, 1140–1151. doi:10.1080/23273798.2019.1642498

Kempen, G., and Harbusch, K. (2002). Performance Grammar: A Declarative Definition. *Lang. Comput.* 45, 148–162. doi:10.1163/9789004334038_013

Kempen, G., and Harbusch, K. (2016). Verb-second Word Order After German Weil 'Because': Psycholinguistic Theory From Corpus-Linguistic Data. *Glossa: A J. Gen. linguistics.* 1 (1), 1–32. doi:10.5334/gjgl.46

König, E., and Lezius, W. (2003). *The TIGER Language: A Description Language for Syntax Graphs, Formal Definition. Tech. Rep*. Universität Stuttgart: Institut für maschinelle Sprachverarbeitung (IMS).

Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Comput. Linguistics.* 40, 121–170. doi:10.1162/COLI_a_00168

Lancioni, G. E., Singh, N. N., O'Reilly, M. F., and Alberti, G. (2019). Assistive Technology to Support Communication in Individuals With Neurodevelopmental Disorders. *Curr. Dev. Disord. Rep.* 6, 126–130. doi:10.1007/s40474-019-00165-x

Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research Methods in Human Computer Interaction*. 2nd Ed Cambridge, MA: Morgan Kaufmann, an imprint of Elsevier. doi:10.1016/B978-0-12-805390-4.00016-9

Lee, D.-S., and Yoon, W. C. (2004). Quantitative Results Assessing Design Issues of Selection-Supportive Menus. *Int. J. Ind. Ergon.* 33, 41–52. doi:10.1016/j.ergon.2003.07.004

Lieske, C., and Siegel, M. (2014). Verstehen Leicht Gemacht. *Technische Kommunikation.* 1, 44–49.

Light, J., McNaughton, D., and Caron, J. (2019). New and Emerging AAC Technology Supports for Children With Complex Communication Needs and Their Communication Partners: State of the Science and Future Research Directions. *Augmentative Altern. Commun.* 35, 26–41. doi:10.1080/07434618.2018.1557251

Light, J., and McNaughton, D. (2012). The Changing Face of Augmentative and Alternative Communication: Past, Present, and Future Challenges. *Augmentative Altern. Commun.* 28, 197–204. doi:10.3109/07434618.2012.737024

Linzen, T., and Baroni, M. (2021). Syntactic Structure from Deep Learning. *Annu. Rev. Linguist.* 7, 195–212. doi:10.1146/annurev-linguistics-032020-051035

Löffler, C. (2015). Leichte Sprache als Chance zur Gesellschaftlichen Teilhabe Funktionaler Analphabeten. *Didaktik Deutsch.* 38, 17–23.

Maaß, C., Rink, I., and Zehrer, C. (2014). Leichte Sprache in der Sprach- und übersetzungswissenschaft. *Susanne Jekat, Heike Elisabeth Jüngst, Klaus Schubert & Claudia Villiger (Hg.), Sprache barrierefrei gestalten. Perspektiven aus der Angewandten Linguistik.* 53, 86.

Mann, W. C., and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Interdiscip. J. Study Discourse.* 8 (3), 243–281. doi:10.1515/text.1.1988.8.3.243

Netzwerk Leichte Sprache (2013). Die Regeln für Leichte Sprache. Available at: https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf (Accessed March 29, 2021).

Ney, J. W. (1980). A Short History of Sentence Combining: Its Limitations and Use. *English Education.* 11, 169–177. doi:10.1177/003368828001100103

Nielsen, J. (1989). "Usability Engineering at a Discount," in Proceedings of the Third International Conference on Human-Computer Interaction on Designing and Using Human-Computer Interfaces and Knowledge Based Systems (USA: Elsevier Science Inc), 394–401. doi:10.5555/92449.92499

Nordquist, R. (2018). An Introduction to Sentence Combining. ThoughtCo. Available at: https://www.thoughtco.com/an-introduction-to-sentence-combining-1692421 (Accessed March 18, 2021).

Nüssli, N. D. (2019). *Übersetzen in die Leichte Sprache: Übersetzungsprobleme, Übersetzungslösungen und Auswirkungen auf das Textverständnis von Menschen mit Downsyndrom: am Beispiel von Texten zum Thema Gesundheit*. Zürich, Switzerland: ZHAW Züricher Hochschule für Angewandte Wissenschaften. doi:10.21256/zhaw-3344

Ogden, C. K. (1930). *Basic English: A General Introduction with Rules and Grammar*, Vol. 1. London, UK: Paul Treber&Co.

Petrova, S., Solf, M., Ritz, J., Chiarcos, C., and Zeldes, A. (2009). Building and Using a Richly Annotated Interlinear Diachronic Corpus: The Case of Old High German Tatian. *Traitement Automatique des Langues.* 50, 47–71.

Portal, H., Schmidt, G., Fernández, R. C., Marcondes, B., Ševerepa, M., Dragičević, V., et al. (2021). Neglect and Discrimination. Multiplied. How Covid-19 Affected the Rights of People With Intellectual Disabilities and Their Families. Inclusion Europe. Available at: https://www.inclusion-europe.eu/wp-content/uploads/2020/11/COVID-report-Final.pdf (Accessed March 09, 2021).

Pottmann, D. M. (2019). "Leichte Sprache and Einfache Sprache" - German plain Language and Teaching DaF German as a Foreign Language. *Studia Linguistica.* 38, 81–94. doi:10.19195/0137-1169.38.6

Reichardt, A., Kruse, N., and Lipowsky, F. (2014). Textüberarbeitung mit Schreibkonferenz oder Textlupe. Zum Einfluss der Schreibumgebung auf die Qualität von schülertexten. *Didaktik Deutsch.* 19, 64–85.

Reis, M. (2013). "Weil-V2"-Sätze und (k)ein Ende? Anmerkungen zur Analyse von Antomo & Steinbach (2010). *Z. für Sprachwissenschaft.* 32, 221–262. doi:10.1515/zfs-2013-0008

Reiter, E., and Dale, R. (2000). Building Natural Language Generation Systems. *Stud. Nat. Lang. Process.* Cambridge UK, Cambridge Univ. Press. doi:10.1017/CBO9780511519857

Rödler, P. (2020). Totale Institution. *Bp.* 59, 345–358. doi:10.30820/0341-7301-2020-4-345

Ross, J. R. (1967). Constraints on Variables in Syntax. Ph.D. Thesis. Cambridge, MA/UASA: MIT.

Saddler, B., and Preschern, J. (2007). Improving Sentence Writing Ability Through Sentence Combining Practice. *Teach. Exceptional Child.* 29, 6–11. doi:10.1177/004005990703900301

Sennrich, R., Volk, M., and Schneider, G. (2013). "Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-Tagging, and Morphological Analysis," in Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2013), Hissar, Bulgaria, 601–609.

Stegmann, R., Telljohann, H., and Hinrichs, E. W. (2000). *Stylebook for the German Treebank in Verbmobil. Tech. Rep.*, 239. Saarbrücken, Germany: DFKI.

Steinmetz, I., and Harbusch, K. (2020). "Enabling Fast and Correct Typing in 'Leichte Sprache' (Easy Language)," in Proceedings of the The Fourth Widening Natural Language Processing Workshop (WINLP 4), Seattle, CA, USA. 64–67.

Suter, J., Ebling, S., and Volk, M. (2016). "Rule-based Automatic Text Simplification for German," in Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016), Zürich, Switzerland, 279–287.

Telljohann, H., Hinrichs, E. W., Kübler, S., Zinsmeister, H., and Beck, K. (2009). *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Tech. Rep*. Tübingen, Germany: Seminar für Sprachwissenschaft, Universität Tübingen.

Temperley, D. (2019). Rare Constructions Are More Often Sentence-Initial. *Cogn. Sci.* 43, e12714. doi:10.1111/cogs.12714

Tritscher, J. (2016). *Evaluation von State of the Art Syntaxparsern auf deutschen Romanen des 16. bis 20. Jahrhunderts. B.S. Thesis*. Würzburg, Germany: Lehrstuhl für Künstliche Intelligenz und Angewandte Informatik an der Universität Würzburg.

Wahlster, W. (2000). *Verbmobil: Foundations of Speech-To-Speech Translation*. Berlin, Germany: Springer. doi:10.1007/978-3-662-04230-4

Waller, A. (2019). Telling Tales: Unlocking the Potential of AAC Technologies. *Int. J. Lang. Commun. Disord.* 54, 159–169. doi:10.1111/1460-6984.12449

Zurstrassen, B. (2015). "Inklusion durch Leichte Sprache? Eine kritische Einschätzung," in *Didaktik der inklusiven politischen Bildung* (Bonn, Germany: Bundeszentrale für politische Bildung), 129–138.

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Harbusch and Steinmetz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

METHODS
published: 12 January 2022
doi: 10.3389/fcomm.2021.689439

Using Adaptive Interaction to Simplify Caregiver's Communication with People with Dementia Who Cannot Speak

Arlene J. Astell[1,2,3]*, *Sarah Shoaran*[4] *and Maggie P. Ellis*[4]

[1] School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom, [2] KITE, University Health Network, Toronto, ON, Canada, [3] Department of Occupational Sciences and Occupational Therapy and Department of Psychiatry, University of Toronto, Toronto, ON, Canada, [4] School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, United Kingdom

Caregivers find it difficult to interact with people with dementia who have lost the capacity for speech. Adaptive Interaction is a simplified approach that uses the nonverbal fundamentals of communication to connect with people who can no longer speak. Here we present Adaptive Interaction as a method for equipping caregivers with these nonverbal skills to increase communication with the people they care for. Six caregivers were each paired with one individual with dementia and trained in Adaptive Interaction. After receiving training in Adaptive Interaction, caregivers identified more communicative behaviours in their interactions partners and engaged in more frequent positive social behaviours and meaningful actions during interactions. These findings suggest that it is possible to equip staff to use simplified communication based on nonverbal fundamentals to connect with people with dementia who can no longer speak.

Keywords: dementia, nonverbal, interaction, social isolation, training, communication

OPEN ACCESS

Edited by:
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

Reviewed by:
Eliseo Diez-Itza,
University of Oviedo, Spain
Sigal Uziel-Karl,
Achva Academic College, Israel

***Correspondence:**
Arlene J. Astell
a.astell@reading.ac.uk

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

Received: *31 March 2021*
Accepted: *15 October 2021*
Published: *12 January 2022*

Citation:
Astell AJ, Shoaran S and Ellis MP (2022) Using Adaptive Interaction to Simplify Caregiver's Communication with People with Dementia Who Cannot Speak. Front. Commun. 6:689439. doi: 10.3389/fcomm.2021.689439

INTRODUCTION

Dementia is a neurodegenerative condition characterised by progressive impairments in cognition that ultimately impact all aspects of an individual's life. There are multiple causes of dementia, of which Alzheimer's disease is the most common and age the biggest risk factor (Alzheimer's Association, 2021). As dementia advances, communication is increasingly affected until people retain little or no speech, relying instead on sounds or vocalizations (Frank, 1994). It has long been established that loss of speech can increase communication difficulties between individuals with dementia and their caregivers, such that the only interactions that take place are during personal care (Bowie and Mountain, 1993).

Lack of meaningful interactions and communication contributes to the increasing social isolation and exclusion of people living with dementia as their illness progresses (Bunn, et al., 2018). Despite being unable to communicate verbally, individuals with dementia retain the "urge to communicate" (Ellis and Astell, 2004; Hughes et al., 2019). Kitwood (1997) as part of his influential work on person-centred dementia care, suggested that when people with dementia are no longer able to communicate with speech, the "fundamentals of communication" that typically act as precursors to the development of speech, can be used. These fundamentals include shared attention, turn taking, eye contact and using and understanding non-verbal communication (Hewitt, 2011; Intensive Interaction Institute, 2021). Using these non-speech-based behaviours to improve communication

between people with dementia, their families and caregivers could be crucial in improving their quality of life and well-being (Hughes, et al., 2019), particularly in the later stages of the illness.

Adaptive Interaction (AI: Ellis and Astell, 2011, 2017) is one such approach to communicating with people living with dementia who can no longer speak. AI was developed from Intensive Interaction (Hewett, 1996; Hewett, 2011; Nind, 1996), a teaching/learning approach developed in the UK in the 1980s for people considered "difficult to reach" (Firth et al., 2013). Specifically, these are children and adults with profound and multiple intellectual disabilities who do not develop speech. Intensive Interaction (II) emphasizes the use of these pre-verbal fundamentals of communication that are present in infant-parent interactions (these foundations of II are typically attributed to the work of Ephraim, 1982). II was developed for promoting communication between people born with profound and multiple intellectual disabilities and teachers or classroom assistants.

Developed initially within an educational context, the application and role of II has expanded beyond the classroom to day centres (Clegg et al., 2020), community living (Samuel et al., 2008) and residential (Firth et al., 2007) settings for people with severe developmental disabilities and autism spectrum disorder (ASD; Fraser, 2011). Additionally, training and use of II has extended to speech and language therapists, clinical psychologists (Firth et al., 2004), and care home staff (Firth, et al., 2007).

In considering the utility and application of II, (Firth, 2009) identified two distinct but related approaches that he termed a "social inclusion process model" and "developmental process model" respectively. He argued that the social inclusion model aims primarily at "inclusively responding to a learning disabled person's communication, however it is expressed. Such social inclusion is seen as being unconditional and at the most personal level" (page 45, Firth, 2009). This is exemplified by practitioners, such as (Caldwell and Jane Horwood 2007) who speak of "connecting" with individuals through a "shared language" and typically describes the increased awareness of the individual's (nonverbal) communication repertoire. The developmental process model regards II as a means to progressive acquisition of communicative skills (Firth, 2009). Here the focus is on II as "a process with the primary aims of increased sociability and communicative skill development through extended experience of Intensive Interaction" (page 45, ibid). Essentially the aim is to connect with individuals and support acquisition of a greater range of communication skills. From the perspective of dementia as a progressive irreversible neurocognitive disorder, the social process model fits very well with the needs of this population.

One of the key principles of Intensive Interaction is to view all behaviour—such as sounds, movements and facial expressions—from the nonverbal communicator as *intentionally communicative*. During interactions, communication partners use the fundamentals of communication to uncover aspects of their partner's communicative repertoire. This uncovering is based primarily on observation of the occurrence of sounds, movements, eye contact, etc, made by the nonverbal partner, to develop an understanding of the ways in which individuals communicate. Close observation means that communication partners can ""be with" this person as they are at present, using their initiatives, gestures, rhythms and sounds to respond in a way that has meaning for them" (Caldwell, 2011). As communicative behaviours are uncovered, the communication partner reflects these back through imitation or repeating a sound or rhythm, building up over time into "wordless conversations" (Barber, 2007). In accordance with the social process model, the intention of learning and adopting the language of the disadvantaged communicator is for the interaction to become shared and meaningful, rather than a list of activities that are carried out with the nonverbal individual (Caldwell, 2011).

The effectiveness of Intensive Interaction is examined through changes in communicative behaviours including elicitations of new behaviours, and increases or reductions in the type, frequency and/or duration of communicative behaviour by both partners. For example, increased contingent smiling (e.g., Argyropoulou and Papoudi, 2012), increased levels of eye contact (e.g., Fraser, 2011) and improved levels of joint attention (e.g., Kellett, 2005). To identify these behaviours, microanalytic observation and coding of video recordings are the major tools for training and implementation of Intensive Interaction (2; Anderson, 2006; Firth, 2021).

Using these methods Watson and Fisher (1997) found increased social behaviours such as "smiling", "vocalizing", "initiation" and "direct eye contact" when compared to other classroom activities where Intensive Interaction was not used, leading them to conclude that II was effective in enhancing the level of engagement. In addition to increased communicative behaviours, Nind (2006) found improved ability to maintain and initiate social contact among people with severe developmental disabilities and ritualistic behaviour using a combination of real-time observation, video analysis and informant measures.

In terms of implementation of II as a communication approach in services for people with severe intellectual disabilities and ASD, there are overall positive findings for clients and staff (Berridge and Hutchinson, 2020). For example, staff can gain increased satisfaction in their relationships with clients, residents or service users (Clegg, et al., 2020). However, some challenges with adoption of II have been encountered among care staff, including reluctance to fully engage with the approach (Firth, 2007). In their evaluation of a large-scale II implementation, Clegg et al. (2020) identified "personal discordance, doubt and discomfort" and the importance of organizational support for "implementation at all levels". These findings have direct relevance for training staff and implementing a non-speech-based approach for people living with advanced dementia.

In adapting II for dementia, Ellis and Astell (2008) conducted a case study with an 81 year old lady—Edie—a care home resident for 5 years. As Edie had gradually ceased being able to use speech, family members and care staff had found communicative interactions with her increasingly difficult. Using the fundamentals of communication, Ellis and Astell uncovered Edie's communication repertoire which included a range of

communicative behaviours. In particular, Edie's use of vocalizations stood out as a distinctive behaviour. By using imitation to reflect this and other behaviours back to Edie, the communication partner was able to engage in an initial interaction lasting 10 min, in which both partners initiated and took turns (Ellis and Astell, 2008). By uncovering Edie's communicative repertoire, Adaptive Interactions was able to enhance quality of communication, demonstrated by developing a "shared language" and "meaningful interactions". A similar study with three individuals living in long-term care reported signs of engagement and interaction and highlighted the opportunity for active participation often denied to bedbound people such as Edie (Harris and Wolverson, 2014).

Adaptive Interaction (AI) was developed specifically to improve communication when people living with dementia can no longer speak (Ellis and Astell, 2011). In a mixed baseline study with five individuals, Adaptive Interaction (AI) techniques were compared with standard interaction (SI) approaches (Ellis and Astell, 2017). In the SI sessions, the communication partner used speech and during the AI sessions used nonverbal channels based on the fundamentals of communication. Each interaction session between the individual with dementia and their conversation partner was video-recorded to allow for a comparison of the two communicative methods. Results of microanalysis indicated an increase in the frequency and duration of communicative behaviours in AI sessions compared to SI sessions. For example, there was an increase in "smiling", "imitation" and "vocalizations" in AI sessions when compared with SI sessions, in which there were longer durations of "neutral" facial expressions, which is something that could indicate lack of engagement or emotion. Crucially, this study confirmed that each person with dementia possessed their own individual communicative repertoire, and that AI methods were able to uncover these repertoires, which could then be used to build up communication with a conversation partner (Ellis and Astell, 2017).

Communication between caregivers and people living with dementia, can be improved through training along with awareness raising and support to improve their wellbeing and quality of life (Surr, et al., 2017). In a feasibility study Ellis and Astell (2011) extended the AI approach to caregivers in a nursing home in order to assess the potential for AI in improving communication between caregivers and people with dementia. The researchers taught three caregivers theoretical and practical elements of the fundamentals of communication, how communication develops, and aspects of verbal and non-verbal behaviours (Ellis and Astell, 2011) over the course of four sessions. Between training sessions, the caregivers were requested to practice a specific AI task each week, such as imitation or focus of eye gaze. Each interaction was observed and video-recorded to understand the impact of the training and support learning. At follow-up, the caregivers reported that they felt "more equipped" to identify communicative behaviours in people with dementia (Ellis and Astell, 2011). This study demonstrated the potential for caregivers to use a simplified nonverbal approach to enhance their communication with people living with dementia.

The present study builds on the previous feasibility study to develop AI as a method to equip caregivers to identify the communicative repertoires of their communication partners. The effectiveness will be explored by examining the impact on the communication behaviour of both parties in the interactions. To achieve this, the study compares the communication repertoires of six caregivers paired with six people with advanced dementia during a baseline interaction and three subsequent interactions recorded whilst the caregivers receive training in AI. The communication behaviour for both partners and their dyadic behaviour is observed for changes during the course of the training that indicate increased engagement and enhanced communication.

The following questions guide the study:

1) Are there changes in the types and/or frequency and/or duration of communicative behaviours across training sessions?
2) Are any new behaviours elicited during training sessions?
3) In order to explore the individuality of each person's communicative repertoire, are there behaviours that differ between pairs of participants in response to AI techniques?

METHODS

Participants

Six people with dementia (two male) were recruited plus six care staff (two male). The six people with dementia were aged between 78–92 years, mean age 84.3 and all had either no use of words or occasional use of single words. All participants came from the same care home in the South of England. Relatives of the six people with dementia gave their consent for them to participate and for their interactions with care staff to be video recorded in accordance with the (Mental Capacity Act 2005) (England and Wales). Each participating caregiver was matched to one individual with dementia who was to be their communication partner throughout the study. The Assistant Manager from the care home also attended the training sessions. The study received ethical approval from the NRES Committee London—Camberwell, St. Giles (Ref: 12/LO/0818) and also the University Teaching and Research Ethics committee (UTREC) at the University of St Andrews.

Study Design

A descriptive-qualitative study was conducted in which Adaptive Interaction training was delivered to staff in the care home over a 3-day period. Interactions between the participating staff and their communication partners were recorded prior to training (baseline condition) and in three further sessions during the training. A microanalysis of communicative behaviours was carried out on the video recordings using the Observer 10.5XT software programme by an independent rate who was not part of the training (second author).

TABLE 1 | Coding categories and their operational definitions for both caregiver and person with dementia (PWD).

Category & Behaviour	Operational definition	Colour
Eye gaze		
Eyes closed	Eyes closed	
Occluded	i.e. eyes out of view of camera	
Eyes/face	Gaze focused on partner's eyes/face	
Partner's body	Gaze focused on partner's body	
Elsewhere	Gaze elsewhere/not at partner	
Facial Expression		
Neutral	No detectable emotion	
Smiling	Smiling	
Frowning	Frowning	
Surprise	Surprised expression	
Other	All other emotional expressions	
Unknown	i.e. face out of view of camera	
Joint movement		
Initiated by caregiver	Caregiver initiates joint movement, e.g. holding partner's hand and moving it towards caregiver	
Initiated by PWD	Person with dementia initiates joint movement	
Vocalization		
Silence	No sound	
Laughter	Laughing	
Speech	Articulation of words	
Vocalization	Sounds/vocalizing	
Other	Any other sound, e.g. cough, heavy breathing	

(Continued on following page)

TABLE 1 | (*Continued*) Coding categories and their operational definitions for both caregiver and person with dementia (PWD).

Physical contact		
Physical contact	One person initiates contact with the other	
No physical contact	No physical contact occurs	
Unknown contact	i.e. out of view of camera	
Body position		
Side to face	One sits to the side of the other's face	
Face to face	Caregiver and PWD face to face	
Side by side	Caregiver and PWD side by side	
Other position	e.g., if one briefly leaves the interaction	
Gestures		
Pointing	Pointing with hand	
Nodding	Nodding head	
Shaking head	Shaking head	
Other head	Making any other head gestures/movement	
Other hand/arm	Making any other hand/arm gestures/movements	
Moving closer	Moving hand/head/body closer to partner	
Other foot/leg	Making a foot or leg gesture/movement	
Moving away	Moving hand/head/body away from partner	
Facial Gesture	Making a facial gesture/movement/expression	
Verbal	Imitates partner verbally	
Non-verbal	Imitates partner non-verbally	

AI Training

The training was delivered over 3 days—2 days together then one further day a month later. The training was interactive and designed to engage participants in a range of activities in pairs and small groups as well as whole group sessions. The training used a range of methods including Powerpoint™ presentations, videos, case studies and peer teaching. Each participant received a folder of teaching materials and guidance in the use of reflective learning techniques. Each training day included practical activities for participants to carry out either before or during the sessions.

The initial session included exploring the difficulties of communicating with people with dementia and sharing their experiences of recording videos. The fundamentals of communication were also introduced. In the second session, Adaptive Interaction was introduced with a video example. The initial baseline videos were then examined in turn to identify one communicative behaviour for each individual with

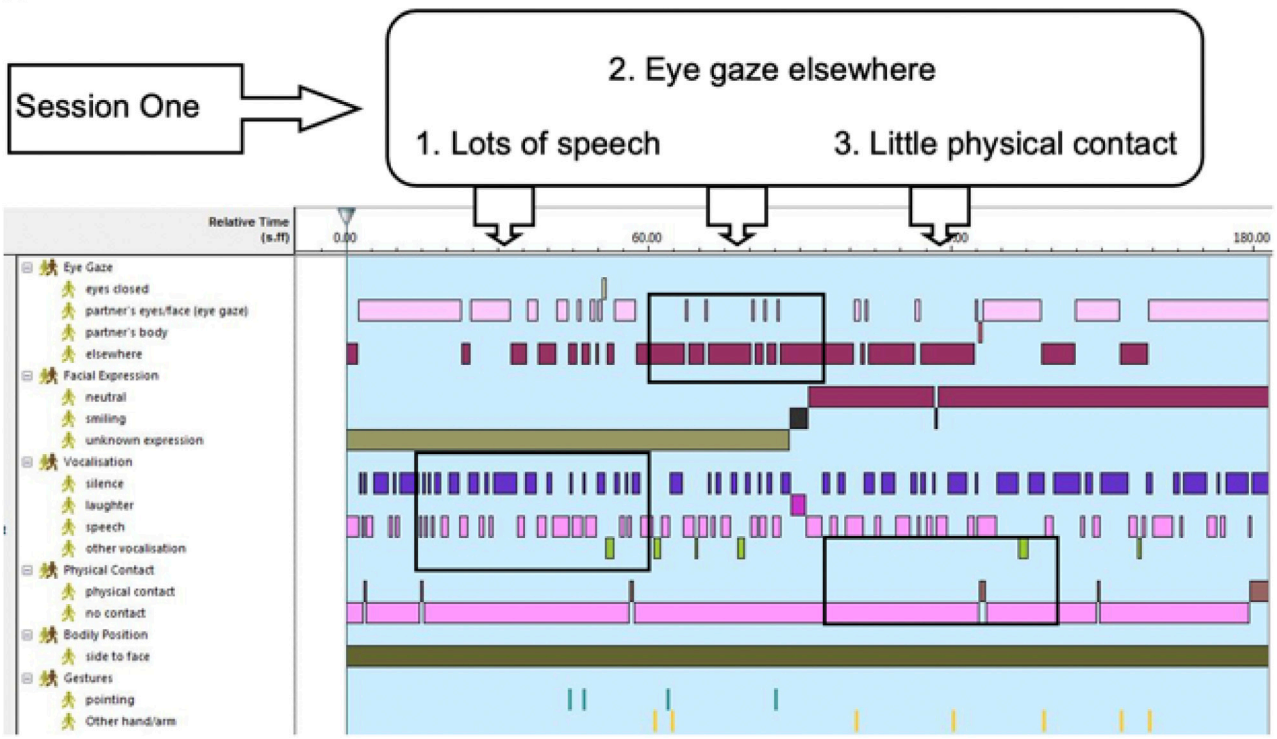

FIGURE 1A | (A) Elena's Baseline (Session 1) communication. **(B)** Renee's Baseline (Session 1) communication. **(C)** Elena's First AI (Session 2) communication. **(D)** Renee's First AI (Session 2) communication. **(E)** Elena's Second AI (Session 3) communication. **(F)** Renee's Second AI (Session 3) communication. **(G)** Elena's Third AI (Session 4) communication. **(H)** Renee's Third AI (Session 4) communication.

dementia which the caregivers then attempted to use in a new recorded interaction with their conversation partner. Day 2 commenced with examining the videoclips from the day before with the caregivers sharing their experiences of their first attempt to use Adaptive Interaction. They applied the fundamentals of communication to the new clips and then discussed how to use what they observed to further develop the interaction. They then conducted another interaction using a further identified behaviour. After recording this interaction, the caregivers discussed more ways they could develop the interactions. They also further discussed their continued experiences of using videos. In the final part of Day 2 the caregivers applied the fundamentals to the latest videoclip to plan how to extend the interaction. They also set goals for the following month by developing an interaction package for each of the six individuals with dementia. Before the follow up session each caregiver was asked to produce a fourth recording of interaction with their conversation partner. At the follow up day the first part focused on reviewing the video-recordings and consolidating the plan for each individual resident. The rest of the day was spent developing a strategy to assess communication in the home to identify residents who could benefit from AI and discussing how to disseminate the AI technique to other caregivers and family members of the nonverbal residents.

Procedure

The six staff were provided with information about the study and a consent form to participate and consent to be videorecorded. They each chose a resident with dementia to work with and were asked to work in pairs to record each other interacting with their resident for 5–10 min. These baseline recordings were reviewed as part of the first AI training activity to identify a single communicative behaviour that could be used in the next session as the basis of nonverbal communication.

The training comprised teaching about the principals of AI and the fundamentals of communication as described above interspersed with practical experience using AI over the three training days. Each day lasted approximately 7 hours resulting in 21 h of training. The final session also included an exercise looking at how to maintain AI for the residents with dementia who cannot speak and identifying potential barriers to be overcome.

Coding Scheme

A coding scheme was developed based on the fundamentals of communication to microanalyse the communicative behaviours of individuals with dementia and their caregivers from the video-recordings. The coding scheme was based on the elements of verbal and nonverbal behaviour found to comprise the communication repertoires of people with advanced dementia

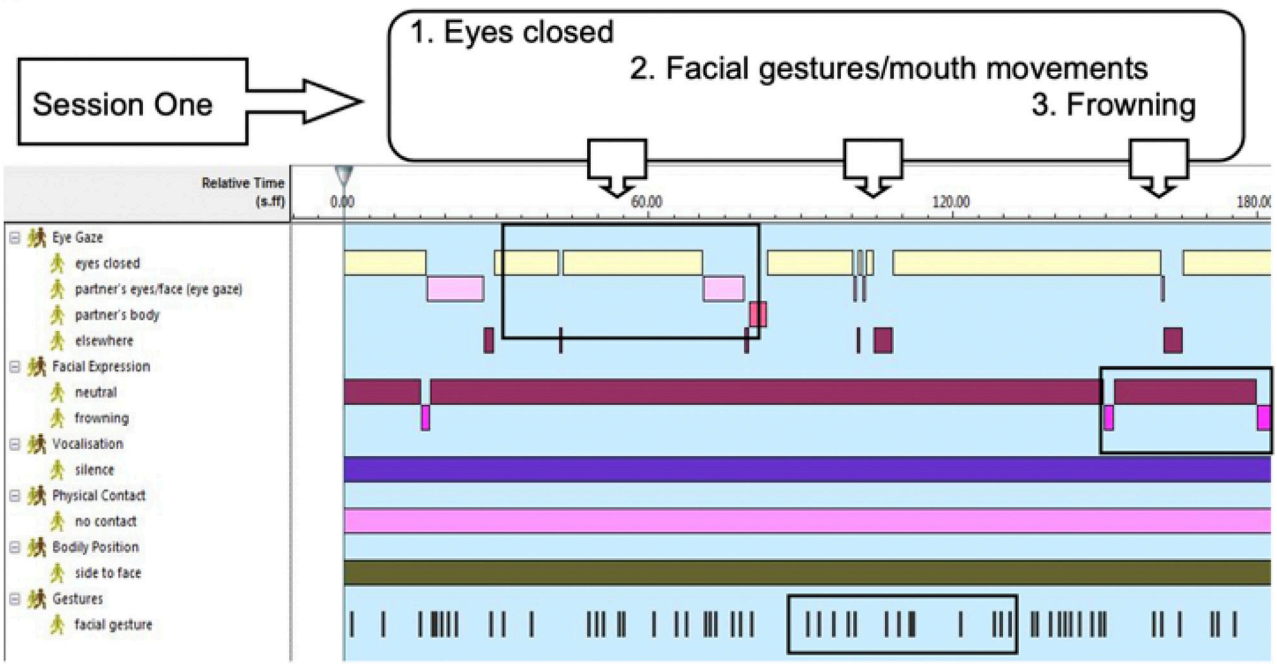

FIGURE 1B | continued

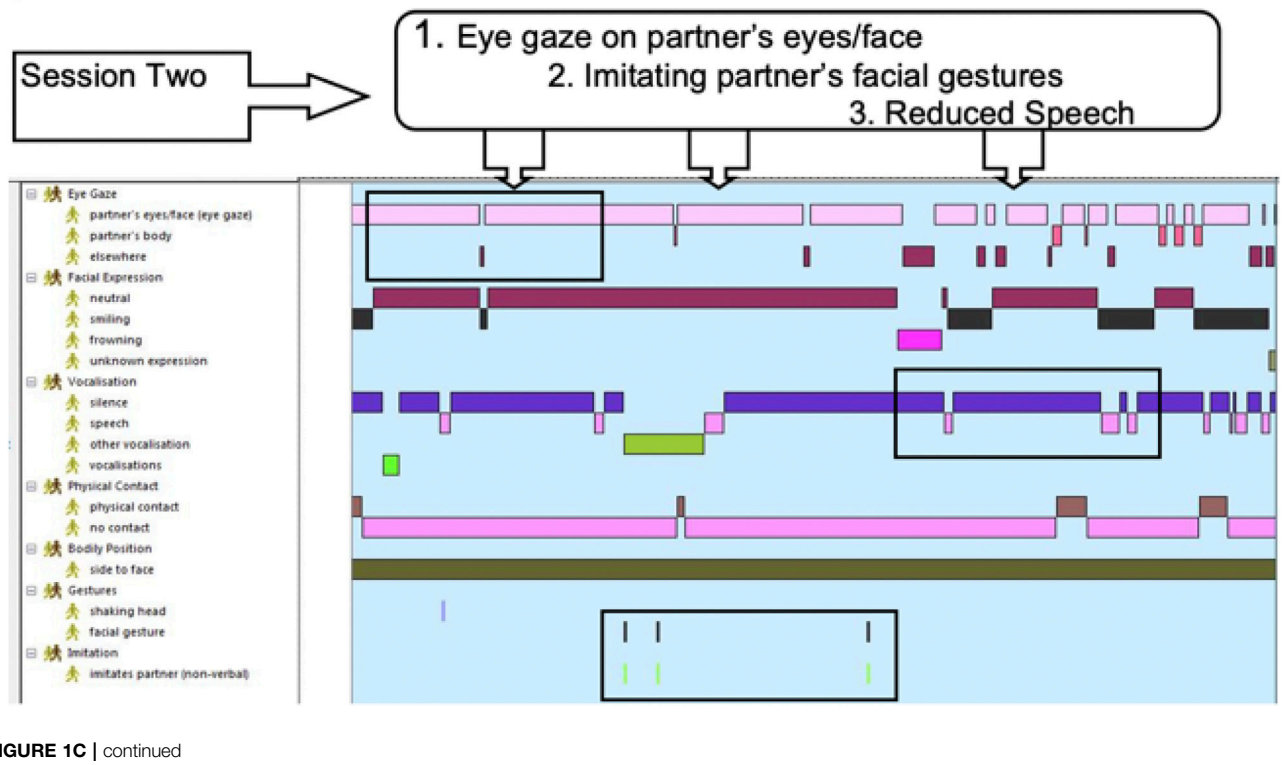

FIGURE 1C | continued

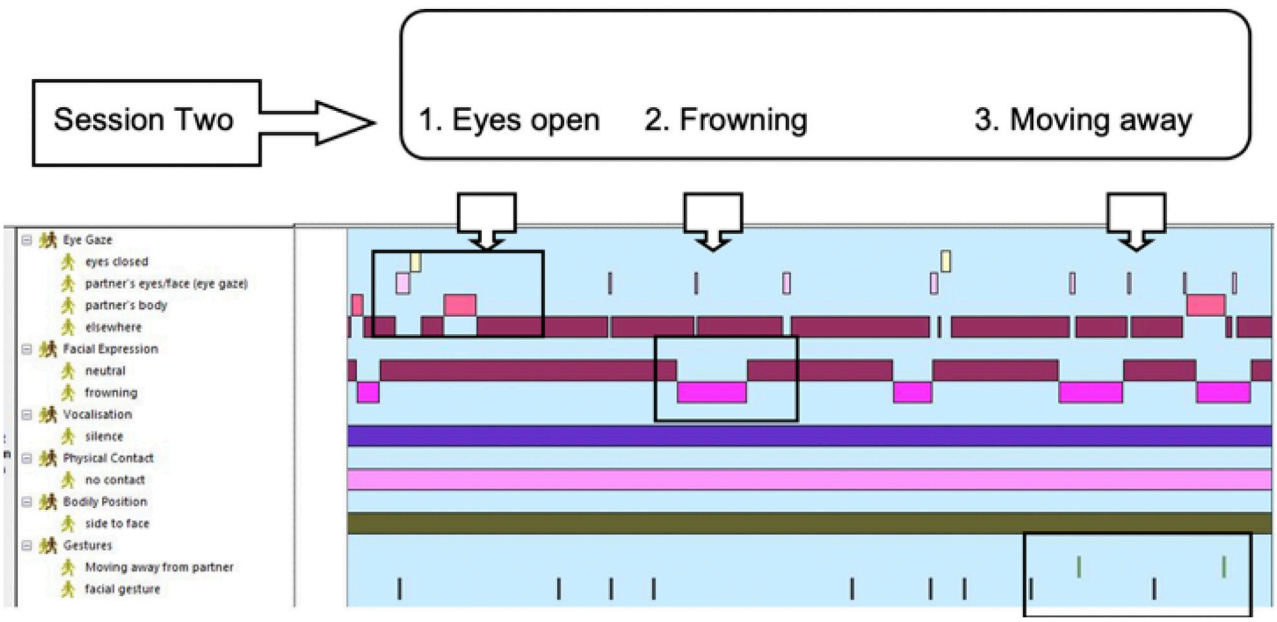

FIGURE 1D | continued

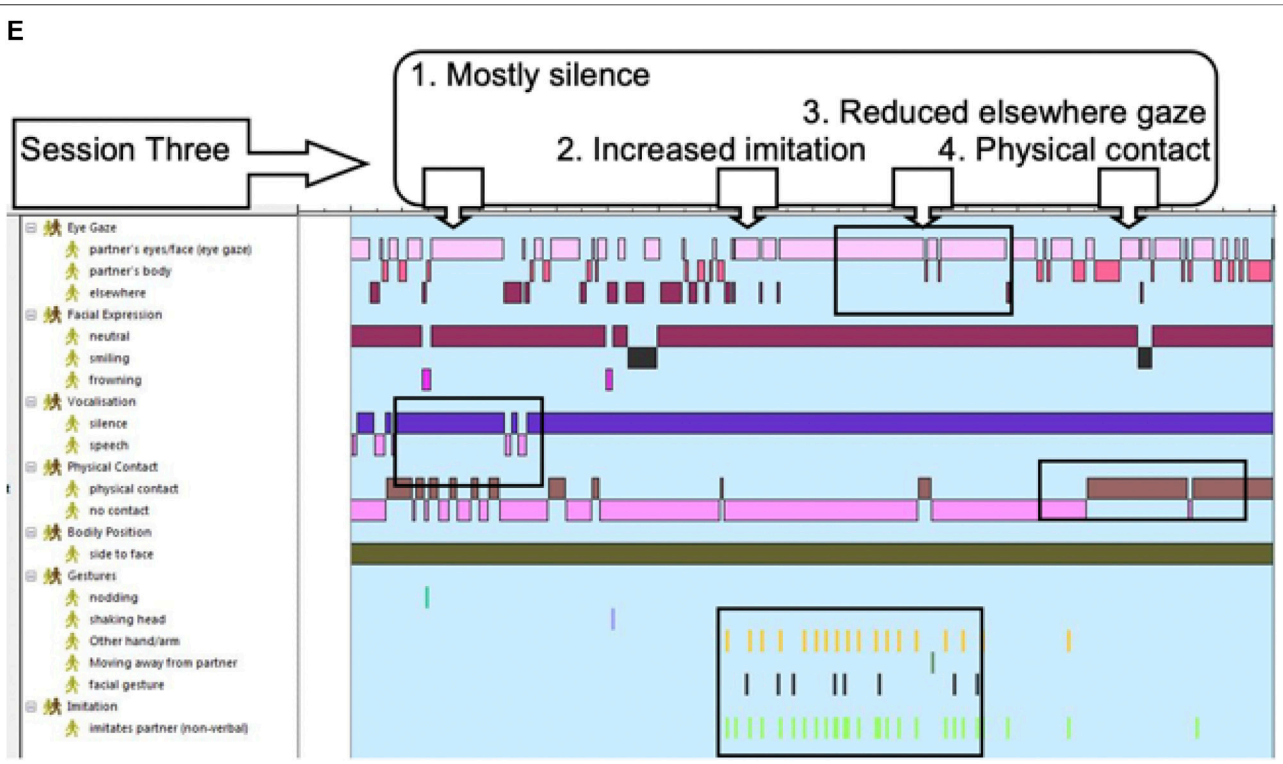

FIGURE 1E | continued

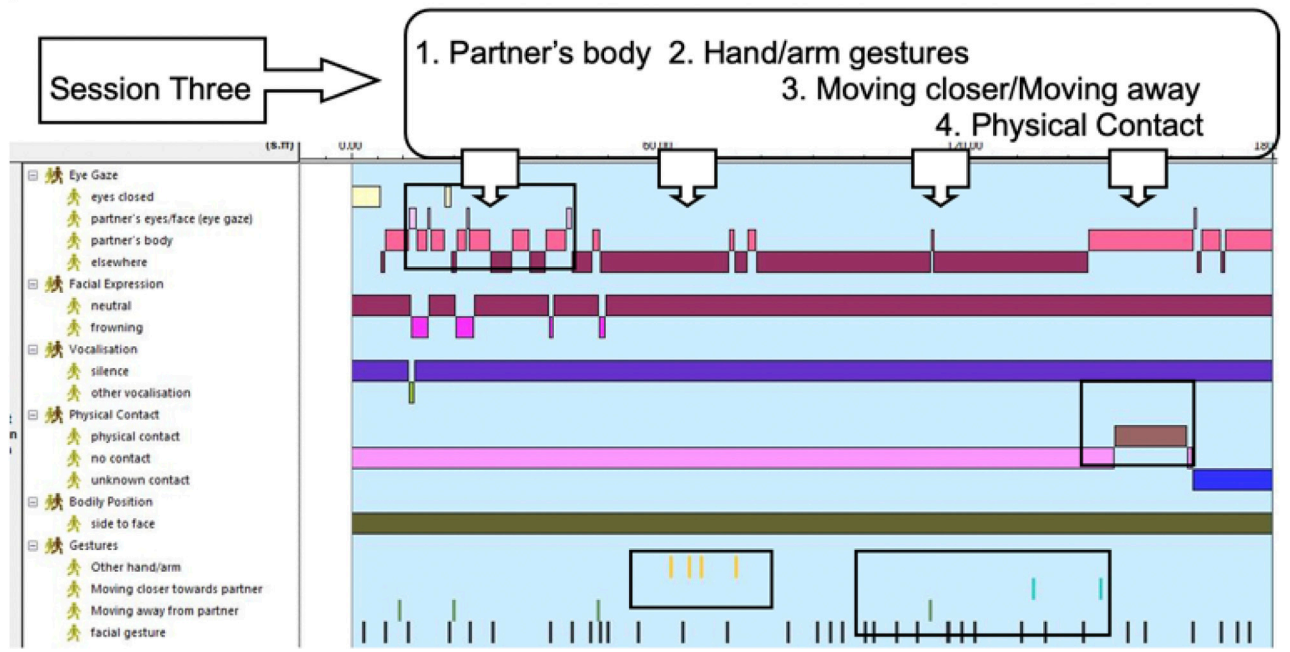

FIGURE 1F | continued

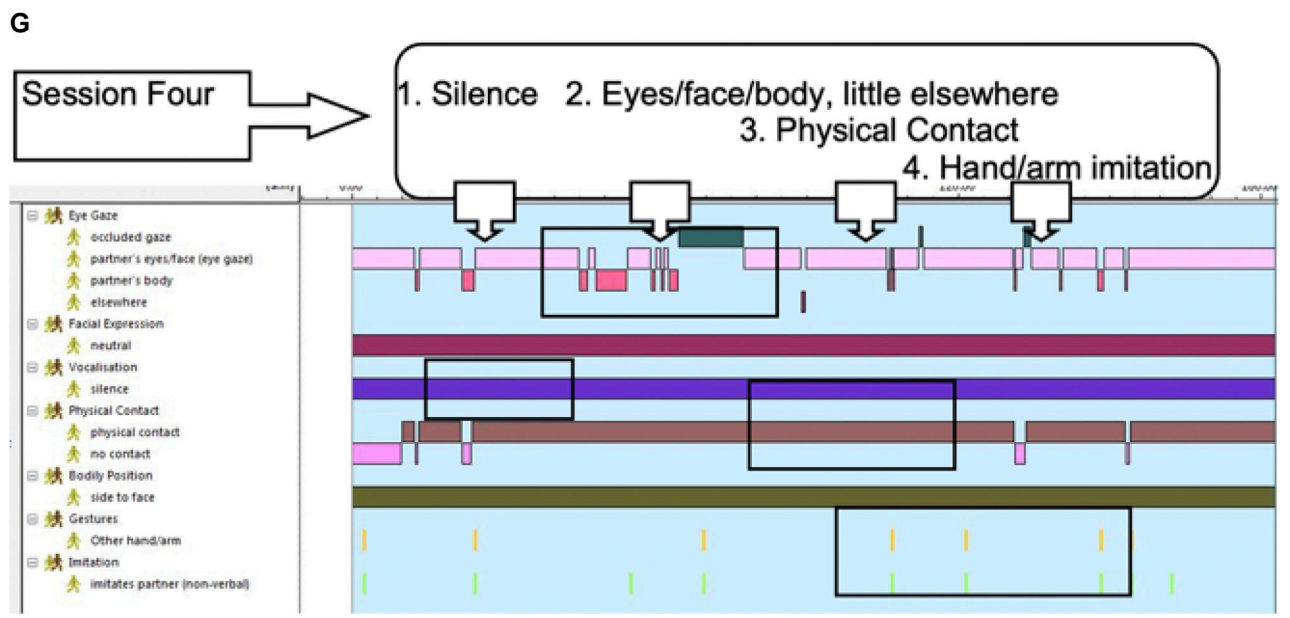

FIGURE 1G | continued

in the previous studies (Astell and Ellis, 2008; Ellis and Astell, 2017). This consisted of thirty-seven behaviours, which were grouped under eight overarching categories, consisting of verbal and nonverbal behaviours: Eye gaze, facial expression, vocalization, physical contact, joint movement, imitation, gestures and body position. The communicative behaviours of individuals with dementia and their caregivers were coded across all four sessions using the following operational definitions (**Table 1**):

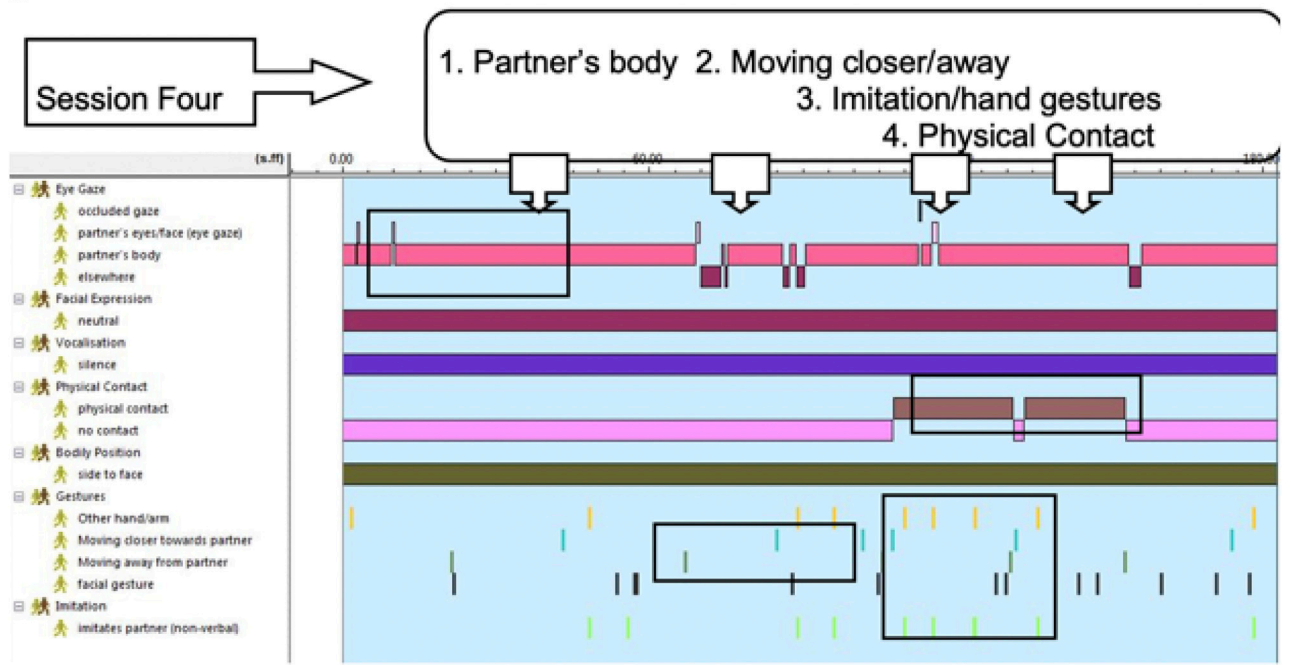

FIGURE 1H | continued

FIGURE 2A | (A) Diane's Baseline (Session 1) communication. **(B)** Derek's Baseline (Session 1) communication. **(C)** Diane's Third AI (Session 4) communication. **(D)** Derek's Third AI (Session 4) communication.

Analysis

Each dyad of participants was recorded across four sessions resulting in twenty-four video recordings, ranging from 72 s to 29 min. The Observer 10.5XT software programme was used to code and analyse communicative behaviours for the first 3 min of each session, where possible. The occurrence of behaviours was coded for frequency of occurrence or duration as appropriate. For example, hand movements were recorded for frequency of occurrence, whereas the duration of hand-holding was measured. Visualizations of the coded behaviour were generated from Observer to display the data.

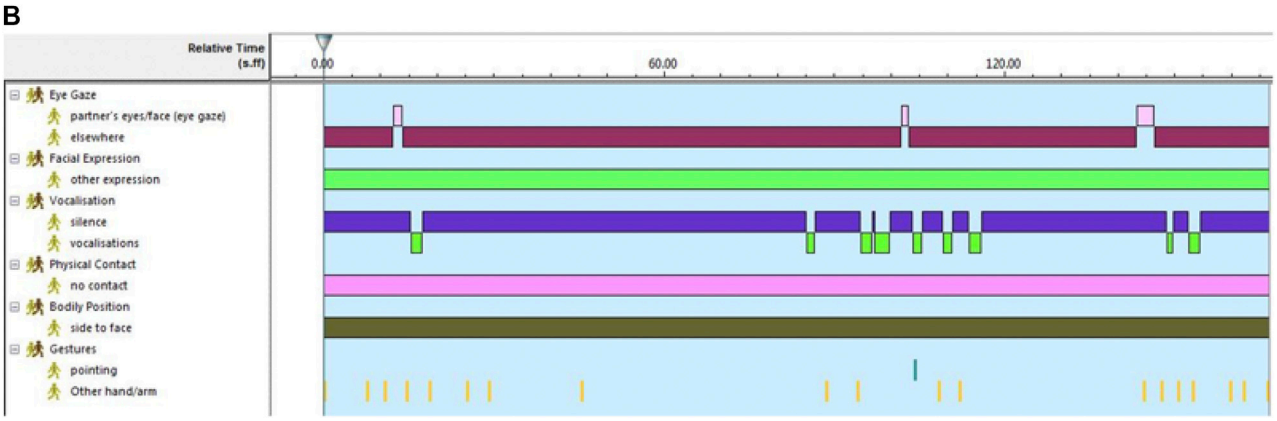

FIGURE 2B | continued

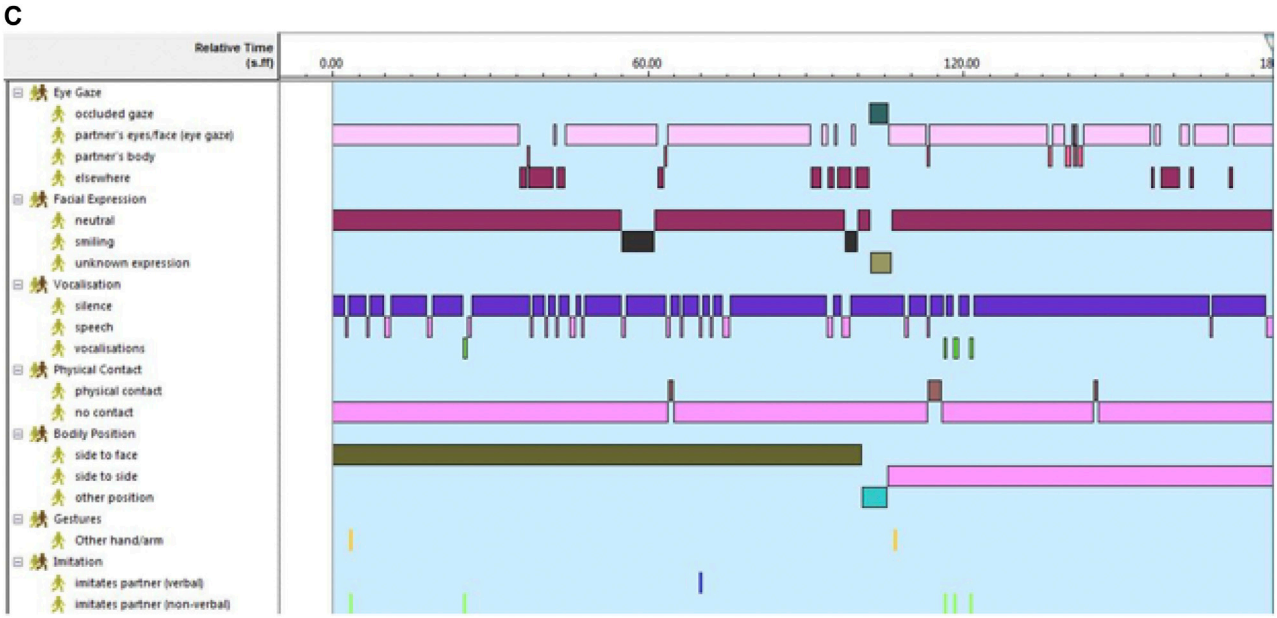

FIGURE 2C | continued

Inter-Rater Reliability

One rater, who was independent of the training, coded each of the twenty-four sessions for communicative behaviours. Three sessions (12%) were selected at random and coded by a second rater, thus allowing kappa values to be calculated. The Cohen's kappa values were as follows: Inter-rater video A (Diane and Derek session 3)—0.67; inter-rater video B (Graham and Marie session 1)—0.58; inter-rater video C (Karen and Jane session 2)—0.66. As such, the inter-rater reliability can be said to range from fair to moderate agreement (Landis and Koch, 1977). The areas of discrepancy were discussed by the two raters, to further clarify the behavioural coding definitions.

RESULTS

The primary purpose of this study was to explore the potential for training caregivers to use AI to identify, learn and use the communicative repertoires of individuals with dementia for whom speech is no longer functional. Changes in the types and/or number of communicative behaviours across sessions for caregivers and people with dementia would indicate that the Adaptive Interaction method impacts communication.

The communicative behaviour of the dyads at baseline and over the course of practicing AI are displayed in visualizations generated by the Observer video analysis software. The communicative behaviours are grouped into eight domains

FIGURE 2D | continued

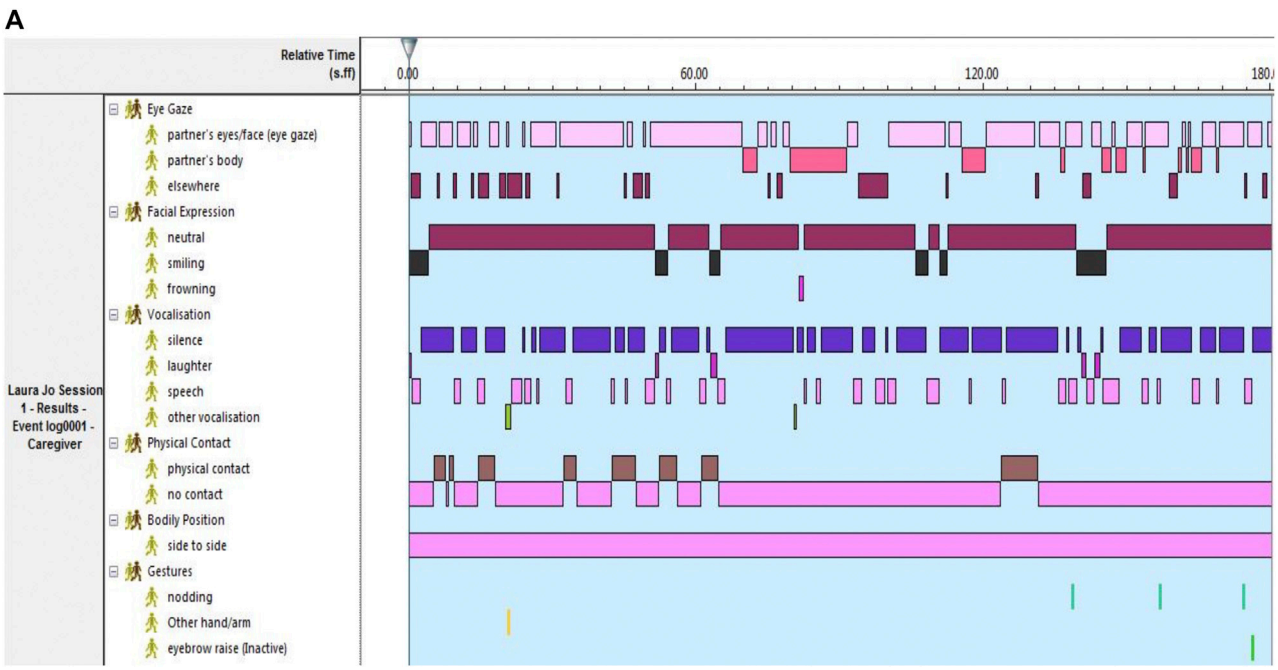

FIGURE 3A | (A) Lorraine's Baseline (Session 1) communication. **(B)** Jen's Baseline (Session 1) communication. **(C)** Lorraine's Third AI (Session 4) communication.

presented in the same order for all participants: eye gaze, facial expression, joint movement, vocalisation, physical contact, bodily position, gestures and imitation (**Table 1**). Each behaviour is represented by a colour on the visualization generated by the software. **Table 1** contains the key for the visualization labelling.

In the visualizations, the behavioural codes recorded in an event log are plotted horizontally against a time axis—each segment is the first 3 minutes of the interaction. The length of a horizontal bar demonstrates the duration of a "state event" behaviour (e.g., direction of eye gaze). The number of occurrences

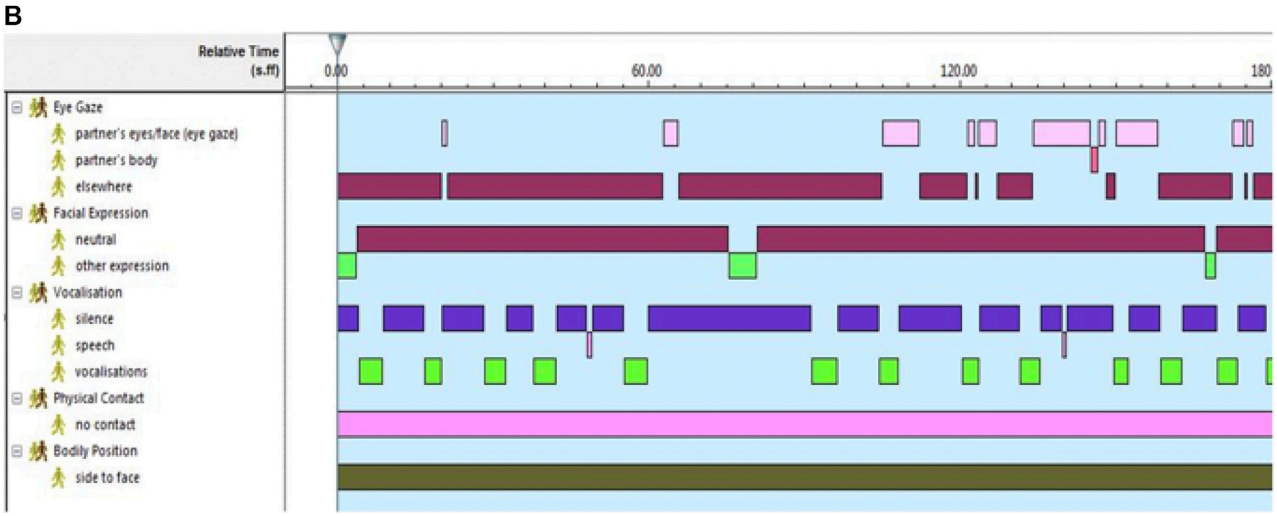

FIGURE 3B | continued

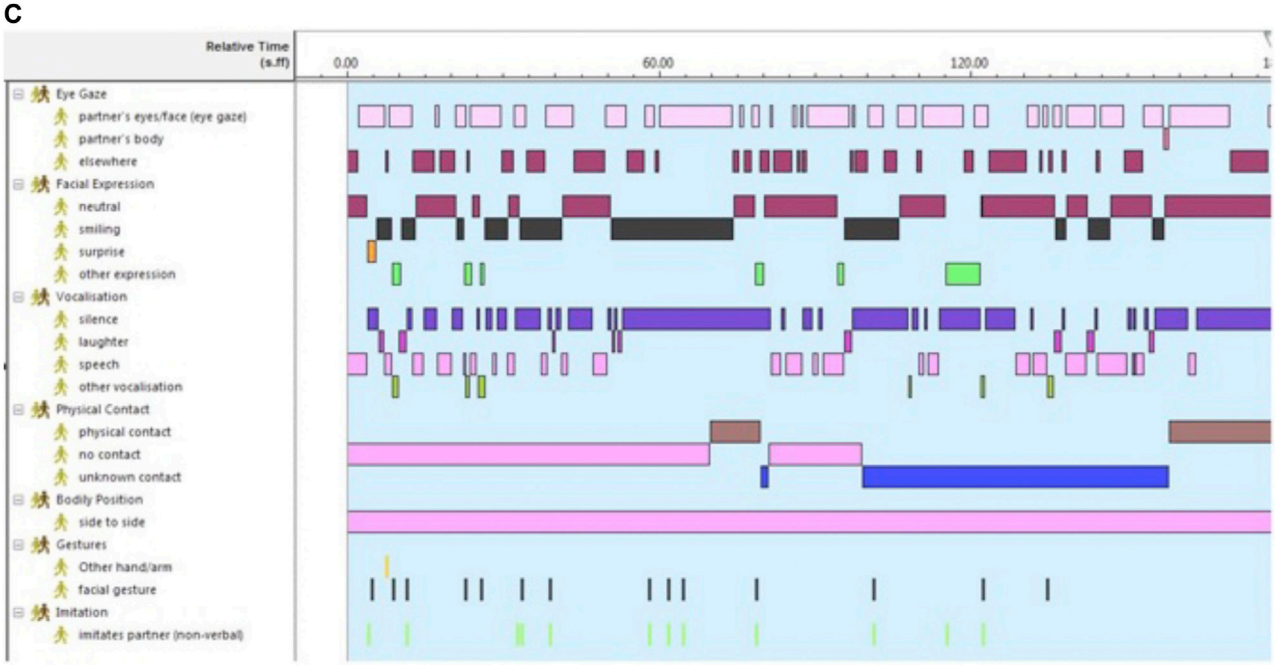

FIGURE 3C | continued

of a vertical segment demonstrates the frequency of a "point event" behaviour (e.g., smiling) (Zhou, et al., 2013).

Each individual repertoire comprises different combinations of communicative behaviours and therefore not all behaviours occur in each visualization. For example, **Figure 3B** Jen's communicative behaviours at baseline fall into five categories—eye gaze (mainly looking elsewhere), facial expression (mainly neutral), vocalization (alternating vocalizing with silence), physical contact (none), and bodily position (side to face). In the fourth interaction (third AI session; **Figure 3D**) gestures and imitation are also observed. Compared to the baseline visualization Jen and Lorraine are sitting side by side, Jen's eye gaze is split between looking at Lorraine and looking elsewhere, her facial expression is expanded to include smiling and looking surprised. She also makes physical contact with Lorraine, gestures and imitates Lorraine, none of which were observed in the baseline recording (**Figure 3B**).

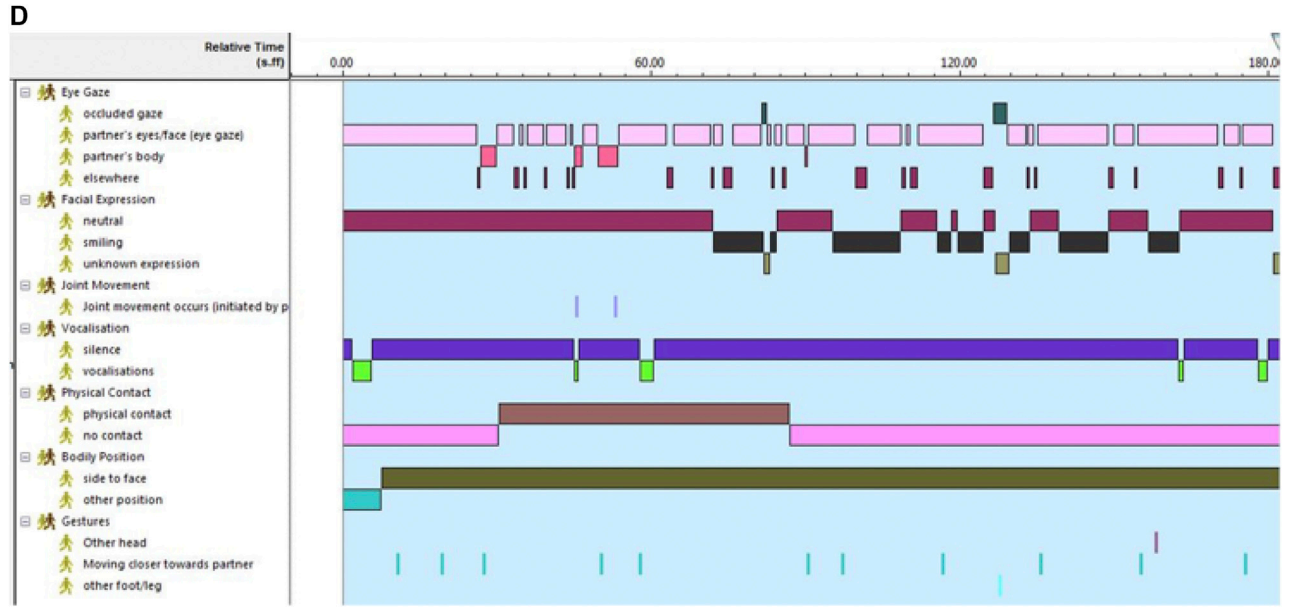

FIGURE 3D | continued

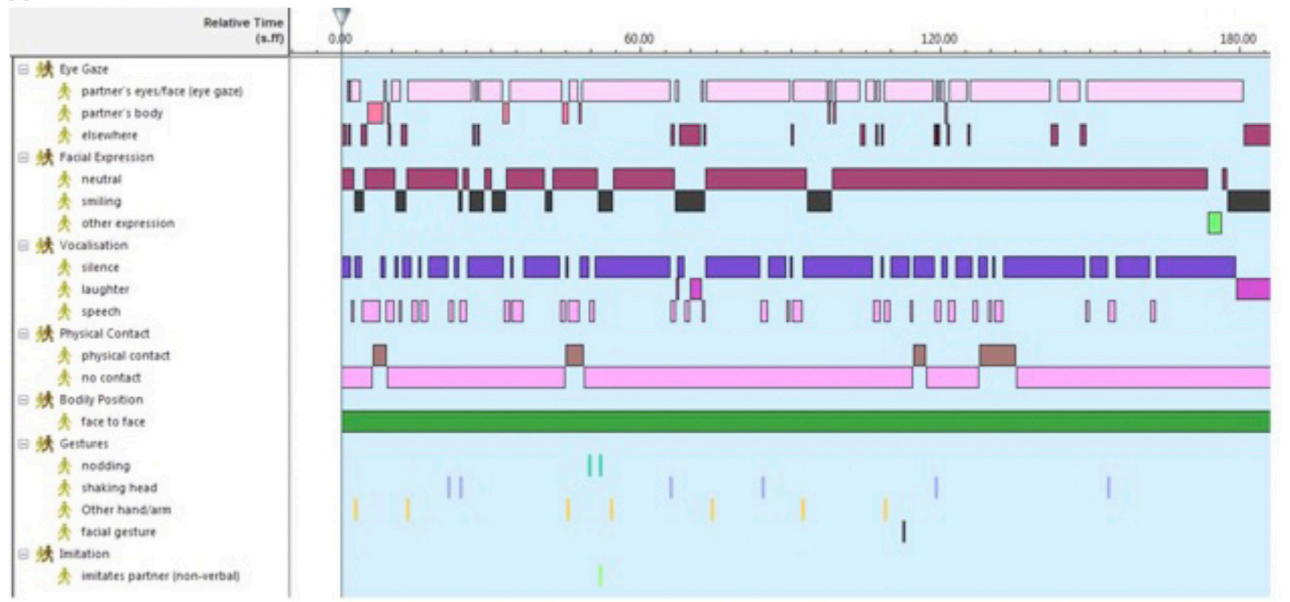

FIGURE 4A | (A) Karen's Baseline (Session 1) communication. **(B)** Jane's Baseline (Session 1) communication. **(C)** Karen's Third AI (Session 4) communication. **(D)** Jane's Third AI (Session 4) communication.

The pattern of change in communicative behaviour across AI sessions provides insight into the point at which the caregiver has identified key components of their partner's repertoire, and how the caregiver then introduces these components into their own communication approach. The findings are discussed in terms of changes in communicative behaviour from baseline across the AI sessions by looking at key behaviours in the dyads. The data are presented as six dyadic case studies where the first named of each pair the caregiver: Dyad 1: Elena and Renee, Dyad 2: Diane and Derek, Dyad 3: Lorraine and Jen, Dyad 4: Karen and Jane, Dyad 5:

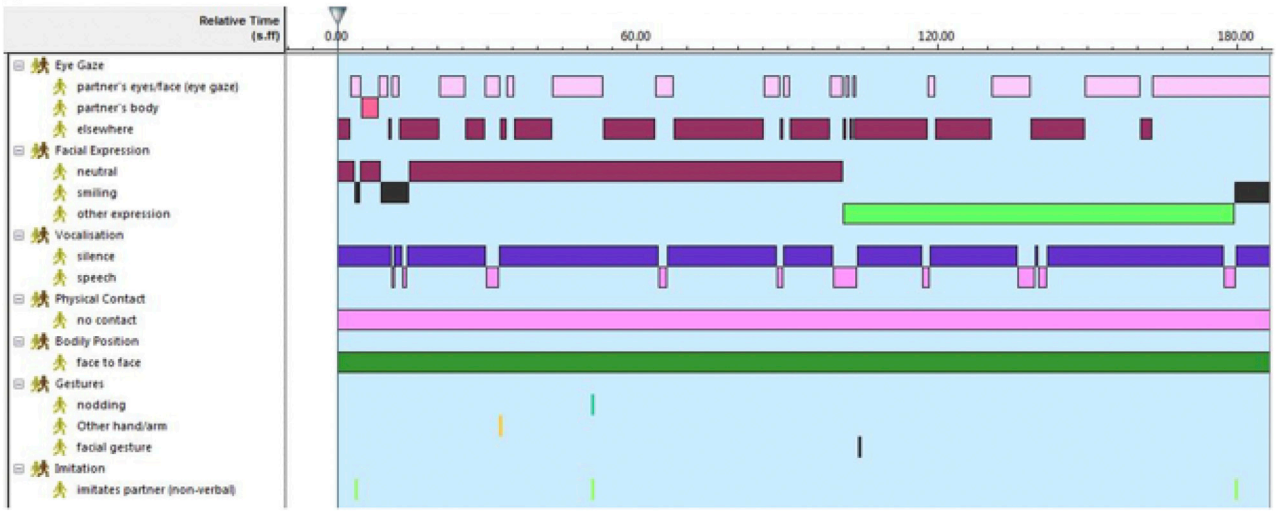

FIGURE 4B | continued

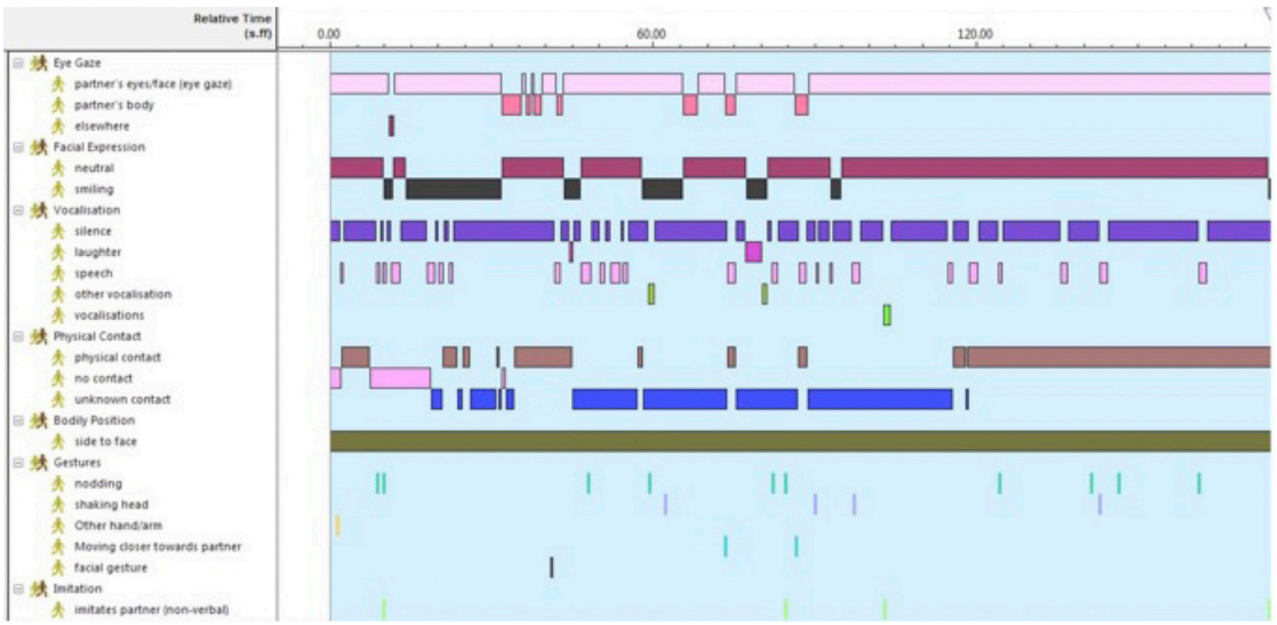

FIGURE 4C | continued

Jake and Ernie, and Dyad 6: Graham and Marie (all names have been changed to protect identities).

Dyad 1: Elena and Renee's Communicative Behaviour Across Sessions

In the baseline condition, much of Elena's communicative behaviour was verbal (44.3%), with speech occurring consistently throughout the session (**Figure 1A**). Elena's eye gaze shifted between Renee's eyes/face to looking elsewhere, with much of the interaction spent looking elsewhere (50%). There was some attempt at initiating physical contact with Renee, with the single brown line indicating the short duration of this contact (4.2%).

Figure 1B illustrates the pattern of Renee's communicative behaviour during the baseline interaction. Renee's eyes were closed for 80% of the videoclip, suggesting a lack of engagement with her caregiver partner. Renee frowned 3 times

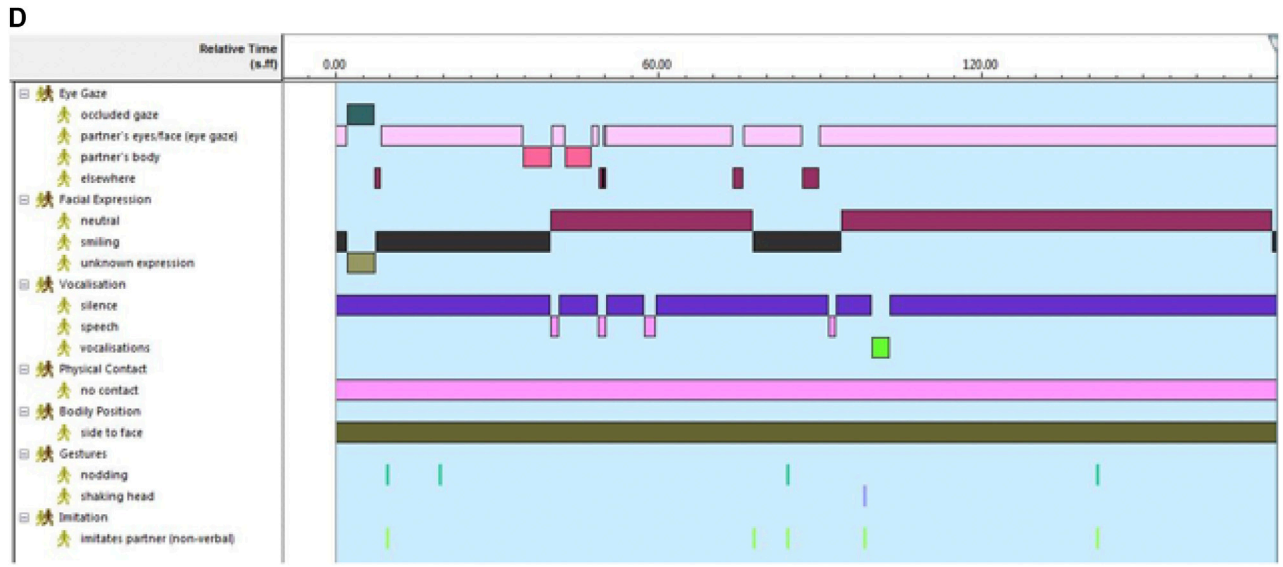

FIGURE 4D | continued

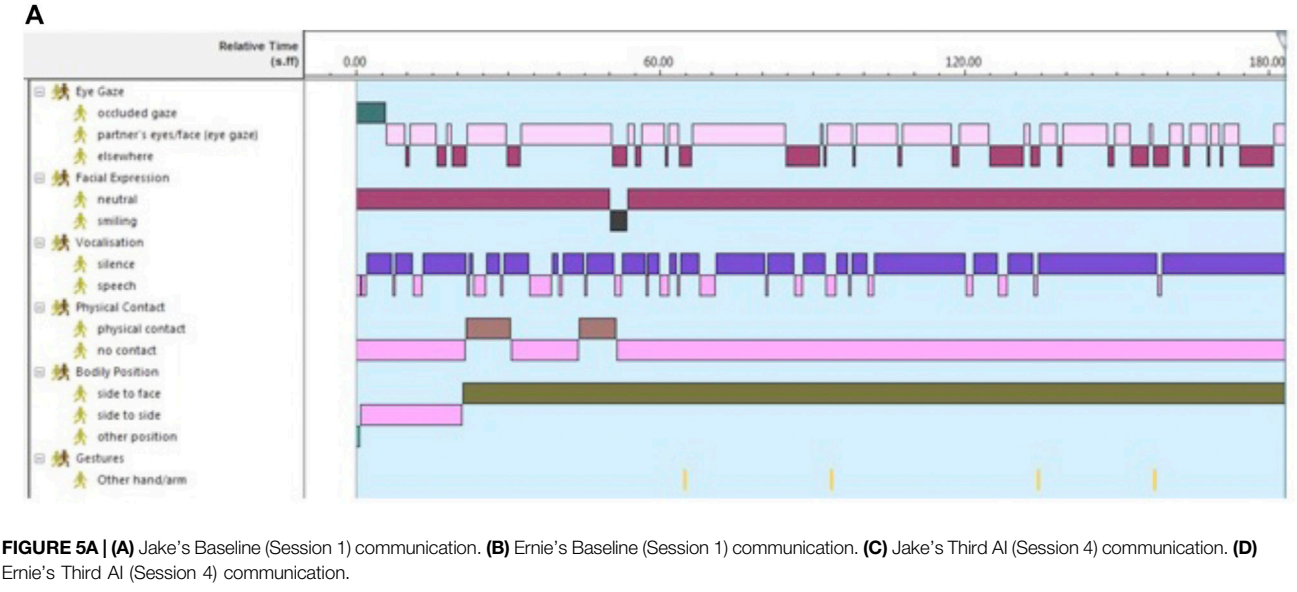

FIGURE 5A | (A) Jake's Baseline (Session 1) communication. **(B)** Ernie's Baseline (Session 1) communication. **(C)** Jake's Third AI (Session 4) communication. **(D)** Ernie's Third AI (Session 4) communication.

as a response to being touched by Elena, which accounted for 4% of the time. Renee's extensive use of facial gestures was evident in this session (55 times) and is something that Elena could potentially reflect back to Renee in subsequent sessions.

Figure 1C presents Elena's first session with Renee after basic training in Adaptive Interaction. Elena was encouraged to uncover one communicative behaviour in Renee's repertoire using the fundamentals of communication. Elena's eye gaze was focused on Renee for most of the video clip (84.8%), in comparison to the 50% elsewhere in the baseline interaction. Elena also attempted to imitate Renee's facial gestures three times. Furthermore, Elena's speech decreased from the baseline condition—down from 44 to 11.8%, suggesting a move towards focusing on the non-verbal elements of Renee's communicative repertoire.

Figure 1D illustrates the pattern of Renee's communicative behaviour during session 2. In a marked change to the baseline condition, Renee's eyes were closed only for a very short duration (2.3%) in this session, and she looked elsewhere for the majority of the interaction (83.4%). The frequency of Renee's facial gestures decreased in comparison to the baseline (9 times) however, the amount of frowning increased (27%). Renee

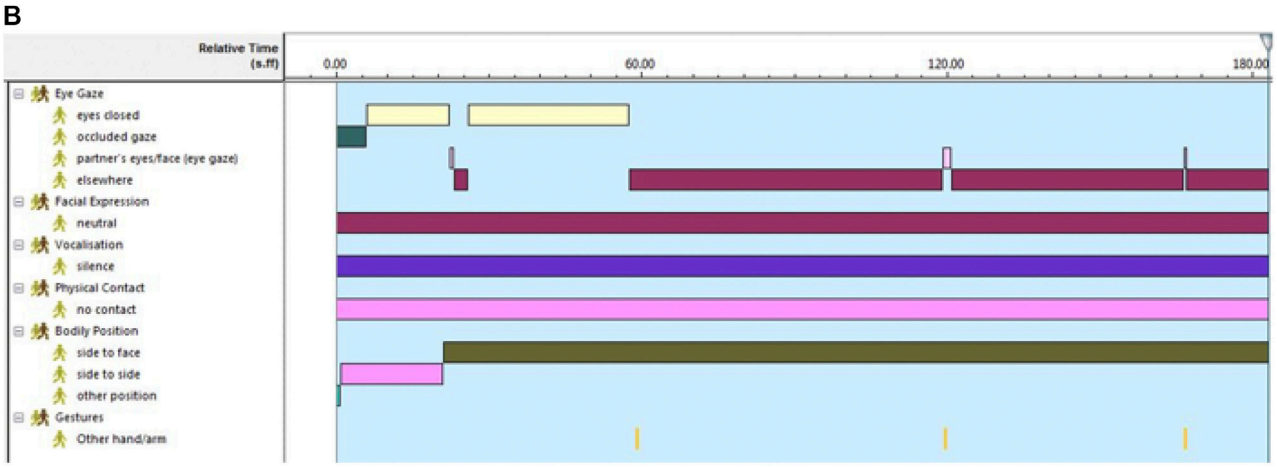

FIGURE 5B | continued

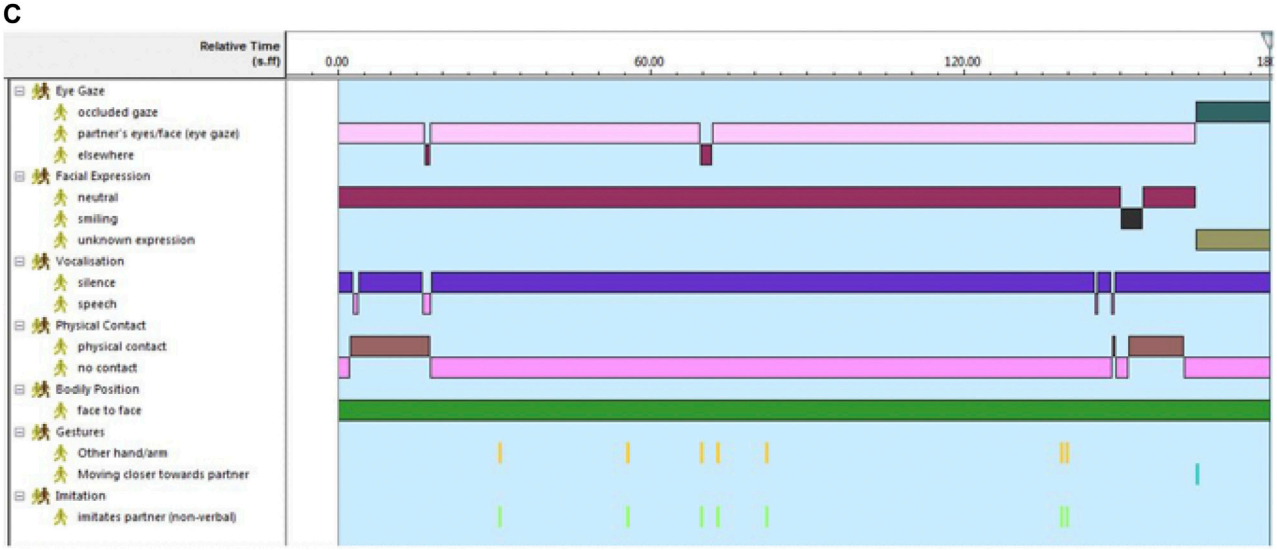

FIGURE 5C | continued

moved away from Elena on two occasions during this session in a response to Elena's initiation of physical contact.

Figure 1E illustrates the turning point of Elena's communicative behaviour towards and with Renee as it occurred during session 3. Elena was mostly silent in this session, as she spoke briefly only at the beginning (4%). The amount of time spent looking elsewhere decreased (14%), and the majority of the time was spent looking at Renee's eyes (67.7%) or her hand movements (18.4%). Physical contact was initiated towards the beginning of the session as Elena began to understand Renee's communicative behaviours, and what she liked and disliked. Elena had been using materials to brush against Renee's skin. After a while, she was able to understand that Renee did not enjoy being touched in this way and so, she ceased physical contact. Later in the video, Elena initiates physical contact with Renee by touching her skirt. The generally unbroken brown bar in **Figure 1E** indicates that Renee engaged in physical contact with Elena, so that physical contact occurred for 31.8% of the interaction. Elena practiced imitative behaviour during this session and spent much time imitating Renee's hand movements on the side of the chair and imitating Renee's finger movements on her skirt (27 times in total).

Figure 1F illustrates the pattern of Renee's communicative behaviour during session 3. A greater percentage of time is spent looking at Elena's body (34.3%), compared to previous sessions

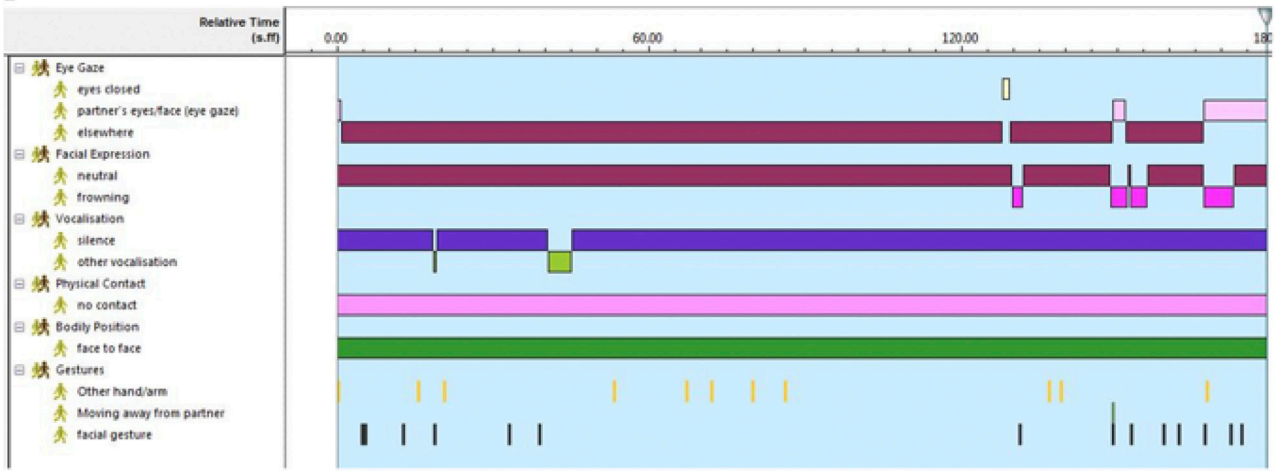

FIGURE 5D | continued

FIGURE 6A | (A) Graham's Baseline (Session 1) communication. **(B)** Marie's Baseline (Session 1) communication. **(C)** Graham's Third AI (Session 4) communication. **(D)** Marie's Third Ai (Session 4) communication.

pink line, **Figure 1F**). Renee introduced hand/arm movements into this session (4 times; yellow line), a behaviour that was not apparent during previous sessions. During this session, Renee moved away and towards Elena in a way that was similar to game-playing, suggesting a desire to interact. Towards the end of the session, Renee initiated physical contact with Elena for the first time (7.8%). This is a marked change from previous sessions.

Figure 1G illustrates Elena's communicative behaviour during the third AI interaction. At this point, Elena has been able to uncover several aspects of Renee's communicative repertoire. Elena was silent throughout this session (100%; purple bar) and focused entirely on the non-verbal elements of communication. There was only one occasion during the interaction in which Elena directed her eye gaze elsewhere (0.5%), suggesting close attentiveness to Renee. Physical contact occurred for the vast majority of this video (91.5%), suggesting that Elena had become more comfortable with Renee and also that Renee had become more comfortable with Elena. In previous videos, Renee would move away when Elena attempted physical contact. This suggests that a connection had developed between both caregiver and person with dementia. Elena also imitated Renee's hand movements during this video (9 times). This imitation differed to that in session 3 in that Elena focused more on Renee's responses as she imitated her. As we will see when we consider Renee's behaviour during this session, there is a great deal of turn-taking in terms of hand movement and imitative behaviour.

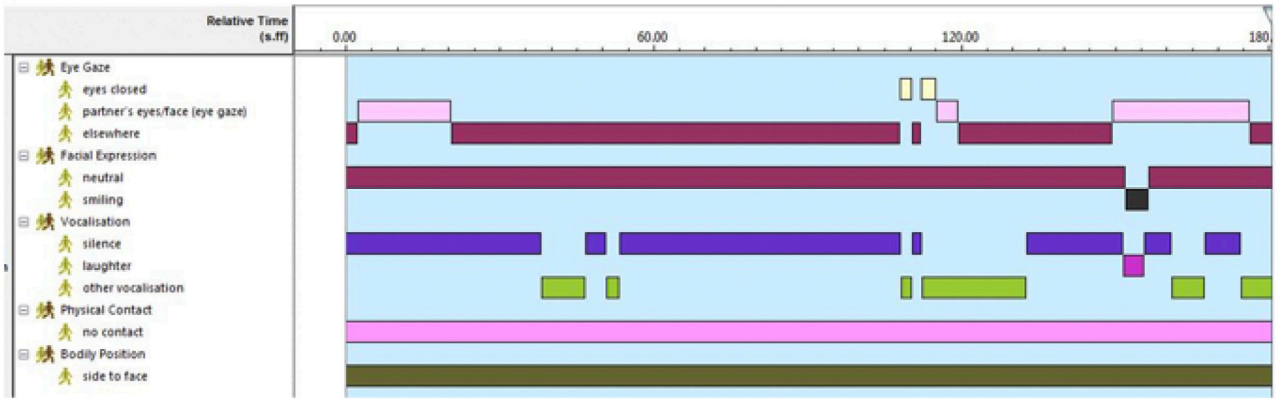

FIGURE 6B | continued

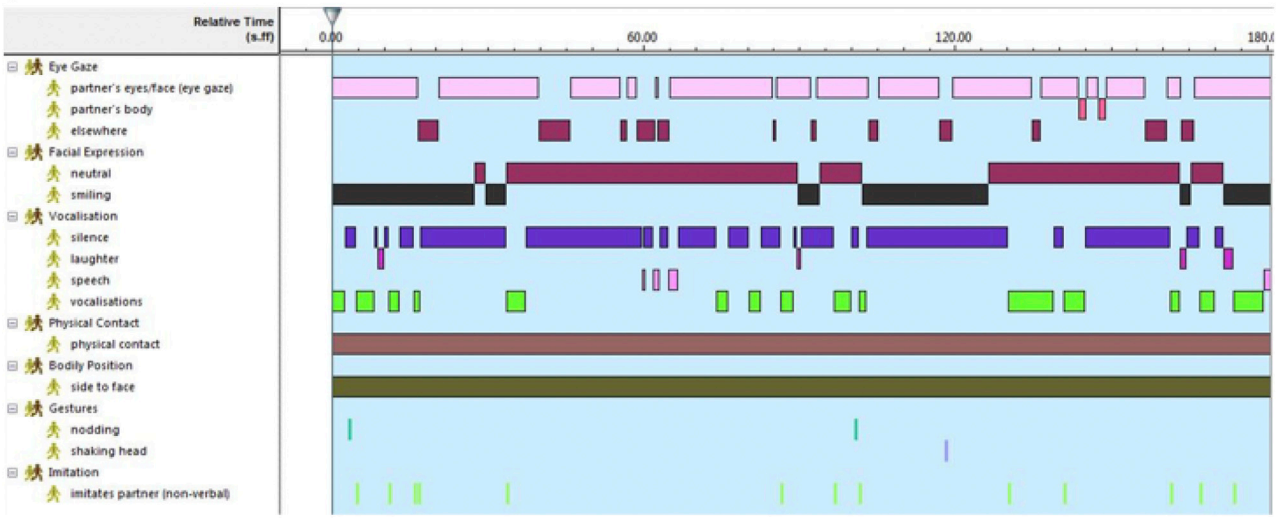

FIGURE 6C | continued

Figure 1H illustrates the pattern of Renee's communicative behaviour during session 4. Renee spent the majority of the interaction looking at Elena's body/hands (93%), and only looked elsewhere for short durations (5.2%). This is a marked change from all three previous sessions and indicates an increase in Renee's engagement with and attentiveness towards Elena. Renee had not previously engaged in imitation, yet she imitated Elena's hand movements on 9 occasions during this session (light green bar, **Figure 1H**). The percentage of time Renee spent initiating physical contact with Elena (23.6%) increased from previous sessions, as did the frequency of moving closer 6) and moving away 5) from Elena, suggesting a willingness and desire to "play" with and interact with her partner.

For space reasons we present the communication behaviour at baseline (session 1) and the third AI (session 4) of the other five dyad cases for comparison. The coding scheme and colours are the same as in the first dyadic case study of Elena and Renee.

Dyad 2: Diane and Derek

During the baseline condition, Diane used speech to try to engage Derek, spending 31.2% of the interaction speaking (**Figure 2A**; second pink line). Often, Diane's use of speech would provoke Derek to vocalize (**Figure 2B**, short light green bars). Her gaze was mostly directed at Derek's face or body, here facial expression was neutral and there was no physical contact (**Figure 2A**).

Derek's vocalizations, which were often very loud, were a key communicative behaviour in his repertoire at baseline (**Figure 2B**). Derek's use of hand/arm gestures was also extensive, gesturing 19 times during the initial baseline condition (**Figure 2B**, dark green line). Derek's engagement in "vocalizations" and "other hand/arm" gestures (yellow lines) was often erratic and gave the impression of him being agitated, which the staff reported that they felt made it difficult to engage in interactions with Derek. As a result, he experienced most

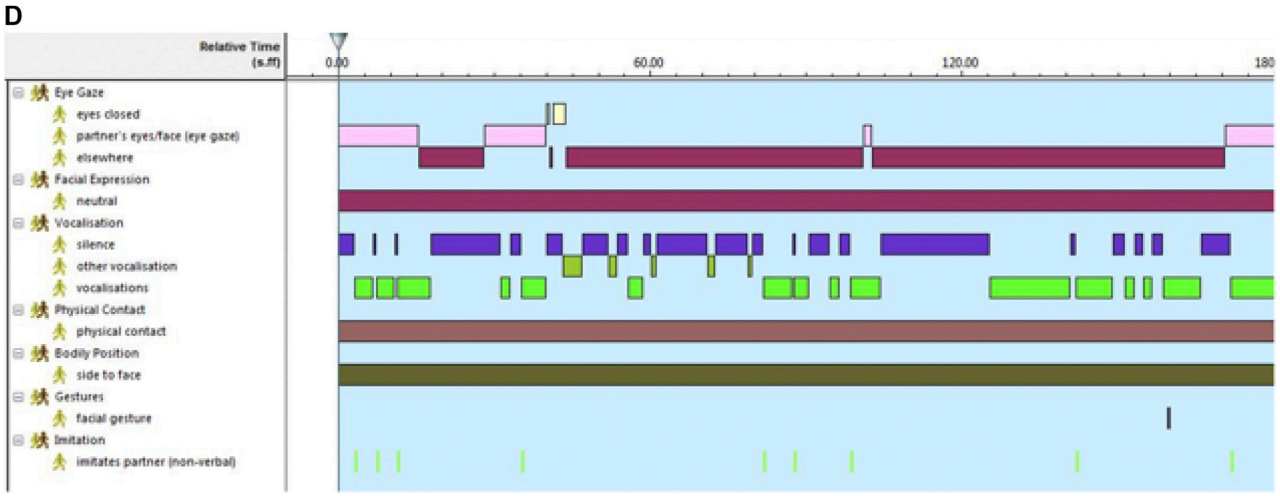

FIGURE 6D | continued

interactions during care activities at the start of the study and was often on his own as the staff feared he would injure other residents or staff through his unpredictable arm waving and gesturing.

After AI training, Diane's use of speech reduced to 10.4% in session 4 (**Figure 2C**, darker pink lines). The percentage of time Diane spent looking at Derek's eyes/face increased from 59.4% in session 1 (**Figure 2A**) to 82.3% in session 4 (**Figure 2C**). In the final session, Dian moved to sit by Derek\s side and initiated some physical contact and nonverbal imitation (**Figure 2C**).

Over the sessions as Diane's use of speech changed Derek's vocalizations also decreased and were of shorter duration and intensity (5.9% in session 4: **Figure 2D**, light green bar). Derek's unpredictable movement had also reduced to only 1 hand/arm gesture during the final session (**Figure 2D**, yellow bar). In the final session the dyad moved part way through to sit side by side and some nonverbal imitation took place between them which was a new component of the interaction (bottom **Figures 2C**,**D**).

Diane's reduced use of speech and increase in eye contact appeared to calm Derek, as indicated by reduced vocalizing and gesturing. Through spending time getting to know Derek's communication repertoire and in particular how he responded to different communicative bids from others, Diane reported that she felt better equipped for initiating interactions with Derek. She was also keen to upskill other staff on how best to engage with him to provide more opportunities for social interaction.

Dyad 3: Lorraine and Jen

Lorraine speaks a lot during the baseline session (**Figure 3A**, second pink bar) and also laughs a few times, but this is not reciprocated by Jen (**Figure 3B**). As she is speaking Lorraine looks at Jen, moving between the face and body, although quite a lot of time she is looking elsewhere (**Figure 3A**, first dark red line).

A distinct aspect of Jen's communicative repertoire was "vocalizations", which were similar to those made by Derek, in respect of being very loud and lengthy in duration. At the start of the project, Jen's vocalizations were identified as a problem in the care home. Jen did have some residual single words and in the baseline condition (3a), Lorraine commented in response to Jen's vocalizations: "I don't understand, use your words". In the baseline session, Jen spent a lot looking away from Lorraine (**Figure 3B**, first dark red bar). She did make two single-word utterances, but the majority of her sound-based communication was loud vocalizing which occurred quite frequently (**Figure 3B**, lower light green bar).

Lorraine continued to use speech during the AI sessions perhaps because she knew that Jen retained one or two individual words (**Figure 3C**, second pink bar). However, during the AI training and practice, a shift occurred during the second AI interaction (session 3; not shown here for space reasons). In this interaction Lorraine and Jen started building a shared language through imitation, physical contact and turn-taking. Specifically, Lorraine initiated physical contact 81.7% of the time and Jen 31% of the time (not shown here). Jen also initiated joint movement twice in this session whilst Lorraine initiated joint movement 6 times. Here, they would pull each other's hands/arm towards one another and would touch heads. Jen "moved closer" to Lorraine 11 times during this session.

As the sessions progressed, Lorraine imitated components of Jen's communicative repertoire including her vocalizations and facial gestures, with the number of imitations increasing steadily from 0 in session 1 (**Figure 3A**) to 12 in session 4 (**Figure 3C**). Lorraine's use of facial gestures increased similarly from 0 in session 1 (**Figure 3A**) to 15 in session 4 (**Figure 3C**, bottom row), suggesting that she became more comfortable using nonverbal communication alongside speech.

As occurred with Derek, the percentage of time Jen spent engaging in vocalizations reduced from 28% in session 1 (**Figure 3B**) to just 1% in session 4 (**Figure 3D**). The level of contentment and attentiveness in the dyad can be seen in the percentage of smiling which increased from 0% in session 1 (**Figure 3B**) to 49.6% in session 4 for Jen (**Figure 3D**), and from

11.2% in session 1 (**Figure 3A**) to 35.8% in session 4 for Lorraine (**Figure 3C**). In the final session, the amount of time Jen spent looking at Lorraine's face (43.4%) was roughly double that in the baseline (22.5%; **Figure 3B**), even though Jen and Lorraine were engaged in an activity which involved looking at a fish tank. A major change is the physical contact Jen makes in the final session (**Figure 3D**, brown bar) compared to none in the baseline session (**Figure 3B**).

During the reflection component of the training, Lorraine was able to share what she had learnt about Jen's communication and specifically her vocalizing. This empowered the team to develop a plan for reducing the occurrence of loud vocalizing, which was creating difficulties within the care home at the start of the training, whilst increasing opportunities for Jen to engage in social interactions, outside of personal care.

Dyad 4: Karen and Jane

Jane was the only participant with dementia who had retained some ability to use more than single words. However, her verbal language would get very muddled, and it was difficult to understand what she was saying and the meaning of her words. Jane's use of facial expressions was one way of understanding her communicative repertoire. Her facial expressions would shift rapidly from "neutral" to "smiling" to looking distressed (coded as "other facial expression"). As Karen recognized Jane's distressed expression, she would often touch and engage in "smiling" to comfort her. In the baseline session, Karen initiated physical contact 8.8% of the time (**Figure 4A**) and she consistently used gestures, including nodding, and shaking her head.

As Jane had some preserved speech, the turn-taking between her and Karen was more apparent at baselines, with Jane frequently looking at Karen's face (**Figure 4B**). Jane did speak during the baseline session, but the interaction was also maintained through nonverbal behaviours such as head nodding and other gestures (**Figure 4B**). Jane also made notable facial expressions including one which was labelled as distress but was not accompanied by other signs that might indicate she was experiencing pain or discomfort.

During the AI training, the percentage of time Karen spent initiating physical contact increased from 8.8% in session 1 (**Figure 4A**) to 48% in session 4 (**Figure 4C**). Karen's use of speech remained at a relatively constant level throughout the sessions (**Figure 4A** and4c), and she consistently used gestures, including nodding and shaking her head, In session 4 Karen moved closer to Jane (**Figure 4C**) and she also made some vocalizations.

Jane's level of contentment and engagement with Karen can be observed in the amount of time spent "smiling" which increased from 7.4% during session 1 (**Figure 4B**) to 30% in session 4 (**Figure 4D**). There was also an increase in the percentage of time Jane was looking at Karen's eyes/face which increased from 44% in session 1 (**Figure 4B**) to 87% in session 4 (**Figure 4D**). Jane used less speech in the final AI session and also vocalized, which appeared to be in response to Karen. Her head nodding and shaking also increased (**Figure 4D**).

Jane and Karen present an interesting example of the familiar situation that occurs when people living with dementia still have some speech, but it is no longer sufficient or functional for supporting interactions. There is usually a tendency to rely on speech, but what emerged during the AI sessions is the importance of attending to and learning an individual's communicative repertoire even when they have some speech, to identify the nonverbal behaviours that can support continued social interaction.

Dyad 5: Jake and Ernie

In the baseline session Jake spoke for around 18% of the time and made no attempts to imitate Ernie (**Figure 5A**). Jake mostly has a neutral facial expression with occasional smiles (**Figures 5A**, **C**). In session 1 Jake attempts physical contact (**Figure 5A**) which Ernie withdraws from. However, by session 4 this has reduced to almost none (**Figure 5C**) as he became more familiar with Ernie's repertoire.

Ernie's started the first session with his eyes closed (26.3% of time; **Figure 5B**) and when he opened them, he mostly looked away from Jake. At baseline a small number of communicative behaviours were evident including hand/arm gestures, and facial gestures such as blinking (**Figure 5B**). He moved away when Jake made physical contact and made no sounds at all (**Figure 5B**).

Jake's speech reduced steadily from 17.8% in session 1 (**Figure 5A**) to 2.2% in session 4 (**Figure 5C**), where he is almost totally silent. After AI training, Jake attempts to imitate Ernie 7 times in the final session 4 (**Figure 5C**, light green bar) relative to no attempts in the baseline (**Figure 5A**). Jake is also more attentive to Ernie in the final session, looking almost constantly at his face (**Figure 5C**), as opposed to the first session where he gazes towards and away from Ernie throughout (**Figure 5A**).

In the final session Ernie kept his eyes open pretty much all of the time, which may be the result of efforts by Jake to engage with Ernie. Ernie still spent most of the time looking away from Jake (**Figure 5D**) Much of this time was spent looking at the camera/researcher, which consistently occurred across the four sessions. Attempts by Jake to touch Ernie would result in Ernie moving away from Jake and/or "frowning". In the final session Ernie vocalized a couple of times, which was a new behaviour and his facial expressions increased (**Figure 5D**).

While people living with dementia who make loud vocalizations (such a Derek and Jen) are often considered "problematic" in care settings, people such as Ernie, who make no sounds (baseline) can be overlooked or their silence interpreted as contentment or satisfaction. The small changes that occurred in Ernie's behaviour suggest that he started to respond to Jake's attentiveness and imitation, which could lead to increased social opportunities as the caregivers understand how to engage and interact with him.

Dyad 6: Graham and Marie

Marie was in the very late stages of dementia and was limited in her movements. Graham was fairly new to caregiving and paid close attention to Marie during the baseline session, looking at her for long periods (**Figure 6A**). He spoke briefly and waited to see if

she responded and also nodded his head, perhaps to reinforce that he was attending to her (**Figure 6A**). Graham initiated physical contact once or twice but primarily tried speech to engage Marie.

A key component of Marie's communicative repertoire was her heavy breathing (**Figure 6B** "other vocalizations"). For most of the time she was silent but there was one instance of laughter in response to Graham laughing (**Figures 6A**,**B** dark pink bar). During the baseline session Marie looked at Graham a few times, including when she was laughing, but mostly she looked elsewhere.

Through the AI training, Graham was able to identify Marie's heavy breathing and imitate this to engage Marie, which resulted in turn-taking during the later sessions. This is evident from the increasing number of instances where Graham imitated Marie from 0 during session 1 (**Figure 6A**) to 17 times during session 4 (**Figure 6C**). The percentage of time Graham spent initiating physical contact also increased from 5.8% in session 1 (**Figure 6A**) to 100% in session 4 (**Figure 6C**). As Graham learnt more about Marie's communication repertoire, his use of speech reduced from 28.3% in session 1 (**Figure 5A**) to 2.6% in session 4 (**Figure 6C**) as he vocalized with Marie (**Figure 6C**, light green bar). Graham continued to use head nodding and shaking to support turn taking.

The number of times Marie imitated Graham also increased steadily from 0 during session 1 (**Figure 6B**) to 13 times during session 4 (**Figure 6D**), highlighting the "turn-taking" that took place. Marie also engaged in physical contact for 100% of the time in session 4 (**Figure 6D**), with Graham commenting that she was "holding my fingers". In the final session Marie had her eyes closed some of the time, during which one of the other staff commented that she looked "blissful". Indeed, the interactions between Graham and Marie were very calm, with the growing sense of connection apparent to observers.

Marie, like Ernie, was a very quiet resident and due to her mobility restrictions was at risk of social exclusion. The recordings of her sessions with Graham and his reflection on them encouraged the staff as a group to think about how they could spend quality time with Marie, who actually died shortly after the end of the training.

Communication Behaviour Summary

Unsurprisingly the caregivers primarily used speech in the baseline interactions to engage the people living with dementia. When they did not respond to speech, they often used physical contact to try to gain their partner's attention. They were self-conscious of being recorded and often uncomfortable trying to have an interaction that did not involve a care act, such as supporting the individual to eat. Many of the interactions took place side to face, as the participants were in bed. Only Karen positioned herself face to face with Jane during the baseline session, where she also gestured and smiled unlike most of her colleagues. Karen was also the only caregiver who used imitation in the baseline session, suggesting that she was comfortable with nonverbal means of communication before the AI training.

During the AI sessions, the use of speech declined considerably as the caregivers started to attend to and reflect their partner's nonverbal communicative behaviours. There was a notable shift in the nature of the interactions, with caregivers focusing intently on their partner and the act of communication. As they concentrated on connecting with their partner, they also lost their self-consciousness. In the video review sessions, the caregivers were excited to point out where they noticed a communicative behaviour, how they imitated or expanded a sound or gesture and the reaction of their partner. Playfulness also emerged, such as Renee moving her hand towards and then away from Elena in the final session and Elena following which she described as a "game".

Two key points emerged from the six dyadic cases. First is that each caregiver was able to learn to use AI to elucidate their partner's communicative repertoire and identify aspects of it to reflect back. These repertoires ranged from loud vocalizing to complete silence, facial expressions, eye gaze direction, and physical movements. This confirms that each person with advanced dementia retained the urge to communicate and that caregivers were able to recognise this which motivated them to continue the interactions.

The second point is that the nature of the interactions changed over time with different interaction patterns apparent in the final AI session (figures "g" and "h" in Dyad 1 and figure "c" and "d" in Dyads 2–6) relative to the baseline interactions (figures "a" and "b" in all dyads). All of the caregivers reduced the amount of speech they used and increased nonverbal behaviours, such as eye gaze, imitation and reflecting back the communicative behaviours of the individual with advanced dementia. For example, Graham identified Marie's heavy breathing as a communicative act and reflected this back to make an intense connection. Whilst caregivers often started to use vocalizing to connect with their partners, for example by imitating a sound or breathing pattern, the loud, vocalizations of two residents—Derek and Jen–were substantially reduced over the AI sessions as other communicative behaviours developed within their dyads. This confirms that it is possible to build a communicative relationship with individuals with advanced dementia who can no longer use speech.

In a separate interview conducted by an independent researcher 3 months after the AI training, the caregivers all said that they found the training excellent and did not have any suggestions to improve it (Dampney-Jay, 2015). Three caregivers reported that watching back videos of themselves interacting was a particularly helpful element of the course and two commented that having opportunities to practice the skills as they learn them was beneficial.

DISCUSSION

This study describes how caregivers can be trained to use Adaptive Interaction, a simplified method, to communicate with people living with dementia who can no longer speak. Introducing the fundamentals of communication to care staff enabled them to identify non speech based communicative behaviours in the

residents they care for. Staff learnt how to attend to the individual residents, to identify an initial communicative behaviour and make a connection with them. The training also equipped them to apply AI to identify the communicative repertoires of people living with advanced dementia and use this as the basis for a meaningful interaction. The study also further extended understanding about the individuality of nonverbal communication repertoires (Ellis and Astell, 2017) and how they can be uncovered with AI.

The findings indicate that it is feasible to train staff in long term dementia care settings to use AI to engage with and interact with nonverbal residents. These individuals, who are often also immobile, are completely reliant on caregivers for social interactions either in or out of their bedrooms (Astell and Ellis, 2006). However, staff often feel that without speech, they can no longer connect with residents. The findings that AI can be used to make a connection and maintain this nonverbally suggests that the Social Inclusion Process (SIP) model of Intensive Interaction (Firth, 2009) on which AI is based, can be applied to people living with advanced dementia. the SIP model emphasises making a connection and learning the language of the individual who does not have speech (Caldwell, 2006). This finding presents a hopeful message for people living with dementia and caregivers as it provides a means of keeping them engaged in the social world (Currie, 2020). This can open up new ways of thinking about how to meet the needs of nonverbal individuals and improve their wellbeing. For example, in this study as the staff became more aware of the ways in which their nonverbal residents communicated, this generated discussions about care planning to increase the wellbeing of residents using AI.

Training in communication skills can raise awareness of the needs of people living with dementia and improve their wellbeing and quality of life (Surr, et al., 2017). Equipping care staff to use AI could overcome some of the barriers identified in previous studies regarding communication with people with advanced dementia (Beer et al., 2012). For example, Beer and colleagues (2012) found that communication training provided to nursing aides increased their awareness of the need for meaningful contact with people living with advanced dementia but did not improve their comfort levels or perceive skills for working with this population. The interactive training approach described here combining case examples, hands-on practice, and reflection may be successful at overcoming barriers to staff competence and confidence at using non speech-based communication.

Additionally, this study included two components aimed at implementing and sustaining AI in long-term care. The first was to include someone from the care home management team in the training, in this case the Deputy Manager, in line with the importance identified by Clegg, et al. (2020) for "organizational support". The second aspect was to include in the training programme a session to develop a strategy 1) to assess communication in the home to identify residents who could benefit from AI and 2) to disseminate AI to other caregivers and family members of the nonverbal residents.

Limitations and Further Research

A single rater coded 100% of the videos, which was partially addressed through this rater being independent of the study and having a second rater code a proportion of the videoclips. The small scale and short duration could be regarded as another limitation, although in long-term care settings, freeing up staff to participate in training is a perennial challenge. Interactions recorded over a longer duration of time (e.g., 6 months) could permit the effects of consolidation over time to be measured. To address this, interviews about the training and consolidation were conducted 3 months after the end by a researcher external to the research team.

CONCLUSION

Adaptive Interaction is a simplified approach to communication that can equip caregivers with the skills to communicate effectively with individuals with dementia who can no longer speak. Caregivers were able to use AI to learn the language of the individuals they care for and adopt nonverbal strategies to connect with them. Each individul living with advanced dementia had their unique communicative repertoire comprising a specific set of nonverbal behaviours. These findings support the utility of AI to elucidate even the most subtle communicative behaviours, whilst looking across the dyads collectively, illustrates the range of such behaviours. The findings also suggest an increase in the quality of communication as indicated by more frequent positive social behaviours and meaningful actions such as eye gaze, turn-taking and initiating physical contact by both partners in the interactions. Adaptive Interaction could be a useful tool for improving the quality of life and wellbeing of people living with advanced dementia who can no longer speak by providing a means of enhancing caregiving relationships. This in turn could improve the job satisfaction and feelings of competence of the people who care for them.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion 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 NRES Committee London—Camberwell, St. Giles (Ref: 12/LO/0818). The ethics committee waived the requirement of written informed consent for participation

AUTHOR CONTRIBUTIONS

AA co-designed the study, delivered the training and prepared the manuscript SS co-developed the video doing scheme, coded the videos and wrote up the results ME co-designed the study, delivered the training, co-developed the video coding scheme.

REFERENCES

Alzheimer's Association (2021). 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 17(3):327-406. Available athttps://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12328 doi:10.1002/alz.12328

Anderson, C. (2006). Early Communication Strategies: Using Video Analysis to Support Teachers Working with Preverbal Pupils. *Br. J. Spec. Education* 33 (3), 114–120. doi:10.1111/j.1467-8578.2006.00426.x

Argyropoulou, Z., and Papoudi, D. (2012). The Training of a Child with Autism in a Greek Preschool Inclusive Class through Intensive Interaction: a Case Study. *Eur. J. Spec. Needs Education* 27 (1), 99–114. doi:10.1080/08856257.2011.640489

Astell, A. J., and Ellis, M. P. (2006). The Social Function of Imitation in Severe Dementia. *Inf. Child. Develop.* 15, 311–319. doi:10.1002/icd.455

Barber, M. (2007). Imitation, Interaction and Dialogue Using Intensive Interaction: tea Party Rules. *Support Learn.* 22 (3), 124–130. doi:10.1111/j.1467-9604.2007.00459.x

Beer, L. E., Hutchinson, S. R., and Skala-Cordes, K. K. (2012). Communicating with Patients Who Have Advanced Dementia: Training Nurse Aide Students, Communicating with Patients Who Have Advanced Dementia: Training Nurse Aide Students. *Gerontol. Geriatr. Education* 33 (4), 402–420. doi:10.1080/02701960.2012.702165

Berridge, S., and Hutchinson, N. (2021). Staff Experience of the Implementation of Intensive Interaction within Their Places of Work with People with Learning Disabilities And/or Autism. *J. Appl. Res. Intellect. Disabil.* 34, 1–15.

Bowie, P., and Mountain, G. (1993). Using Direct Observation to Record the Behaviour of Long-Stay Patients with Dementia. *Int. J. Geriat. Psychiatry* 8, 857–864. doi:10.1002/gps.930081009

Bunn, F., Lynch, J., Goodman, C., Sharpe, R., Walshe, C., Preston, N., et al. (2018). Improving Living and Dying for People with Advanced Dementia Living in Care Homes: a Realist Review of Namaste Care and Other Multisensory Interventions. *BMC Geriatr.* 18, 303. doi:10.1186/s12877-018-0995-9

Caldwell, P. (2005). *Finding You, Finding Me: Using Intensive Interaction to Get in Touch with People with Sever Learning Disabilities Combined with Autistic Spectrum Disorder*. London: Jessica Kingsley Publisher.

Caldwell, P. (2011). *Intensive Interaction: Using Body Language to Communicate*. UK. JODD. Available at http://www.intellectualdisability.info/how-to-guides/articles/intensive-interaction-using-body-language-to-communicate.

Caldwell, P., and Horwood, J. (2007). *From Isolation to Intimacy*. London: Jessica Kingsley Publisher.

Clegg, J., Black, R., Smith, A., and Brumfitt, S. (2020). Examining the Impact of a City-wide Intensive Interaction Staff Training Program for Adults with Profound and Multiple Learning Disability: a Mixed Methods Evaluation. *Disabil. Rehabil.* 42 (2), 201–210. doi:10.1080/09638288.2018.1495769

Currie, L. (2020). Adaptive Interaction as a Method of Non-verbal Communication in Individuals with Dementia. *Independent Nurse.* 29(4):176-180. Feb 11th Available at https://www.independentnurse.co.uk/year-of-the-nurse/adaptive-interaction-as-a-method-of-non-verbal-communication-in-individuals-with-dementia/224051/. doi:10.12968/pnur.2018.29.4.176

Dampney-Jay, G. (2015). *Improving Nonverbal Communication between Caregivers and People with Advanced Dementia: the Effect of Staff Training in Intensive Interaction on Quality of Life*. Unpublished DClinPsych Thesis. Canterbury, United Kingdom: Canterbury Christchurch University.

Ellis, M., and Astell, A. J. (2004). The Urge to Communicate in Severe Dementia. *Brain Lang.* 91 (1), 51–52. doi:10.1016/j.bandl.2004.06.028

Ellis, M. P., and Astell, A. J. (2011). Intensive Interaction for People with Dementia: A New Approach to Communication? *J. Demen. Care* 19 (3), 24–26.

Ellis, M. P., and Astell, A. J. (2008). *Promoting Communication with People with Severe Dementia*. in *Techniques for Promoting Social Engagement in Individuals with Communicative Impairments* Editor S. Zeedyk (London: Jessica Kingsley Publishers).

Ellis, M. P., and Astell, A. J. (2017). Communicating with People with Dementia Who Are Nonverbal: the Creation of Adaptive Interaction. *PLOS ONE* 12 (8), e0180395.

Ephraim, G. W. (1982). *Developmental Process in Mental Handicap: A Generative Structure Approach*. Unpublished PhD Thesis. Bristol, United Kingdom: Brunel University.

Firth, G. (2021). *Connecting with Intensive Interaction*. Available at https://connectingwithintensiveinteraction.com/category/research/research-summaries-with-adult-participants/.

Firth, G. (2009). A Dual Aspect Process Model of Intensive Interaction. *Br. J. Learn. Disabilities* 37 (1), 43–49. doi:10.1111/j.1468-3156.2008.00505.x

Firth, G., Elford, H., Crabbe, M., and Leeming, C. (2007). Intensive Interaction as a Novel Approach in Social Care: Care Staff's Views on the Practice Change Process. *J. Appl. Res. Intellect. Disabil.* 21, 58–69. doi:10.1111/j.1468-3148.2007.00369.x

Firth, G., Elford, H., Crabbe, M., and Leeming, C. (2004). Intensive Interaction: Changing Views and Relationships? *Draft Final Report*. Leeds: Leeds Mental Health NHS Trust.

Firth, G., Poyser, C., and Guthrie, N. (2013). Training Care Staff in Intensive Interactions. *Learn. Disabil. Pract.* 16 (10), 14–19. doi:10.7748/ldp2013.12.16.10.14.e1499

Frank, E. M. (1994). Effect of Alzheimer's Disease on Communication Function. *J. South Carolina Med. Assoc.* 90 (9), 417–423.

Fraser, C. (2011). Can Adults on the Autism Spectrum Be Affected Positively by the Use of Intensive Interaction in Supported Living Services? *Good Autism Pract.* 12 (2), 37–42.

Harris, C., and Wolverson, E. (2014). Intensive Interaction: to Build Fulfilling Relationships. *J. Demen. Care* 22 (6), 27–30.

Hewett, D. (1996). "How to Do Intensive Interaction," in *Interactive Approaches to Teaching: A Framework for INSET*. Editors M. Collis. and P. Lacey (London: David Fulton).

Hewett, D. (2011). Intensive Interaction. *Theoretical Perspectives*. London, UK: SAGE.

Hughes, S., Woods, B., Algar-Skaife, K., and Hedd Jones, C. (2019). Understanding Quality of Life and Well-Being for People Living with Advanced Dementia. *Nurs. Old. People* 31 (2), 18–24. doi:10.7748/nop.2019.e1129

Intensive Interaction Institute (2021). *The Fundamentals of Communication*. Spring Arbor University. Available at https://www.intensiveinteraction.org/find-out-more/fundamentals-of-communication/.

Kellet, M. (2005). Catherine's Legacy: Social Communication Development for Individuals With Profound Learning Difficulties and Fragile Life Expectancies. *Brit. J. Spec. Edu.* 32 (3), 116–122. doi:10.1111/j.0952-3383.2005.00383.x

Kitwood, T. (1997). Dementia Reconsidered. *The Person Comes first Maidenhead*. Maidenhead, United Kingdom: Open University Press.

Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159–174. doi:10.2307/2529310

Mental Capacity Act (2005). Available at: https://www.legislation.gov.uk/ukpga/2005/9/contents.

Nind, M. (1996). 'Efficacy of Intensive Interaction: Developing Sociability and Communication in People with Severe and Complex Learning Difficulties Using an Approach Based on Caregiver-Infant Interaction'. *Eur. J. Spec. Educ. Needs* 11 (1), 48–66.10.1080/0885625960110104.

Nind, M. (2006). Efficacy of Intensive Interaction: Developing Sociability and Communication in People with Severe and Complex Learning Difficulties Using an Approach Based on Caregiver-infant Interaction. *Eur. J. Spec. Needs Education* 11 (1), 48–66. doi:10.1080/0885625960110104

Samuel, J., Nind, M., Volans, A., and Scriven, I. (2008). An Evaluation of Intensive Interaction in Community Living Settings for Adults with Profound Intellectual Disabilities. *J. Intellect. Disabil.* 12 (2), 111–126. doi:10.1177/1744629508090983

Surr, C. A., Gates, C., Irving, D., Oyebode, J., Smith, S. J., Parveen, S., et al. (2017). Effective Dementia Education and Training for the Health and Social Care Workforce: A Systematic Review of the Literature. *Rev. Educ. Res.* 87 (5), 966–1002. doi:10.3102/0034654317723305

Watson, J., and Fisher, A. (1997). Research Section: Evaluating the Effectiveness of Intensive Interactive Teaching with Pupils with Profound and Complex Learning Difficulties. *Br. J. Spec. Education* 24, 80–87. doi:10.1111/1467-8527.00020

Zhou, Y., Forbes, G., and Humphris, G. M. (2013). The Behaviour of Preschool Children Receiving Fluoride Varnish Application in a Community Setting. *Br. Dental J.* 215 (7), E11. doi:10.1038/sj.bdj.2013.990

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Astell, Shoaran and Ellis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

BRIEF RESEARCH REPORT
published: 23 February 2022
doi: 10.3389/fcomm.2022.706718

Automatic Text Simplification for German

Sarah Ebling, Alessia Battisti, Marek Kostrzewa, Dominik Pfütze, Annette Rios, Andreas Säuberli and Nicolas Spring*

Department of Computational Linguistics, University of Zurich, Zurich, Switzerland

The article at hand aggregates the work of our group in automatic processing of simplified German. We present four parallel (standard/simplified German) corpora compiled and curated by our group. We report on the creation of a gold standard of sentence alignments from the four sources for evaluating automatic alignment methods on this gold standard. We show that one of the alignment methods performs best on the majority of the data sources. We used two of our corpora as a basis for the first sentence-based neural machine translation (NMT) approach toward automatic simplification of German. In follow-up work, we extended our model to render it capable of explicitly operating on multiple levels of simplified German. We show that using source-side language level labels improves performance with regard to two evaluation metrics commonly applied to measuring the quality of automatic text simplification.

Keywords: simplified language, automatic text simplification, automatic sentence alignment, sentence alignment gold standard, German, neural machine translation

OPEN ACCESS

Edited by:
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

Reviewed by:
Horacio Saggion,
Pompeu Fabra University, Spain
Julia Fuchs,
Laboratoire STL - Savoirs, Textes et Langage (UMR 8163), France

****Correspondence:***
Sarah Ebling
ebling@cl.uzh.ch

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

Received: *07 May 2021*
Accepted: *18 January 2022*
Published: *23 February 2022*

Citation:
Ebling S, Battisti A, Kostrzewa M, Pfütze D, Rios A, Säuberli A and Spring N (2022) Automatic Text Simplification for German. Front. Commun. 7:706718. doi: 10.3389/fcomm.2022.706718

1. INTRODUCTION

Simplified language[1] is a variety of standard language characterized by reduced lexical and syntactic complexity, the addition of explanations for difficult concepts, and clearly structured layout. Two tasks deal with automatic processing of simplified language: automatic readability assessment and automatic text simplification (Saggion, 2017).

Automatic text simplification was initiated in the late 1990s (Chandrasekar et al., 1996; Carroll et al., 1998) and since then has been approached by means of rule-based and statistical methods. As part of a rule-based approach, the operations carried out typically include replacing complex lexical and syntactic units with simpler ones (Chandrasekar et al., 1996; Siddharthan, 2002; Gasperin et al., 2010; Bott et al., 2012; Drndarević and Saggion, 2012). A statistical approach (Specia, 2010; Zhu et al., 2010) generally conceptualizes the simplification task as one of converting a standard-language into a simplified-language text using machine translation techniques on a sentence level. The success of such approaches is contingent on the availability of high-quality sentence alignments.

Research on automatic text simplification is comparatively widespread for languages such as English (Zhu et al., 2010), Spanish (Saggion et al., 2015), Portuguese (Aluisio and Gasperin, 2010), French (Brouwers et al., 2014), Italian (Barlacchi and Tonelli, 2013), and other languages. For German, only few contributions exist. Research on simplified German has gained momentum in

[1]The term "simplified language" is used to denote the sum of all "comprehensibility-enhanced varieties of natural languages" (Maaß, 2020, p. 52), i.e., what is commonly termed "Easy Language" (German *leichte Sprache*) and "Plain Language" (German *einfache Sprache*). Maaß (2020, p. 52) mentions "easy-to-understand language" as an umbrella term subsuming these varieties. However, in this contribution, we prefer the term "simplified language" to emphasize the notion of the result of a simplification process.

recent years due to a number of legal and political developments in German-speaking countries, such as the introduction of a set of regulations for accessible information technology (*Barrierefreie-Informationstechnik-Verordnung, BITV 2.0*) in Germany, the approval of rules for accessible information and communication (*Barrierefreie Information und Kommunikation, BIK*) in Austria, and the ratification of the United Nations Convention on the Rights of Persons with Disabilities (CRPD) in Germany, Austria, and Switzerland. In addition, two volumes on Easy Language appeared in the "Duden" series (Bredel and Maaß, 2016a,b), further highlighting the relevance of the topic. See Maaß (2020, Chapter 2.3) for a comprehensive overview of the situation in Germany.

The article at hand aggregates the work of our group in automatic processing of simplified German. We present four parallel corpora compiled and curated by our group. We report on the creation of a gold standard of sentence alignments from the four sources for evaluating five alignment methods on this gold standard. We used two of the corpora as a basis for the first sentence-based neural machine translation (NMT) approach toward automatic simplification of German. In follow-up work, we extended our model to render it capable of explicitly operating on multiple levels of simplified German.

More specifically, the contributions of the article at hand are:

- Overview of four parallel (standard/simplified German) corpora, of which automatically generated sentence alignments for one source are available for research purposes
- Gold standard of sentence alignments from the four sources
- Evaluation of automatic sentence alignment methods based on the gold standard
- First sentence-based NMT approach toward automatic simplification of German
- First multi-level simplification approach for German

The remainder of this article is structured as follows: Section 2.1 discusses approaches to automatic sentence alignment in the context of text simplification. Section 2.2 discusses parallel standard-/simplified-language corpora available for language pairs other than standard German/simplified German. Section 2.3 presents previous approaches to automatic text simplification. Sections 3 to 5 present our own contributions, consisting of compiling four standard German/simplified German parallel corpora (Section 3), creating a gold standard for automatic sentence alignment (Section 4) against which to measure existing automatic sentence alignment methods, and performing automatic text simplification on a sentence level (Section 5). Section 6 offers a conclusion and an outlook on future research.

2. PREVIOUS WORK

2.1. Automatic Sentence Alignment

Sentence alignment between standard- and simplified-language texts is an instance of monolingual sentence alignment. As such, it is unable to rely on well-established heuristics of bilingual sentence alignment based on, for example, sentence length (Gale and Church, 1991). The relation between source and target sentences in a standard-language/simplified-language document pair can be of the following types:

- **1:1**, i.e., one standard-language sentence corresponding to one simplified-language sentence
- **n:1** (with n>1), i.e., more than one standard-language sentence reduced to a single simplified-language sentence
- **1:n** (with n>1), i.e., one standard-language sentence split up into multiple simplified-language sentences
- **n:m** (with n>1 and m>1), i.e., more than one standard-language sentence corresponding to more than one simplified-language sentence
- **1:0**, i.e., a standard-language sentence omitted in the simplified-language text
- **0:1**, i.e., a simplified-language sentence inserted compared to the standard-language text

This is visualized in **Figure 1**. Also shown in this figure is an example of a crossing alignment, i.e., an alignment where the order of information of the standard-language text is not the same as that of the simplified-language text (non-monotonicity).

A number of tools have been developed specifically for sentence alignment in the context of text simplification; among them are MASSAlign (Paetzold et al., 2017), CATS (Customized Alignment for Text Simplification) (Štajner et al., 2018), and LHA (Large-scale Hierarchical Alignment for Data-driven Text Rewriting) (Nikolov and Hahnloser, 2019).

MASSAlign is a hierarchical algorithm that uses a vicinity-driven approach. It employs a heuristic according to which the order of information is consistent on the standard- and simplified-language sides, allowing for reduction of the search space. In a first step, MASSAlign searches for alignments between paragraphs, and in a second, for sentence alignments within the aligned paragraphs. The tool employs a similarity matrix with a bag-of-words TF-IDF model with maximum TF-IDF cosine similarity as a similarity metric. The paragraph alignment uses three levels of vicinity: (1) 1:1, 1:n, and n:1 alignments; (2) single-unit skips (where units can be sentences or paragraphs); and (3)

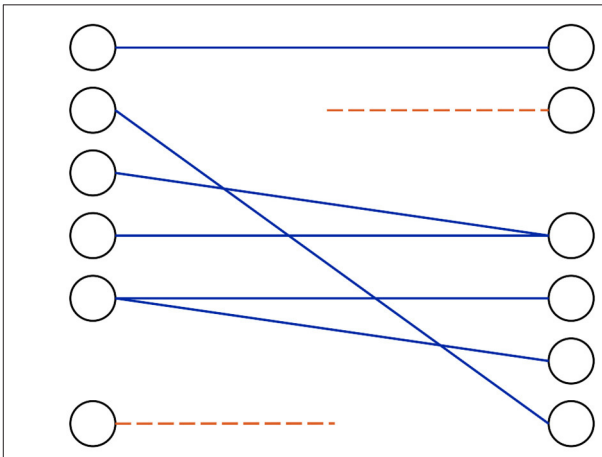

FIGURE 1 | Schematic depiction of source/target sentence relations in standard-language/simplified-language parallel texts.

long-distance unit skips. Sentence alignment relies on (1) and (2) only.

Like MASSAlign, CATS is capable of aligning paragraphs and sentences in two steps. The tool offers three similarity strategies, a lexical (character-n-gram-based, CNG) and two semantic similarity strategies. The two semantic similarity strategies, WAVG (Word Average) and CWASA (Continuous Word Alignment-based Similarity Analysis), both require pretrained word embeddings. WAVG averages the word vectors of a paragraph or sentence to obtain the final vector for the respective text unit (sentence or paragraph). CWASA is based on the alignment of continuous words using directed edges. CATS offers two different alignment strategies: MST (Most Similar Text) and MST-LIS (MST with Longest Increasing Sequence) to allow for 1:n alignment.

LHA uses a hierarchical alignment approach with two steps: Firstly, document alignment is performed based on document embeddings and an approximate nearest neighbor search using the Annoy library[2]. Annoy exhibits a low memory footprint *via* usage of static files as indexes. Secondly, sentence embeddings and an inter-sentence similarity matrix are used to extract K nearest neighbors for each source and target sentence. The tool further uses a variation of MST-LIS from CATS to model sentence splitting and compression.

Vecalign (Thompson and Koehn, 2019) and alignment based on SBERT (Reimers and Gurevych, 2020) were introduced in the context of bilingual sentence alignment. SBERT modifies the pretrained BERT network (Devlin et al., 2019) by using siamese and triplet network structures to arrive at sentence embeddings that may then be compared using cosine similarity. Vecalign is a method based on the similarity of the average sentence embedding with cosine similarity as the scoring function.

Table 1 characterizes the five alignment methods MASSAlign, CATS, LHA, SBERT, and Vecalign along the following aspects:

- **All source sentences aligned:** whether the alignment method in its default setup force-aligns every source sentence or bases the decision whether to align a source sentence on a similarity threshold (cutoff)[3]
- **Concatenation:** whether the alignment method concatenates multiple sentences into one and aligns them as one
- **Crossing alignments:** whether the alignment method allows for abandoning the monotonicity restriction, i.e., supports crossing alignments (cf. **Figure 1**)
- **Alignment type:** which relations between source and target sentences are ultimately supported by the method.

TABLE 1 | Overview of mono- and bilingual sentence alignment tools and methods.

Tool	All source sentences aligned	Concatenation	Crossing alignments	Alignment type
MASSAlign	Cutoff	Yes	No	n:m
CATS	Yes	No	Yes	n:1
LHA	Cutoff	No	Yes	n:m
SBERT	Cutoff	No	Yes	n:1
Vecalign	Yes	Yes	No	n:m

2.2. Sentence-Aligned Parallel Corpora

Automatic text simplification *via* (sentence-based) machine translation as outlined in Section 1 requires pairs of standard-language/simplified-language texts aligned at the sentence level, i.e., parallel corpora. A number of parallel corpora have been created to this end. Gasperin et al. (2010) compiled the PorSimples Corpus consisting of Brazilian Portuguese texts (2,116 sentences), each with two different levels of simplifications ("natural" and "strong,") resulting in around 4,500 aligned sentences. Bott and Saggion (2012) produced the Simplext Corpus consisting of 200 Spanish/simplified Spanish document pairs, amounting to a total of 1,149 (Spanish) and 1,808 (simplified Spanish) sentences (approximately 1,000 aligned sentences).

A large parallel corpus for text simplification is the Parallel Wikipedia Simplification Corpus (PWKP) compiled from parallel articles of the English Wikipedia and the Simple English Wikipedia (Zhu et al., 2010), consisting of about 108,000 sentence pairs. The difference in vocabulary size between the English and the simplified English side of the PWKP Corpus amounts to 18%[4]. Application of the corpus has been criticized for various reasons (Štajner et al., 2018); the most important among these is the fact that Simple English Wikipedia articles are often not translations of articles from the English Wikipedia. Hwang et al. (2015) provided an updated version of the corpus that includes a total of 280,000 full and partial matches between the two Wikipedia versions.

Another frequently used data collection, available for English and Spanish, is the Newsela Corpus (Xu et al., 2015) consisting of 1,130 news articles, each simplified into four school grade levels by professional editors. The difference in vocabulary size between the English side and the simplest level (Simple-4) is 50.8%.

The above-mentioned PorSimples and Newsela corpora present standard-language texts simplified into multiple levels, thus accounting for a recent consensus in the area of simplified-language research, according to which a single level of simplified language is not sufficient; instead, multiple levels are required to account for the heterogeneous target usership.

2.3. Automatic Text Simplification

Specia (2010) introduced statistical machine translation to the automatic text simplification task, using data from a small parallel corpus (roughly 4,500 parallel sentences) for Portuguese. Coster and Kauchak (2011) used the PWKP Corpus in its original form (cf. Section 2.2) to train an MT system. Xu et al. (2016) performed

[2]https://github.com/spotify/annoy (last accessed: May 5, 2021).

[3]Note that for CATS, the alignment direction is from simplified language to standard language; hence, CATS searches for one or more standard-language sentences for each simplified-language sentence.

[4]Vocabulary size as an indicator of lexical richness is generally taken to correlate positively with complexity (Vajjala and Meurers, 2012).

syntax-based MT on the English/simplified English part of the Newsela Corpus (cf. Section 2.2).

Nisioi et al. (2017) pioneered NMT models for text simplification, performing experiments on both the Wikipedia dataset of Hwang et al. (2015) and the Newsela Corpus for English, with automatic alignments derived from CATS (cf. Section 2.1). The authors used LSTMs as instances of Recurrent Neural Networks (RNNs).

More recent contributions to ATS include explicit edit operation modeling (Dong et al., 2019), graded simplification (Nishihara et al., 2019), multi-task learning (Guo et al., 2018; Dmitrieva and Tiedemann, 2021), weakly supervised (Palmero Aprosio et al., 2019), and unsupervised approaches (Surya et al., 2019; Kumar et al., 2020; Laban et al., 2021). These approaches are largely limited to English (Al-Thanyyan and Azmi, 2021) due to a lack of training data in other languages.

Säuberli et al. (2020) presented the first approach to text simplification for German using (sentence-based) NMT models. As data, they used an early version of the APA Corpus (cf. Section 3.2) amounting to approximately 3,500 sentence pairs.

The most commonly applied automatic evaluation metrics for text simplification are BLEU (Papineni et al., 2002) and SARI (Xu et al., 2016). BLEU, the *de-facto* standard metric for machine translation evaluation, computes token n-gram overlap between a hypothesis and one or multiple references. A shortcoming of BLEU with respect to automatic text simplification is that it rewards hypotheses that do not differ from the input. By contrast, SARI was designed to punish such output. It does so by explicitly considering the input and rewarding tokens in the hypothesis that do not occur in the input but in one of the references (addition) and tokens in the input that are retained (copying) or removed (deletion) in both the hypothesis and one of the references. More precisely, SARI computes the arithmetic average of n-gram precision and recall of the three rewrite operations addition, copying, and deletion, specifically rewarding simplifications that are dissimilar from the input. The metric was shown to exhibit "reasonable correlation with human evaluation on the text simplification task" (Xu et al., 2016).

Table 2 displays BLEU and SARI scores for previous sentence-level simplification approaches for different languages.

TABLE 2 | Automatic evaluation scores for sentence-level ATS approaches (PBMT, phrase-based SMT; SBMT, syntax-based MT).

References	Language	Approach	Scores
Specia (2010)	Portuguese	SMT	60.75 BLEU
Coster and Kauchak (2011)	English	SMT	60.46 BLEU
Wubben et al. (2012)		PBMT	67.79 BLEU (Nisioi et al., 2017)
			34.07 SARI (Nisioi et al., 2017)
Xu et al. (2016)	English	SBMT	73.62 BLEU (Nisioi et al., 2017)
			38.59 SARI (Nisioi et al., 2017)
Nisioi et al. (2017)	English	NMT	87.50 BLEU
Štajner and Nisioi (2018)	English	NMT	Newsela:
			89.49 BLEU
			36.48 SARI
			PWKP:
			84.69 BLEU
			35.78 SARI
Säuberli et al. (2020)	German	NMT	9.75 BLEU
			36.88 SARI

3. COMPILING DATA FOR AUTOMATIC PROCESSING OF SIMPLIFIED GERMAN

This section reports on our contributions in building and curating four parallel corpora for use in automatic text simplification for German.

3.1. Web Corpus

Klaper et al. (2013) created the first parallel corpus for German/simplified German, consisting of 256 texts each (approximately 70,000 tokens) downloaded from the Web. Battisti et al. (2020) extended the corpus such that it contained more parallel data, newly contained monolingual-only data (simplified German), and newly contained information on text structure (e.g., paragraphs, lines), typography (e.g., font type, font style), and images (content, position, and dimensions)[5]. The parallel part of the corpus is useful for automatic text simplification *via* machine translation (cf. Section 2.3), the monolingual-only part for automatic readability assessment, which is not the focus of this article. In addition, monolingual-only data can also be leveraged as part of machine translation through applying back-translation, a data augmentation technique.

The corpus is compiled from PDFs and webpages collected from Web sources in Germany, Austria, and Switzerland. Information on the underlying guidelines for creating simplified German is not available, as the data was collected automatically. The sources mostly represent websites of governments, specialized institutions, and non-profit organizations. The documents cover a range of topics, such as politics (e.g., instructions for voting), health (e.g., what to do in case of pregnancy), and culture (e.g., introduction to art museums). The corpus contains 6,217 documents, of which 5,461 are monolingual-only, and 378 are available in both standard German and simplified German. The 378 parallel documents amount to 17,121 sentences on the standard German and 21,072 sentences on the simplified German side. Compared to their German counterparts, the simplified German texts in the parallel data have clearly undergone a process of lexical simplification: The vocabulary is smaller by 51% (33,384 vs. 16,352 types), which is comparable to the rate of reduction reported in Section 2.2 for the Newsela Corpus (50.8%).

[5]The importance of the latter type of information has repeatedly been stressed, e.g., for automatic readability assessment (Bredel and Maaß, 2016a; Arfé et al., 2018; Bock, 2018).

TABLE 3 | Examples from the Austria Press Agency (APA) corpus (Säuberli et al., 2020).

Original	*15,2 Prozent der Österreicher genießen Gemüe mehrmals am Tag, 41,0 Prozent öfter täglich und der Rest seltener.* ('15.2 percent of Austrians enjoy vegetables several times a day, 41.0 percent often daily and the rest less often.')
B1	*Nur rund 15 Prozent der Österreicher essen mehrmals am Tag Gemüse.* ('Only around 15 percent of Austrians eat vegetables several times a day.')
Original	*Jedes Kalb erhält spätestens sieben Tage nach der Geburt eine eindeutig identifizierbare Lebensnummer, die in Form von Ohrmarken beidseitig eingezogen wird.* ('At the latest seven days after birth, each calf is given a unique identification number, which is recorded on ear tags on both sides.')
B1	*In Österreich bekommt jedes Kalb spätestens 7 Tage nach seiner Geburt eine Nummer, mit der man es erkennen kann.* ('In Austria, at the latest 7 days after birth, each calf receives a number, with which it can be identified.')
Original	*US-Präsident Donald Trump hat in seiner mit Spannung erwarteten Rede zur Lage der Nation seine politischen Prioritäten betont, ohne große wirtschaftliche Initiativen vorzustellen.* ('In his eagerly awaited State of the Union address, U.S. President Donald Trump stressed his political priorities without presenting any major economic initiatives.')
B1	*US-Präsident Donald Trump hat am Dienstag seine Rede zur Lage der Nation gehalten.* ('U.S. President Donald Trump gave his State of the Union address on Tuesday.')
Original	*Sie stehe noch immer jeden Morgen um 6.00 Uhr auf und gehe erst gegen 21.00 Uhr ins Bett, berichtete das Guinness-Buch der Rekorde.* ('She still gets up at 6:00 a.m. every morning and does not go to bed until around 9:00 p.m., the Guinness Book of Records reported.')
B1	*Sie steht auch heute noch jeden Tag um 6 Uhr in der Früh auf und geht um 21 Uhr schlafen.* ('Even today, she still gets up at 6 every morning and goes to bed at 9.')

3.2. APA Corpus

A second corpus built by our group, which is a parallel corpus throughout, consists of news items of the Austria Press Agency (*Austria Presse Agentur*, APA) with their simplified versions.[6] At APA, four to six news items per day covering the topics of politics, economy, culture, and sports are manually simplified into two language levels, B1 and A2, following guidelines by *capito*, the largest provider of simplification services (translations and translators' training) in Austria, Germany, and Switzerland[7]. **Table 3** shows standard German/simplified German (B1) examples from the corpus (Säuberli et al., 2020). The corpus contains a total of 2,426 distinct documents. This amounts to 60,732 standard-language sentences, 30,328 sentences at level B1, and 30,432 sentences at A2. We generated sentence alignments with LHA (cf. Section 2.1), arriving at 10,268 alignments for B1 and 9,456 for A2. The sentence alignments are made available for research purposes[8].

3.3. Wikipedia Corpus

This parallel corpus was created by automatically translating 150,064 articles of the Simple English Wikipedia (cf. Section 2.2) to German using DeepL[9] [10]. The synthetically created "simplified German" articles were then aligned on a document level with their standard German counterparts from the German Wikipedia[11] using interlanguage links, resulting in 106,126 parallel documents with 6,933,192 standard German sentences and 1,077,992 "simplified German" sentences.

3.4. Capito Corpus

As a provider of simplification services, capito produces a high number of professional simplifications for a variety of documents and text genres. This includes but is not limited to booklets, information texts, websites and legal texts, which are manually simplified into one or more levels following the capito guidelines. The simplification levels in this corpus include B1, A2, and A1. We extracted simplified German documents along with their standard German counterparts, amounting to 1,055 document pairs for B1, 1,546 for A2, and 839 for A1. The documents contain a total of 183,216 standard-language sentences, 68,529 sentences at level B1, 168,950 sentences at level A2, and 24,243 sentences at level A1. Aligning the sentences with LHA (cf. Section 2.1) yielded 54,224 sentence pairs for B1, 136,582 for A2, and 10,952 for A1.

[6]Note that news items are among the most frequent sources of simplification (Caseli et al., 2009; Klerke and Søgaard, 2012; Bott and Saggion, 2014; Goto et al., 2015; Xu et al., 2015).

[7]https://www.capito.eu/ (last accessed: August 4, 2020). *capito* distinguishes between three levels along the Common European Framework of Reference for Languages (CEFR) Council of Europe (2009): A1, A2, and B1. Each level is linguistically operationalized, i.e., specified with respect to linguistic constructions permitted or not permitted at the respective level.

[8]https://zenodo.org/record/5148163 (last accessed: October 14, 2021).

[9]https://www.deepl.com/translator (last accessed: May 5, 2021). Simple English Wikipedia authors are instructed to "use Basic English words and shorter sentences", where Basic English refers to the variety introduced by Ogden (1944) that consists of 850 words on the lexical side.

[10]The Simple Wikipedia dump of 12/12/2019 was used, https://dumps.wikimedia.org/simplewiki/ (last accessed: April 26, 2021).

[11]Obtained by using the CirrusSearch dump as of 14/09/20, https://dumps.wikimedia.org/other/cirrussearch/ (last accessed: May 5, 2021).

TABLE 4 | Overview of the four parallel corpora for standard German/simplified German.

Corpus	No. of parallel documents	No. of original sentences	No. of simplified sentences		
			B1	A2	A1
Web Corpus	378	17,121		21,072	
APA Corpus	2,426	60,732	30,328	30,432	n.a.
Wikipedia Corpus	106,126	6,933,192		1,077,992	
Capito Corpus	2,279	183,216	68,529	168,950	24,243

For the Web Corpus and the Wikipedia Corpus, information on language levels is not available.

Table 4 presents an overview of the four data sources.

4. SENTENCE ALIGNMENT GOLD STANDARD AND EVALUATION OF AUTOMATIC SENTENCE ALIGNMENT METHODS

This section reports on the manual creation of a gold standard for sentence alignment based on a subset of the four corpora introduced in Section 3. We subsequently evaluate the five automatic sentence alignment methods presented in Section 2.1 against this gold standard to allow us to select the most accurately aligned sentences as data to train our translation models in Section 5. For more details on this evaluation, see Spring et al. (2021a).

4.1. Method

To create a gold standard against which to measure the performance of the different automatic sentence alignment methods introduced in Section 2.1, we selected approximately 1,500 simplified-language sentences from each of the four sources described in Section 3: the Web Corpus (where 36 documents amount to approximately 1,500 simplified sentences), APA Corpus (134 documents), Wikipedia Corpus (198 documents), and the capito Corpus (42 documents), as summarized in **Table 5**. Two annotators independently aligned the simplified sentences to their standard-language counterparts, considering all of the alignment types shown in Section 2.1. In case of n:1 or 1:n alignments, the annotators assigned a list of labels of length n to either the standard- or simplified-language sentence. In case of 1:0 or 0:1 alignments, the annotators assigned a placeholder label to the empty standard- or simplified-language sentence. Inter-annotator agreement (Cohen's Kappa) for all corpora was between 0.730 and 0.924 (cf. **Table 6**). To create a single version of the gold standard, an arbitrator took the final decision in cases where the two annotators disagreed.

4.2. Results

The alignment methods presented in Section 2.1 were used with their default settings and embeddings (where applicable)[12] to align sentences in the pairs of standard-language and simplified-language documents that make up the gold standard. Alignment was performed in both directions, simple to complex and vice versa, and the set of the extracted alignments for both directions was used. This made it possible to evaluate the alignment methods extracting n:1 alignments, even though the gold standard is n:m. Evaluation was performed with the Vecalign scoring script[13]. The scoring script made it possible to evaluate the diverse alignments that naturally occur in text simplification in a standardized way by converting all alignments to a collection of 1:1 alignments.

The results of evaluating the performance of the five alignment methods (MASSAlign, CATS, LHA, SBERT, Vecalign; with CATS featuring three sub-methods) against the gold standard are shown in **Table 7** (Spring et al., 2021a). Lower CEFR levels (available in the capito and APA data) proved harder to align and in general corresponded to lower F1 scores. The alignment task becomes harder with increasing distance from standard German, as simplification requires more modifications to the text. Also, on

TABLE 5 | Overview of the gold standard of sentence alignments for standard German/simplified German.

Corpus	Parallel documents	Original sentences	Simplified sentences
Web	36	1,454	1,440
APA	134	3,388	1,497
Wikipedia	198	11,668	1,530
capito	42	2,428	1,482
Total	410	18,938	5,949

TABLE 6 | Cohen's Kappa per data source.

Web	Wikipedia	capito A1	capito A2	capito B1	APA A2	APA B1
0.887	0.922	0.924	0.730	0.873	0.885	0.886

TABLE 7 | F1 scores of sentence alignment evaluation from Spring et al. (2021a).

Tool	Web	Wikipedia	Capito A1	Capito A2	Capito B1	APA A2	APA B1
MASSAlign	0.175	0.130	0.096	0.228	0.112	0.076	0.129
LHA	**0.339**	**0.170**	0.099	0.321	**0.513**	**0.150**	**0.213**
SBERT	0.218	0.104	0.205	0.348	0.321	0.119	0.136
CATS C3G	0.029	0.037	0.045	0.037	0.032	0.078	0.077
CATS CWASA	0.024	0.035	0.039	0.031	0.024	0.066	0.072
CATS WAVG	0.022	0.031	0.032	0.028	0.026	0.053	0.061
Vecalign	0.188	0.073	**0.215**	**0.392**	0.160	0.085	0.099
Mean	0.142	0.083	0.104	0.198	0.17	0.09	0.112

F1 score represents the mean between precision (how many of the alignments that were extracted by an alignment method were correct in the gold standard) and recall (how many of the alignments in the gold standard were retrieved by an alignment method). Bold: best-performing configuration per column.

[12]One of the tools, CATS, for example, offers an n-gram-based alignment approach that does not employ embeddings of any kind.

[13]https://github.com/thompsonb/vecalign (last accessed: April 26, 2021).

lower CEFR levels, elaborations and explanations are increasingly common. Generally, the alignment methods performed best on the Web and capito data, with average F1 scores being considerably higher. The low overall scores on the Wikipedia data could be explained by the fact that it is the dataset with the largest disparity between the number of standard German and simplified German sentences (cf. Section 3). Regarding the alignment methods, LHA performed best on five out of the seven datasets. It is also the method with the highest F1 scores on average. On capito A1 and capito A2, Vecalign reached the highest scores.

5. SENTENCE-BASED AUTOMATIC TEXT SIMPLIFICATION

This section reports on our work in training NMT models on two of the data sources introduced in Section 3. For more details, the reader is referred to Spring et al. (2021b).

5.1. Method

For these experiments, we used the APA and the capito corpora introduced in Sections 3.2 and 3.4, respectively, amounting to 19,724 sentence alignments for the APA Corpus (10,268 for B1 and 9,456 for A2) and 201,758 for the capito Corpus (54,224 for B1, 136,582 for A2, and 10,952 for A1), produced with LHA (cf. Section 2.1).

Our baseline models were trained on all available training data across all levels, i.e., these models were language-level-agnostic. They performed generic simplification because they had no explicit method to determine the desired level of simplification on the target side. We trained transformer models (Vaswani et al., 2017) with five layers, four attention heads, 512 hidden units in transformer layers, and 2048 hidden units in transformers feed forward layers. Embedding dropout and label smoothing were set to 0.3. We used BLEU for early stopping on a held-out development set with a patience of 10 checkpoints. We trained with a shared vocabulary (20,000 BPE operations). All our experiments were carried out in sockeye (Hieber et al., 2018).

Our experimental models made use of source-side labels corresponding to the desired CEFR level of the target sentence. These labels allow the model to make a distinction between the different CEFR levels and thus to simplify into different complexity levels. Among others, labels have been used in a variety of tasks such as domain adaption (Kobus et al., 2017), multilingual translation (Johnson et al., 2017), and making better use of back-translation (Caswell et al., 2019). Apart from these modifications to the training data, the model architecture and all hyperparameters were identical to the baseline models and they used the same vocabulary of 20k.

To evaluate our models, we used a test set that consists of 500 parallel sentences each for A1, A2, and B1, which were randomly sampled from the combined corpus.

5.2. Results

The BLEU and SARI scores of our two models on the test sets are presented in **Table 8**. The SARI values of our baseline model are comparable to the results of Säuberli et al. (2020) (cf. **Table 2**), who used a preliminary version of the APA corpus of approximately 3,500 sentence pairs (cf. Section 2.3), but our baseline achieved higher BLEU scores in the range of 13.4 to 16.3. The experimental model reached improved scores for both metrics. The use of source-side labels boosted performance in terms of BLEU on A1 and B1, with the new values in the range of 14.1 to 17.2. The BLEU score did not improve for A2, which was the level with the highest amount of parallel data available (cf. Section 3). This indicates that the addition of source-side labels may be especially helpful in low resource settings, as, on the other hand, A1 and B1, for both of which there was substantially less data, reached higher scores with the experimental model. In terms of SARI, the addition of source-side labels led to considerable improvements for all levels, with the new scores lying in the range of 41.53 to 43.12.

TABLE 8 | BLEU and SARI scores of the different models.

Model	A1		A2		B1	
	BLEU	**SARI**	**BLEU**	**SARI**	**BLEU**	**SARI**
Baseline	13.4	36.26	**14.4**	36.11	16.3	34.53
APA+capito multi	**14.2**	**43.12**	14.1	**41.53**	**17.2**	**41.81**

Bold: best-performing configuration per column.

6. CONCLUSION AND OUTLOOK

This article has presented the work of our group in automatic processing of simplified German. We have given an overview of four parallel corpora compiled and curated by our group: the Web, APA, Wikipedia, and capito corpora. Moreover, we have reported on the creation of a gold standard of sentence alignments from the four sources for evaluating five alignment methods on this gold standard (MASSAlign, CATS, LHA, SBERT, Vecalign; with CATS featuring three sub-methods). We found that LHA performed best on five out of the seven datasets (Web, Wikipedia, capito A1, capito A2, capito B1, APA A2, APA B1). It was also the method with the highest average F1 scores (on capito A1 and capito A2, Vecalign reached the highest absolute scores). In general, for the multi-level sources (capito and APA), lower CEFR levels proved harder to align and corresponded to lower F1 scores. Intuitively, the alignment task becomes harder with increasing distance from standard German, as simplification requires more modifications to the text. Also, on lower CEFR levels, elaborations and explanations are increasingly common. Generally, the alignment methods performed best on the Web and capito data, with average F1 scores being considerably higher. The low overall scores on the Wikipedia data can be explained by the fact that it is the dataset with the largest disparity between the number of standard German and simplified German sentences.

We used the LHA alignments as a basis for the first sentence-based neural NMT approach toward automatic simplification of German (baseline model), and we proposed a model that is capable of explicitly operating on multiple levels of simplified German. We showed that compared to our baseline model, this multi-level experimental model reached improved scores for both automatic evaluation metrics, BLEU and SARI. Specifically,

performance improved on all levels with respect to SARI and on A1 and B1 with respect to BLEU (A2 is the level with the highest amount of parallel data available).

We plan to further investigate the potential of the various alignment methods by varying the embedding strategies and the cutoff values used. In doing so, we expect to further increase the performance of our text simplification approaches according to automatic metrics. In addition, we plan to evaluate the output of future models with the help of human experts and to investigate the comprehensibility of the output among the target groups, e.g., persons with cognitive impairments.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data from the commercial provider of simplification services (capito) is not publishable. Sentence alignments based on APA have, however, been made publicly available here: https://zenodo.org/record/5148163.

AUTHOR CONTRIBUTIONS

SE as the group leader was involved in the creation of the parallel corpora, the sentence alignment gold standard, and the text simplification experiments. AS, NS, and AR carried out the text simplification experiments. AB was the primary person responsible for the Web corpus and one of the annotators of the sentence alignment gold standard. DP was the second annotator. NS was the arbitrator and performed the evaluation of the sentence alignment methods relative to the gold standard. MK provided the sentence alignments. All authors contributed to the article and approved the submitted version.

FUNDING

This submission received funding through the *capito automatisiert* project funded by the Austrian Research Promotion Agency (*Österreichische Forschungsförderungsgesellschaft*, FFG) General Programme under grant agreement number 881202.

ACKNOWLEDGMENTS

The authors are greatly indebted to the Austria Presse Agentur (APA) and CFS GmbH for providing data consisting of standard German documents with their simplified counterparts.

REFERENCES

Al-Thanyyan, S. S., and Azmi, A. M. (2021). Automated text simplification: a survey. *ACM Comput. Surveys (CSUR)* 54, 1–36. doi: 10.1145/3442695

Aluisio, S. M., and Gasperin, C. (2010). "Fostering digital inclusion and accessibility: the porsimples project for simplification of portuguese texts," in *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas* (Los Angeles, CA), 46–53.

Arfé, B., Mason, L., and Fajardo, I. (2018). Simplifying informational text structure for struggling readers. *Read. Writ.* 31, 2191–2210. doi: 10.1007/s11145-017-9785-6

Barlacchi, G. and Tonelli, S. (2013). "ERNESTA: a sentence simplification tool for children's stories in Italian," in *Proceedings of the 14th Conference on Intelligent Text Processing and Computational Linguistics (CICLing)* (Samos), 476–487.

Battisti, A., Pfütze, D., Säuberli, A., Kostrzewa, M., and Ebling, S. (2020). "A corpus for automatic readability assessment and text simplification of german," in *Proceedings of The 12th Language Resources and Evaluation Conference* (Marseille: European Language Resources Association), 3295–3304.

Bock, B. (2018). *"Leichte Sprache"–Kein Regelwerk. Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeiSA-Projekt*. Technical Report, Universität Leipzig.

Bott, S., and Saggion, H. (2012). "Automatic simplification of Spanish text for e-Accessibility," in *Proceedings of the 13th International Conference on Computers Helping People with Special Needs (ICCHP)* (Linz), 527–534.

Bott, S., and Saggion, H. (2014). Text simplification resources for Spanish. *Lang. Resour. Eval.* 48, 93–120. doi: 10.1007/s10579-014-9265-4

Bott, S., Saggion, H., and Figueroa, D. (2012). "A hybrid system for Spanish text simplification," in *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies* (Montreal, QC), 75–84.

Bredel, U. and Maaß, C. (2016a). *Leichte Sprache: Theoretische Grundlagen. Orientierung für die Praxis* (Berlin: Duden).

Bredel, U., and Maaß, C. (2016b). *Ratgeber Leichte Sprache: Die wichtigsten Regeln und Empfehlungen für die Praxis* (Berlin: Duden).

Brouwers, L., Bernhard, D., Ligozat, A.-L., and François, T. (2014). "Syntactic sentence simplification for French," in *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* (Gothenburg) 47–56.

Carroll, J., Minnen, G., Canning, Y., Devlin, S., and Tait, J. (1998). "Practical simplification of English newspaper text to assist aphasic readers," in *Proceedings of the AAAI'98 Workshop on Integrating AI and Assistive Technology* (Madison, Wisconsin), 7–10.

Caseli, H. M., Pereira, T. F., Specia, L., Pardo, T. A. S., Gasperin, C., and Aluisio, S. M. (2009). "Building a Brazilian Portuguese parallel corpus of original and simplified texts," in *10th Conference on Intelligent Text Processing and Computational Linguistics* (Mexico City), 59–70.

Caswell, I., Chelba, C., and Grangier, D. (2019). "Tagged back-translation," in *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)* (Florence: Association for Computational Linguistics), 53–63,

Chandrasekar, R., Doran, C., and Srinivas, B. (1996). "Motivations and methods for text simplification," in *Proceedings of the 16th Conference on Computational Linguistics* (Copenhagen), 1041–1044.

Coster, W., and Kauchak, D. (2011). "Learning to simplify sentences using wikipedia," in *Proceedings of the Workshop on Monolingual Text-To-Text Generation (MTTG)* (Portland, OR), 1–9.

Council of Europe (2009). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Cambridge: Cambridge University Press).

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Minneapolis, MN: Association for Computational Linguistics), 4171–4186.

Dmitrieva, A., and Tiedemann, J. (2021). "Creating an aligned Russian text simplification dataset from language learner data," in *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing* (Kiyv: Association for Computational Linguistics), 73–79.

Dong, Y., Li, Z., Rezagholizadeh, M., and Cheung, J. C. K. (2019). "EditNTS: an neural programmer-interpreter model for sentence simplification through explicit editing," in *Proceedings of the 57th Annual Meeting of the Association*

for Computational Linguistics (Florence: Association for Computational Linguistics), 3393–3402.

Drndarević, B., and Saggion, H. (2012). "Towards automatic lexical simplification in Spanish: an empirical study," in *Proceedings of the First Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* (Montreal, QC), 8–16.

Gale, W. A., and Church, K. W. (1991). "A program for aligning sentences in bilingual corpora," in *29th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference* (Berkeley, CA), 177–184.

Gasperin, C., Maziero, E., and Aluisio, S. M. (2010). "Challenging choices for text simplification," in *Computational Processing of the Portuguese Language. Proceedings of the 9th International Conference, PROPOR 2010* (Porto Alegre), 40–50.

Goto, I., Tanaka, H., and Kumano, T. (2015). "Japanese news simplification: task design, data set construction, and analysis of simplified text," in *Proceedings of MT Summit XV* (Miami, FL), 17–31.

Guo, H., Pasunuru, R., and Bansal, M. (2018). "Dynamic multi-level multi-task learning for sentence simplification," in *Proceedings of the 27th International Conference on Computational Linguistics* (Santa Fe, NM: Association for Computational Linguistics), 462–476.

Hieber, F., Domhan, T., Denkowski, M., Vilar, D., Sokolov, A., Clifton, A., et al. (2018). "The sockeye neural machine translation toolkit at AMTA 2018," in *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (AMTA)*, Vol. 1. 200–207.

Hwang, W., Hajishirzi, H., Ostendorf, M., and Wu, W. (2015). "Aligning sentences from standard wikipedia to simple wikipedia," in *Proceedings of NAACL-HLT* (Denver, CO), 211–217.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., et al. (2017). Google's multilingual neural machine translation system: enabling zero-shot translation. *Trans. Assoc. Comput. Linguist.* 5, 339–351. doi: 10.1162/tacl_a_00065

Klaper, D., Ebling, S., and Volk, M. (2013). "Building a German/simple german parallel corpus for automatic text simplification," in *ACL Workshop on Predicting and Improving Text Readability for Target Reader Populations* (Sofia), 11–19.

Klerke, S., and Søgaard, A. (2012). "DSim, a Danish parallel corpus for text simplification," in *Proceedings of Language Resources and Evaluation Conference (LREC)* (Istanbul), 4015–4018.

Kobus, C., Crego, J., and Senellart, J. (2017). "Domain control for neural machine translation," in *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017* (Varna: INCOMA Ltd.), 372–378.

Kumar, D., Mou, L., Golab, L., and Vechtomova, O. (2020). "Iterative edit-based unsupervised sentence simplification," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, eds D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault (Association for Computational Linguistics), 7918–7928.

Laban, P., Schnabel, T., Bennett, P., and Hearst, M. A. (2021). "Keep it simple: unsupervised simplification of multi-paragraph text," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (Association for Computational Linguistics), 6365–6378.

Maaß, C. (2020). *Easy Language – Plain Language – Easy Language Plus. Balancing Comprehensibility and Acceptability Easy – Plain – Accessible*, Vol. 3 (Berlin: Frank & Timme).

Nikolov, N. I., and Hahnloser, R. (2019). "Large-scale hierarchical alignment for data-driven text rewriting," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)* (Varna: INCOMA Ltd.), 844–853.

Nishihara, D., Kajiwara, T., and Arase, Y. (2019). "Controllable text simplification with lexical constraint loss," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (Florence: Association for Computational Linguistics), 260–266.

Nisioi, S., Štajner, S., Ponzetto, S. P., and Dinu, L. P. (2017). "Exploring neural text simplification models," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Vancouver, BC), 85–91.

Ogden, C. K. (1944). *Basic English: A General Introduction With Rules and Grammar*. Vol. 29. ed K. Paul (London, UK: Trench, Trubner).

Paetzold, G., Alva-Manchego, F., and Specia, L. (2017). "Massalign: alignment and annotation of comparable documents," in *Proceedings of the IJCNLP 2017, Tapei, Taiwan, November 27 - December 1, 2017, System Demonstrations*, esd S. Park and T. Supnithi (Tapei: Association for Computational Linguistics) 1–4.

Palmero Aprosio, A., Tonelli, S., Turchi, M., Negri, M., and Di Gangi, M. A. (2019). "Neural text simplification in low-resource conditions using weak supervision," in *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation* (Minneapolis, MI: Association for Computational Linguistics), 37–44.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (Philadelphia, PA), 311–318.

Reimers, N., and Gurevych, I. (2020). "Making monolingual sentence embeddings multilingual using knowledge distillation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Punta Cana: Association for Computational Linguistics), 4512–4525.

Saggion, H. (2017). *Automatic Text Simplification*. (San Rafael, CA: Morgan & Claypool Publishers).

Saggion, H., Štajner, S., Bott, S., Mille, S., Rello, L., and Drndarević, B. (2015). "Making it simplext: implementation and evaluation of a text simplification system for Spanish," *ACM Trans. Accessible Comput. (TACCESS)* 6:14.

Säuberli, A., Ebling, S., and Volk, M. (2020). Benchmarking data-driven automatic text simplification for German. "In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)* (Marseille: European Language Resources Association), 41–48.

Siddharthan, A. (2002). "An architecture for a text simplification system," in *Proceedings of the Language Engineering Conference* (Hyderabad), 64–71.

Specia, L. (2010). "Translating from complex to simplified sentences," in *Computational Processing of the Portuguese Language. Proceedings of the 9th International Conference, PROPOR 2010* (Porto Alegre), 30–39.

Spring, N., Pfütze, D., Kostrzewa, M., Battisti, A., Rios, A., and Ebling, S. (2021a). "Comparing sentence alignment methods for automatic simplification of german texts," in *Presentation Given at the 1st International Easy Language Day Conference (IELD)*, (Germersheim).

Spring, N., Rios, A., and Ebling, S. (2021b). "Exploring German multi-level text simplification," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, 1339–1349.

Štajner, S., Franco-Salvador, M., Rosso, P., and Ponzetto, S. P. (2018). "CATS: a tool for customized alignment of text simplification corpora," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (Miyazaki: European Language Resources Association (ELRA)).

Štajner, S., and Nisioi, S. (2018). "A detailed evaluation of neural sequence-to-sequence models for in-domain and cross-domain text simplification," in *Proceedings of the 11th Language Resources and Evaluation Conference* (Miyazaki).

Surya, S., Mishra, A., Laha, A., Jain, P., and Sankaranarayanan, K. (2019). "Unsupervised neural text simplification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Florence), 2058–2068.

Thompson, B., and Koehn, P. (2019). "Vecalign: improved sentence alignment in linear time and space," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong: Association for Computational Linguistics), 1342–1348.

Vajjala, S., and Meurers, D. (2012). "On improving the accuracy of readability classification using insights from second language acquisition," in *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP (NAACL HLT '12)* (Montreal, QC), 163–173.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc.), 5998–6008.

Wubben, S., van den Bosch, A., and Krahmer, E. (2012). "Sentence simplification by monolingual machine translation," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Jeju Island: Association for Computational Linguistics), 1015–1024.

Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: new data can help. *Trans. Assoc. Comput. Linguis.* 3, 283–297. doi: 10.1162/tacl_a_00139

Xu, W., Napoles, C., Pavlick, E., Chen, Q., and Callison-Burch, C. (2016). Optimizing statistical machine translation for text simplification. *Trans. Assoc. Comput. Linguist.* 4, 401–415. doi: 10.1162/tacl_a_00107

Zhu, Z., Bernhard, D., and Gurevych, I. (2010). "A monolingual tree-based translation model for sentence simplification," in *Proceedings of the International Conference on Computational Linguistics* (Beijing), 1353–1361.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ebling, Battisti, Kostrzewa, Pfütze, Rios, Säuberli and Spring. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fpsyg.2022.707630

Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian

Dominique Brunato, Felice Dell'Orletta and Giulia Venturi*

ItaliaNLP Lab, Institute for Computational Linguistics "A. Zampolli" (ILC-CNR), Pisa, Italy

In this paper, we present an overview of existing parallel corpora for Automatic Text Simplification (ATS) in different languages focusing on the approach adopted for their construction. We make the main distinction between manual and (semi)–automatic approaches in order to investigate in which respect complex and simple texts vary and whether and how the observed modifications may depend on the underlying approach. To this end, we perform a two-level comparison on Italian corpora, since this is the only language, with the exception of English, for which there are large parallel resources derived through the two approaches considered. The first level of comparison accounts for the main types of sentence transformations occurring in the simplification process, the second one examines the results of a linguistic profiling analysis based on Natural Language Processing techniques and carried out on the original and the simple version of the same texts. For both levels of analysis, we chose to focus our discussion mostly on sentence transformations and linguistic characteristics that pertain to the morpho-syntactic and syntactic structure of the sentence.

Keywords: text simplification, aligned corpora, linguistic complexity, Italian language, corpus construction

OPEN ACCESS

Edited by:
Sigal Uziel-Karl,
Achva Academic College, Israel

Reviewed by:
Marianna Marcella Bolognesi,
University of Bologna, Italy
Andras Kornai,
Computer and Automation Research Institute (MTA), Hungary

****Correspondence:***
Dominique Brunato
dominique.brunato@ilc.cnr.it

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

Received: *10 May 2021*
Accepted: *07 January 2022*
Published: *08 March 2022*

Citation:
Brunato D, Dell'Orletta F and Venturi G (2022) Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian. Front. Psychol. 13:707630. doi: 10.3389/fpsyg.2022.707630

1. INTRODUCTION AND MOTIVATION

Automatic Text Simplification (ATS) is the Natural Language Processing (NLP) task aimed at reducing linguistic complexity of texts, especially at the lexical and syntactic levels, while preserving their original content (Bott and Saggion, 2014; Shardlow, 2014; Alva-Manchego et al., 2020a). It has long attracted the attention of different research communities that address the issue of generating a simplified version of an input text from two broad perspectives. The first perspective can be called "machine-oriented," in that it conceives the task as a pre-processing step useful to improve the performance of other NLP tasks by providing an input that is easier to analyze for, e.g., syntactic parsing (Chandrasekar et al., 1996), Machine Translation (Štajner and Popović, 2016), Information Extraction (Klebanov et al., 2004; Niklaus et al., 2016), or Semantic Role Labeling (Vickrey and Koller, 2008) systems.

The second perspective is "human-oriented" and is concerned with the production of texts equally accessible for a wide variety of readers, also including less-skilled ones. In this respect, ATS is tightly intertwined with the Automatic Readability Assessment (ARA) task (Collins-Thompson, 2014), as they both share the primary objective of identifying and modeling properties of linguistic complexity within text according to cognitive and psycho-linguistic evidence on human sentence

processing. However, while ARA allows discriminating between difficult-to-read and easy-to-read texts, ATS takes a step further which is to automatically convert the former into the latter. In this sense, it can be viewed as a sort of monolingual translation process. The two tasks have also in common the idea that linguistic complexity is a property highly related to the final reader. From this perspective, ATS studies have been devoted to define strategies to simplify texts for people with cognitive disabilities (Bott and Saggion, 2014), language impairments, e.g., aphasia (Carroll et al., 1998), dyslexia (Rello et al., 2013), or deafness (Inui et al., 2003), limited literacy skills (Aluísio et al., 2008) or a low proficiency level in a second language (Petersen and Ostendorf, 2007; Crossley et al., 2012). As most of our daily interaction with society, government, and other institutions require access to information conveyed by text, making content accessible also for this kind of people is ultimately a strategy to promote social inclusion. With this purpose in mind, numerous initiatives have been pursued in recent years by private and public organizations aimed at developing educational and assistive technologies for the benefit of human readers. A well-known one concerns the most popular free online encyclopedia, Wikipedia, which has been offering an easy–to–read version of its contents since 2003, although only limited to the English language[1].

Another coarse distinction that has characterized the field of ATS is related to the methodological framework. Independently from the machine- or human-oriented purpose for which ATS is carried out, early computational approaches to the task were mostly based on hand-crafted rules targeting specific complex constructions informed by theoretical, cognitive, and computational linguistics literature. A special focus was put on the syntactic level with specific rules addressing the simplification of relative clauses, appositions, subordination, passive voice (Chandrasekar et al., 1996; Siddharthan, 2002, 2010). These methods can reach high precision and can potentially account for the maximum linguistic information, but they are extremely time-consuming and tend to cover only a few lexical and syntactic constructions. To overcome these drawbacks, much of current research is shifting toward data-driven techniques, most recently based on neural sequence-to-sequence models (Nisioi et al., 2017), which can automatically acquire from corpora the transformations occurring to sentences when they are manually simplified. The first fundamental requirement to allow the application of data-driven methods is the availability of large–scale monolingual parallel corpora, i.e., corpora containing the original and the simplified version of the same text possibly aligned at the sentence level. This is the main goal of the initiative launched by Wikipedia devoted to leveraging volunteers to create pages more easy-to-be read by everyone, both children and adults who are learning English. The efforts resulted in the first resource used for ATS purposes that include portions of English (EW) Wikipedia automatically aligned to Simple English Wikipedia (SEW) ones. It represented a benchmark because of its size and availability, mostly used by many ATS studies allowing both the development of machine learning algorithms and the evaluation of the quality of the automatic simplification results (Alva-Manchego et al., 2020b).

However, beyond size, there is a widespread consensus in the community that the quality of the simplified language resources is of fundamental importance and needs to be investigated in detail. This issue can be easily explained by drawing parallelism with the translation studies: in order to assess how good are simplified language resources, there should be a *native simplified–language speaker* who masters the target and the source language thus guaranteeing the quality of the translated texts (Siddharthan, 2014). But it is not the case since, as we introduced above, simplified texts should be differently conceived to reach a large number of target readers. This means that different varieties of simplified language exist with different characteristics that should be learned by ATS systems. In light of these considerations, the suitability of the Wikipedia-based resource for ATS applications has become quite debated (Bach et al., 2011; Woodsend and Lapata, 2011; Yasseri et al., 2012) and brought to the creation of new resources for the English language. To date, the most important one is the Newsela corpus (Xu et al., 2015), which contains original sentences extracted from newspaper articles and their simplified version at different readability levels by professional editors. Unlike parallel corpora derived from the original and simple English Wikipedia, the manually performed simplification guarantees very consistent alignments at the sentence level, as well as high quality of the linguistic transformations undergone by the original sentences (Xu et al., 2015; Scarton et al., 2018). As shown in the following section, during the last few years other resources have been built, but the Newsela corpus represents nowadays the most comprehensive benchmark for the English language since it includes the widest and qualitatively checked range of simplification operations.

The number of outcomes in ATS research, both in terms of large-scale corpora and available systems, mostly concerns a highly-resource language like English. The picture is different in other languages, for which a preliminary step in the development of ATS systems has been typically represented by the collection of monolingual parallel corpora either from scratch, asking experts (e.g., teacher, translators, speech therapists) to simplify texts for a specific target or aligning existing resources of original and simplified versions of the same texts. It is worth noticing that, with only a few exceptions (see Section 2 for details), these corpora are smaller than the ones available for English and this has made it hardly feasible to use them as training data for pure ATS systems based on machine learning methods. It is the reason why similar resources were primarily collected to be used as reference corpora to identify the most frequent simplification operations occurring in manually-simplified texts or to train rule-based systems covering limited sets of simplification phenomena, as in the case of, e.g., Italian (Barlacchi and Tonelli, 2013), Basque (Aranzabe et al., 2013), French (Brouwers et al., 2014), and German (Suter et al., 2016). Despite the low amount of data, useful methodological insights come from the Statistical Machine Translation (SMT) community. Monolingual SMT approaches were firstly tested for example for the Portuguese language (Specia, 2010), and more recently, neural MT architectures are

[1]https://simple.wikipedia.org/wiki/Main_Page

used to develop neural text simplification systems for Italian (Aprosio et al., 2019), German (Sauberli et al., 2020), and Russian (Dmitrieva and Tiedemann, 2021). In these cases, the problem of data scarcity is alleviated with data augmentation techniques or methods to generate new synthetic data that are added to the original too small training corpora. However, this poses several issues related to the quality of the training data and, as a consequence, of the resulting automatically simplified sentences.

1.1. Our Contribution

In the scenario outlined so far, the main purpose of this study is to provide a deeper investigation on the effects that different approaches to ATS resources may have on the simplified corpora, with a special emphasis on languages other than English. Specifically, we focus on Italian since it is the only language, among the less-resourced ones, for which not only there are resources representative of the manual and the (semi)automatic approach, but they are also large enough to allow a significant comparison. Rather than proposing a method to assess the impact of the approaches used to build a resource on the performance of ATS systems, our investigation intends to assess whether and to what extent different approaches to the construction of ATS resources can affect the linguistic characteristics of simplified sentences to their original versions. In this sense, the purpose of the study is to contribute to the discussion on the quality of ATS resources by providing a fine-grained analysis of the linguistic phenomena characterizing parallel corpora available for the same language but built with different approaches. The investigation is twofold and it consists, on the one hand, in the study of the distribution of transformations (mainly syntactic structures) that the original sentences undergo when they are simplified; on the other hand, it is based on the results of a linguistic profiling analysis carried out with NLP-based techniques that allow comparing the original/simplified sentences by accounting for the distribution of a wide range of morpho-syntactic and syntactic characteristics automatically extracted from the linguistically annotated pairs of sentences. The first type of comparison requires the contribution of human experts who explicitly annotate the considered resource for a set of sentence transformations, while the second one is completely automatic.

The rest of the paper is organized as follows: Section 2 reports a survey of existing ATS corpora for different languages, drawing a main distinction according to the approach adopted in their construction, i.e., manual vs. (semi-)automatic. Section 3 describes the simplification operations that have been detected across these corpora to classify the major structural transformations involved in the process of sentence simplification. Our two-fold methodology devoted to comparing original and simplified sentences is described in Section 4, where we show how it can be used to study which kind of sentence transformations characterize a manually and an automatically derived corpus and whether the distribution of the linguistic characteristics of the corpus is correlated with the building approach.

2. MANUAL VS. (SEMI-)AUTOMATIC APPROACH TO CORPORA CONSTRUCTION

In this section, we present an overview of the main ATS corpora existing for different languages. As shown in **Table 1**, we classified them into two main typologies: the ones manually built and the ones built adopting automatic (or semi-automatic) methods. For each corpus, we further provide the following details: the textual genre, the language, the target user, and the dimension.

As it can be seen, the majority of corpora are manually derived and they were developed assuming a human-oriented perspective. Within this group, the only exception is represented by Collados (2013), who created a parallel corpus of 3,000 original and manually simplified sentences for the Spanish language to be used as a pre–processing steps for NLP tasks.

2.1. Manual Approach

According to this approach, the simplification process is typically performed by qualified linguists expert in text simplification or professionals (e.g., speech therapists, translators) and it starts from a previously chosen text, which is considered complex for a specific readership and simplified by the expert to improve user's comprehension. The original and simplified versions of this text are then paired either automatically or manually. As anticipated in Section 1, as a universal *native simplified–language speaker* does not exist (Siddharthan, 2014), manually-built corpora differ with respect to the expertise of the "human simplifier." With the intent of grouping together human simplifiers sharing a common methodology to text simplification, in the literature, it has been drawn the main distinction between two manual simplification strategies: the "structural" and the "intuitive" one, according to Allen's definition (Allen, 2009). The former uses predefined graded lists (covering both word and structural levels) or traditional readability formulas based on shallow proxies of language complexity, the latter is dependent on the professional's intuition on which sentence transformations are needed to reduce the linguistic complexity of a text for a given user, e.g., the author's teaching experience and personal judgments about the comprehension ability of learners. These two strategies have been explicitly taken into account for the collection of Italian and Basque corpora. Specifically, Brunato et al. (2015) compiled two Italian corpora of aligned complex/simple sentences: the *Terence* corpus, a collection of 1,060 pairs of sentences produced by a pool of experts (i.e., linguists and psycholinguists) in the framework of a past EU project Terence[2] as representative of the "structural" strategy. The experts manually simplified 32 short novels for children aged 7–11 affected by text comprehension difficulties by following a predefined guideline tackling the simplification at three separate textual dimensions, i.e., global coherence, local cohesion and, lexicon/syntax. The intuitive strategy is represented by the so–called *Teacher* corpus, a collection of 24 pairs of original/simplified texts, collected from specialized educational websites that offer free resources for teachers on different textual genres, from famous Italian

[2]https://cordis.europa.eu/project/id/257410

TABLE 1 | Monolingual parallel corpora of original/simplified sentences classified with respect to the type of approach adopted for their construction, the language, the textual genre, the target (GP, general purpose; CHI, children; LL, language learners; L2LL, L2 language learners; PLI, people with language impairments; PLL, people with low literacy level; NLP, NLP tasks; CS, crowd-sourcing), and the size of corpus.

Manual approach			
Language	**Textual genre**	**Target**	**Dimension**
ENG (Pellow and Eskenazi, 2014)	Everyday documents	GP	200 sentence pairs
ENG (Xu et al., 2015)	Newspapers	CHI	56,037 original sentences
ENG (Barzilay and Elhadad, 2003)	Encyclopedia Britannica	CHI	2,600 easy-to-read documents
ENG (Allen, 2009)	Classroom materials	LL	178,967 of simplified words
ENG (Petersen and Ostendorf, 2007)	Newspapers	LL	2,539 original sentences
ENG (Xu et al., 2016)	Wikipedia	CS	2,359 original sentences
ENG (Alva-Manchego et al., 2020a)	Wikipedia	CS	2,359 original sentences
Many (Orasan et al., 2013)	Miscellanea	PLI	320 original sentences
SPA (Bott and Saggion, 2014)	Newspapers	PLI	145 simplified sentences
SPA (Collados, 2013)	Newspapers	NLP	300 simplified sentences
FRE (Brouwers et al., 2014)	Narrative texts	L2LL	83 original sentences
FRE (Grabar and Cardon, 2018)	Encyclopedic, scientific, clinical texts	GP	4,596 sentence pairs
FRE (Gala et al., 2020)	L1 student materials	PLI	52,704 tokens
DAN (Klerke and Søgaard, 2012)	Newspapers	L2LL	3,701 document pairs
POR (Caseli et al., 2009)	Newspapers	PLL	2,116 original sentences
POR (Aluísio et al., 2008)	Popular science articles	PLL	882 original sentences
GER (Klaper et al., 2013)	Websites	PLI	7,755 original sentences
GER (Sauberli et al., 2020)	Newspapers	L2LL	3,616 sentence pairs
JPN Goto et al. (2015)	Newspapers	L2LL	2,885 sentence pairs
EUS Gonzalez-Dios et al. (2017)	Popular science articles	L2LL	227 original sentences
RUS Dmitrieva and Tiedemann (2021)	Literary texts	L2LL	69,737 original sentences
ITA Tonelli et al. (2016)	Administrative texts	GP	157 original sentences
ITA Brunato et al. (2015)	Children's literature	PLI	1,060 sentence pairs
ITA Brunato et al. (2015)	Educational material	L2LL	1,356 original pairs
(Semi)Automatic Approach			
ENG Kauchak (2013)	Wikipedia	GP	167K sentence pairs
ENG Kajiwara and Komachi (2016)	Wikipedia	GP	492,993 sentence pairs
ENG Zhu et al. (2010)	Wikipedia	GP	108,016 sentence pairs
ENG Narayan et al. (2017)	Wikipedia	GP	5,546 original sentences
ENG Woodsend and Lapata (2011)	Wikipedia	GP	14,831 sentence pairs
ENG Botha et al. (2018)	Wikipedia	GP	1,004,944 original sentences
ENG Pavlick and Callison-Burch (2016)	Miscellanea	CS	4.5 million of simplifying paraphrase rules
ITA Tonelli et al. (2016)	Wikipedia	GP	530 original sentences
FRE Brouwers et al. (2014)	Wikipedia	L2LL	72 original sentences
FRE Cardon and Grabar (2020)	Wikipedia	GP	297,494 sentence pairs
ITA Brunato et al. (2016)	Web corpus	GP	63,000 sentence pairs

novels to handbooks for high school on diverse subjects (e.g., history, geography). In this case, texts were simplified by a school teacher who aimed at making them easier–to–read for students, especially L2 learners. Similarly, for the Basque language, Gonzalez-Dios et al. (2017) gathered a collection of documents belonging to the scientific popularization domain manually simplified by a court translator according to easy-to-read guidelines (as representative of the "structural" strategy) and by a teacher based on her/his experience (as representative of the "intuitive" strategy).

For what concerns the textual genre dimension, corpora of newspaper articles are largely predominant. This is the case of the *Newsela* corpus (Xu et al., 2015) which includes a collection of news articles, each one manually simplified at four distinct levels. The multiple simplification versions are a direct consequence of the primary aim of the corpus, which was conceived to help teachers prepare curricula that match the English language skills required at each grade level. A similar purpose is shared by Goto et al. (2015), who simplified a corpus of news for learners of Japanese as a second language. Newswire is also the main genre

of the 200 articles contained in the Spanish corpus by Bott and Saggion (2014), which were manually simplified by professionals for people affected by Down's syndrome. Similarly, Klerke and Søgaard (2012) compiled a corpus of aligned original/simplified news intended for adults with cognitive impairment and adult learners of Danish; Caseli et al. (2009) made available a corpus of 104 Brazilian newspaper articles, developed in the framework of the *PorSimples* project, which were manually simplified adopting two different types of approaches to simplification, i.e., *natural* and *strong*, to attend the needs of people (adults and children) with different levels of literacy. Beyond newswire texts, other corpora contain texts from a mix of genres, and for some languages also genre–specific resources have been released covering, e.g., biomedical texts (Grabar and Cardon, 2018), science articles as in Gonzalez-Dios et al. (2017), children's literature, as in Allen (2009) and Brunato et al. (2015), and administrative texts such as the *SIMPITIKI-Admin* corpus, a collection of 157 Italian sentences extracted from documents issued by the Trento Municipality to rule building permits and kindergarten admittance and manually simplified by a linguist expert in text simplification (Tonelli et al., 2016).

All these examples also highlight that the target reader population is a fundamental factor in driving the construction of ATS corpora. As we introduced at the beginning of this paper, people who need assistive technologies and learners at different levels of proficiency, both in the native and in a second language, are the two main groups of readers targeted by ATS. The first group is addressed for example by the FIRST project (Orasan et al., 2013), a project launched in 2011 aimed at developing tools and resources to assist people with autism spectrum disorders (ASD). A main outcome of the project was a multilingual corpus of 25 texts (for a total of 320 sentences) available in three different languages (English, Spanish, Bulgarian) and covering a wide range of topics, which were manually simplified by professionals (teachers, psychologists, speech and language therapists, and psychiatrists) to improve reading and comprehension of ADS people. The second main group of target readers was taken into account, for example, in the framework of the PorSimples project for the Portuguese language (Aluísio et al., 2008; Caseli et al., 2009) devoted to developing text tools for promoting digital inclusion and accessibility mainly for native language people with low literacy levels. For German, a newspaper corpus compiled by Sauberli et al. (2020) was manually simplified for second language learners at two language levels, B1 and A2 (based on the Common European Framework of Reference for Languages standard), and exploited to test state-of-the-art neural machine translation techniques. More recently, specific attention to children facing problems in reading was paid by Gala et al. (2020), who compiled a corpus of literary and scientific materials available for students in French primary schools and manually simplified them at different linguistic levels, i.e., lexical, morpho-syntactic, syntactic, and discourse. Their final goal was to test the simplified materials with poor-reading and dyslexic children to assess the impact of simplification operations in reducing reading errors.

A survey of very recent works dealing with the collection of parallel corpora also highlights the exploitation of a new manual strategy, in addition to the two main ones discussed so far (i.e., structural and intuitive). It relies on crowd-sourcing techniques to collect human simplified versions of original sentences and so far it has been applied to the English Wikipedia pages. The two most notable corpora obtained in this way are TURKCORPUS (Xu et al., 2016) and ASSET (Alva-Manchego et al., 2020a), differing at the level of rewriting operations adopted to obtain the human simplified version of the original sentences. Another motivation driving the introduction of this new strategy concerns the use of the collected resources as benchmarks to define new, more human-oriented, metrics able to evaluate the ability of ATS systems to generate easy-to-read sentences that not only preserve the original meaning but also sound fluent and simply according to the correlation with human judgments.

2.2. (Semi-)Automatic Approach

This second type of approach to building ATS corpora gathers together strategies that even with minor differences allow searching, in a large reference resource, texts which are equivalent in meaning but different at the level of linguistic complexity. The multiple versions of these texts can be written independently, so they are not strictly parallel; they are aligned in a later stage, generally at the sentence level, using word-level (Barzilay and Elhadad, 2003; Nelken and Shieber, 2006; Coster and Kauchak, 2011) or sentence-level (Bott and Saggion, 2011) similarity functions. As already mentioned in the introduction, the most typical example of automatically (or semi-automatically) derived corpora were obtained by aligning articles from the standard and the simple version of the English Wikipedia. It is the case of the corpora described by Kauchak (2013), by Kajiwara and Komachi (2016), and by Zhu et al. (2010) for English. A similar attempt has been pursued by Brouwers et al. (2014), who semi-automatically aligned 20 articles from the French Wikipedia with their equivalents in Vikidia, a small online encyclopedia intended for young readers which gathers more accessible articles than Wikipedia, both in terms of language and content[3].

Different use of the Wikipedia resource for the construction of monolingual parallel corpora has been shown by Woodsend and Lapata (2011), and more recently by Tonelli et al. (2016) and by Botha et al. (2018). They started from the same assumption that the multiple revisions underlying Wikipedia articles can be used to collect reliable resources for ATS purposes. The resulting corpora are made of aligned sentence pairs where the complex sentence is the one occurring in a previous version of a Wikipedia article and the simple one is the outcome of all edit operations involving a sentence split (Botha et al., 2018), or only those marked by the Wikipedia's contributor as a simplification or grammatical correction (Woodsend and Lapata, 2011; Tonelli et al., 2016).

Beyond Wikipedia data, other text sources were explored. For instance, Narayan et al. (2017) started from the dataset described in Gardent et al. (2017), where each item consists of a set of RDF triples (corresponding to an abstract meaning representation) and one or more texts that verbalize the triples and contain one or more sentences. They used it to automatically create the WEBSPLIT corpus, a very large dataset of 1,066,115

[3]http://fr.vikidia.org

distinct pairs of complex/simple sentences where each distinct complex sentence is associated with multiple (2–7) simpler versions sharing the same abstract meaning representation. The rationale was to segment each verbalization of the same original (complex) sentence into multiple sentences to build a resource useful to be used to train an ATS system able to perform a subset of sentence transformations involved in the simplification process, namely sentence split and rephrase. The construction of a resource containing instances of a single type of simplification rule was Pavlick and Callison-Burch (2016)'s goal, who semi-automatically built the *Simple Paraphrase Database* a corpus of 4.5 million lexical paraphrases devoted to the development of lexical ATS systems.

A further implementation of the automatic approach was proposed by Brunato et al. (2016). To our knowledge, it is the first one not based on Wikipedia and not specifically devised for the English language. Considering the scarcity of a large quantity of aligned data in languages other than English, the authors proposed an approach that does not rely on any kind of pre-existing parallel corpora: this makes such an approach highly scalable and language agnostic. The authors followed the intuition that sentences conveying the same information but with a different level of complexity can be extracted from a large–scale, monolingual corpus of heterogeneous genres and domains, such as the web corpus. According to these premises, they conceived a semi-unsupervised methodology to detect and pair sentences with overlapping lexicon (thus, guaranteeing that the pair had the same meaning) but showing structural transformations of different types. The two sentences of the same pair were then ranked for linguistic complexity, which was calculated according to the score automatically assigned by a readability assessment tool, i.e., the "simple" sentence of the pair was the one assigned with a lower readability score. The approach was tested for the Italian language, resulting in a corpus, named PaCCSS–IT (*Parallel Corpus of Complex-Simple Sentences for ITalian*), which contains about 63,000 pairs of complex/simple sentences.

3. AN OVERVIEW OF THE MAIN SIMPLIFICATION OPERATIONS ACROSS CORPORA

As we mentioned in the introduction, parallel corpora have a strong application value since they primarily serve as training and evaluation resources for ATS systems, being them rule-based or, especially in more recent years, based on deep learning. This means that if an automatic system learns a model of "simple" language from available training corpora, we expect that it would apply it to generate new simplified texts. Therefore, comparing how the complex and the simple version of a sentence vary is crucial to assess the quality of these resources and their suitability for ATS purposes. In this section, we thus take a closer look at the most representative types of transformations occurring in the corpora previously described. To perform this analysis, we moved from the observation that many of the existing parallel resources were annotated with a set of simplification rules aimed at identifying the specific types of linguistic phenomena changing between the original and the simplified version of a sentence. However, as pointed out by Bott and Saggion (2014), these phenomena are not necessarily comparable since the classifications of simplification operations can vary according to language- and genre-specific properties or to the needs of the expected readership. This is the reason why we focus here on a representative set, without the ambition to report an exhaustive list. In particular, we chose to analyse the ones that better fit with the main focus of our investigation, thus considering those rules that have an impact on the morpho-syntactic and syntactic structure of the simplified sentences, and we deliberately paid less attention to the numerous types of transformations affecting the use of words at the lexical level. Indeed, although lexical properties represent a very important and well-investigated aspect in text simplification (Paetzold and Specia, 2017), accounting for them would open an orthogonal but different area of research, with several other variables to be considered, largely inspired by cognitive models on the organization of the mental lexicon, such as word frequency, word length, familiarity, concreteness, imageability, age of acquisition (Cutler, 1983). Moreover, while the English language can rely on large-scale machine readable dictionaries curated by experts where entries are labeled for many of these properties [see, e.g., the Medical Resource Council (MRC) Psycholinguistic Database (Wilson, 1988)], less-resourced languages have to cope with the unavailability, or rather poorer coverage, of such lexical databases; this makes it necessary to supply them with more traditional resources, such as word frequency and word familiarity lists drawn from large corpora. As described in Section 4.2, the only lexical aspect we took into account in this study as a marker of lexical complexity is word frequency considering a representative lexical resource of Italian.

Split: breaking down long sentences into shorter ones is probably one of the most studied simplification operations, also from the point of view of its computational treatment (Siddharthan, 2002; Collados, 2013; Narayan et al., 2017). Typical candidates for splitting are coordinate clauses (introduced by coordinating conjunctions, colons, or semicolons), subordinate clauses (e.g., non-restrictive relative clauses, as in the example below), appositive and adverbial phrases. Nevertheless, some real examples detected across ATS parallel corpora showed that human experts do not exploit the split rule as much as expected (Brunato et al., 2015; Xu et al., 2015). A complex sentence may be judged more comprehensible than a simple one, for instance because it contains a subordinate clause that provides the necessary background information to understand the main clause.

O: Mamma Gorilla sembrava completamente distrutta per le cure che dava al suo vivace cuccioletto Tito, **che stava giocando vicino alle grosse sbarre di acciaio che circondavano il recinto**. [lit. Mummy Gorilla looked completely worn out from looking after her lively baby, Tod, **who was playing by the thick steel bars that surrounded the enclosure.**]

S: Mamma Gorilla sembrava proprio distrutta per le cure che dava al suo vivace cuccioletto Tito. *Tito stava giocando vicino*

alle grosse sbarre di acciaio che erano intorno alla loro area. [lit. Mummy Gorilla looked completely worn out from looking after her lively baby Tod. *Tod was playing by the thick steel bars that surrounded the enclosure*.]

(Terence corpus, Brunato et al., 2015)

Merge: this operation joins two (or more) original sentences into a unique sentence, thus it has the opposite effect of a split. Despite adding more propositions per sentence could make it harder to process (Kintsh and Keenan, 1973), such an operation sometimes allows writers to avoid unnecessary repetition, to clarify the logical order of events with explicit connectives, and to improve sentence variety, with a positive effect on the reader's comprehension.

O: **Gli ebrei debbono consegnare le biciclette. Gli ebrei non possono salire in tram, gli ebrei non possono più andare in auto**. [lit. **Jews have to hand over their bikes. Jewish are not allowed to get in the tram. Jewish are not allowed to drive cars**.]
S: *Gli ebrei non possono più andare in bicicletta, non possono salire in tram e non possono andare in auto*. [lit. *Jews have to hand over their bikes, are not allowed to get in the tram and are not allowed to drive cars*.]

(Teacher corpus, Brunato et al., 2015)

Reordering: another possible strategy to simplify texts consists in changing the position of the elements in a sentence, possibly yielding the unmarked order of that language, which is associated with easier comprehension and earlier acquisition (Slobin and Bever, 1982). As shown by the examples here reported, reordering can affect single words, phrases, or entire clauses.

O: **In 1962**, Steinbeck received the Nobel Prize for Literature.
S: Steinbeck won the Nobel Prize in Literature *in 1962*.

(English Wikipedia corpus, Coster and Kauchak, 2011)

O: Aireak hegazkinaren inguruan duen jokabidea zoruak alda dezake, **hegaldia oso baxua denean**. [lit. The soil can change the behavior that the air has around the plane, **when the flight is very low**.]
S: *Hegaldia oso baxua denean* zoruak hegazkinaren inguruko airearen jokabidea alda dezake. [lit. *When the flight is very low*, the soil can change the behavior that the air has around the plane.]

(Basque corpus, Gonzalez-Dios et al., 2017)

O: Ringraziandola per la sua cortese attenzione, **resto in attesa di risposta**. [lit. Thanking you for your kind attention, **I look forward to your answer**.]
S: *Resto in attesa di una risposta* e ringrazio vivamente per l'attenzione. [lit. *I look forward to your answer* and I thank you greatly for your attention.]

(PaCCSS–IT corpus, Brunato et al., 2016)

Insert: the process of simplification may even result in a longer sentence because of the insertion of words or phrases that provide additional information to the original sentence and possibly reduce the inference load of a text. Despite the cognitive literature suggests reducing the inference load of a text, especially when it targets less skilled or low-knowledge readers (Ozuru et al., 2009), it is difficult to predict what an author will add to the original sentence to make it clearer. The sentence can be elliptical, i.e., syntactically compressed, and the difficulty depends on the ability to retrieve the missing arguments, which are then made explicit as a result of the simplification. The following examples show a case of insertion of the main verb and a subject, respectively. The insertion of a subject has to be intended as the transformation of a covert subject into a lexical noun phrase, which is an option available in null-subject languages (e.g., Italian).

O: Escuela Segura, un compromiso municipal con la protecciońn integral de los escolares. lit. [Safe School: a municipal promise for the full protection of school kids.]
S: Escuela Segura *es* un programa municipal para la protecciońn de los escolares. [lit. Safe School *is* a municipal promise for the full protection of school kids.]

(Spanish corpus, Bott and Saggion, 2014)

O: Curiosa com'era, si avvicinò per osservarla meglio, prima timidamente, poi con più coraggio. [lit. Curious as she was, (she) moved closer to watch it better, shyly at first, than more courageously.]
S: Curiosa com'era, *Ernesta* si avvicinò per guardarla meglio, prima con paura, poi con più coraggio. [lit. Curious as she was, *Ernestine* moved closer to watch it better, timidly at first, than more courageously.]

(Terence corpus, Brunato et al., 2015)

Delete: removing redundant information has proven to be another effective strategy to simplify a text. Like insertion, it is difficult to predict which words could be removed, although we can predict that simplified sentences would contain fewer adjunct phrases (e.g., adverbs or adjectives). In null-subject languages, a particular case of deletion is the substitution of a lexical noun phrase subject with a covert pronoun, especially when the latter points to a referent which is highly prominent in the context, as shown by the last example.

O: The crust and **underlying relatively rigid** mantle make up the lithosphere.
S: The crust and mantle make up the lithosphere.

(English Wikipedia corpus, Coster and Kauchak, 2011)

O: Poi la nuvoletta aggiunse, con molta tristezza, che purtroppo **lei** stava partendo, come ogni anno. [lit. Then the little cloud said, with much sadness, that unfortunately **she** was leaving, like every year.]
S: La nuvoletta, un po' triste, disse che stava partendo, come tutti gli anni. [lit. The little

cloud, a bit sad, said that (she) was leaving, like every year.]

(Terence corpus, Brunato et al., 2016)

Transformations: the macro-class of sentence transformations (or sentence changes) is articulated into more fine-grained operations representative of specific linguistic phenomena, which affect the lexical, morpho-syntactic, or syntactic structure. Here follows a list of major sentence transformations:

– *Lexical substitution*: the substitution of a complex word with an easier synonym is a feasible way to reduce the linguistic complexity of a text. Much research in ATS has been done on lexical simplification trying to automatize this process, e.g., by relying on electronic resources, such as WordNet (DeBelder and Moens, 2010), word frequency lists (Drndarevic et al., 2012) or simpler paraphrases (Kriz et al., 2018)[4]. However, corpus analysis highlighted that a complex word can be substituted by a multi-word paraphrase rather than a synonym or even explained by a gloss, especially for the technical terms.

O: Dopo la scoperta del cadavere di Lily Kimble, la polizia comincia a interessarsi al caso. Dal canto loro, i Reed capiscono che finchè l'assassino non sarà **al fresco**, la loro vita sarà in pericolo. [lit. After the discovery of Lily Kimble's body, the police begin to take an interest in the case. For their part, the Reeds realize that until the killer is not **in the pen**, their lives are in danger.]

S: Dopo la scoperta del cadavere di Lily Kimble, la polizia comincia a interessarsi al caso. Dal canto loro, i Reed capiscono che finchè l'assassino non sarà *in carcere*, la loro vita sarà in pericolo. [lit. After the discovery of Lily Kimble's body, the police begin to take an interest in the case. For their part, the Reeds realize that until the killer is not *in jail*, their lives are in danger.]

(Simpitiki-Admin corpus, Tonelli et al., 2016)

O: Poiché era indeciso su quale fosse il bidone giusto, chiese ad un **passante** di indicargli il bidone dove buttare la carta. [lit. Since he was unsure about which dustbin was the right one, he asked to a **passer-by** to point him the dustbin to throw paper.]

S: Poiché non sapeva quale fosse il bidone giusto, chiese ad un *signore che passava* dove era il bidone per la carta. [lit. Since he didn't know which dustbin was the right one, he asked to a *man who was walking* where the paper dustbin was.]

(Terence corpus, Brunato et al., 2015)

–*Anaphoric phenomena*: under this class, there were marked transformations involving the substitution of a referential pronoun in the original sentence with its full lexical antecedent (a definite noun phrase or a proper noun) or vice versa. As shown by the examples that follow, in several cases a transformation at one level triggers rearrangements at other levels of the sentence, which are necessary for the grammatically of the simplified output. For instance, replacing a direct or indirect object pronoun (which is preverbal in some languages like Italian) with its full lexical antecedent not only changes the grammatical category of the element but also affects syntactic order, since full nominal objects follow the verb.

O: Il passante gli spiegò che, per arrivare al bidone, **doveva contare ben 5 bidoni a partire dal semaforo**. [lit. The passer-by explained him that, to get to the dustbin, **he had to count exactly 5 dustbins starting from the traffic light.**]

S: Il signore spiegò a Ugolino che *doveva contare 5 bidoni a partire dal semaforo*, per arrivare al bidone della carta. [lit. The man explained Little Hugh that *he had to count 5 dustbins starting from the traffic light* to get to the wastepaper dustbin.]

(Terence corpus, Brunato et al., 2015)

O: Anche Federico Fellini, all'epoca ancora giovane e sconosciuto, aiuterà **Aldo Fabrizi** nella sceneggiatura. [lit. Also Federico Fellini, still young and little known at that time, will help **Aldo Fabrizi** in the script.]

S: Anche Federico Fellini, all'epoca ancora giovane e sconosciuto, *lo* aiuterà nella sceneggiatura. [lit. Also Federico Fellini, still young and little known at that time, will help *him* in the script.]

(Simpitiki-admin corpus, Tonelli et al., 2016)

– *Nominalization phenomena*: these transformations target a nominalization (or a support verb construction), which is replaced by the simple verb from which it derives, or conversely, a simple verb which is changed into a nominal phrase headed by the corresponding derivative noun.

O: Il giorno **della partenza**, i bambini salutarono i loro genitori durante la colazione. [lit. On the day of their **parents' departure**, the children said their goodbyes to their parents over breakfast.]

S: Il giorno *in cui i genitori partirono*, i bambini li salutarono durante la colazione. [lit. The day *that their parents left*, the children said them goodbye over breakfast.]

(Terence corpus, Brunato et al., 2015)

O: Un computer simile è presente nel film Pixar Wall-e: il computer della nave spaziale axiom **si chiama** AUTO e governa la nave assieme al capitano. [lit. A similar computer is present in the Pixar Wall-e movie: the axiom spacecraft computer **is called** AUTO and rules the ship with the captain].

S: Un computer simile è presente nel film Pixar Wall-e: il computer della nave spaziale axiom *di nome* AUTO e governa la nave assieme al capitano. [lit. A similar computer is present in the Pixar Wall-e movie: the axiom spacecraft computer *with name* AUTO and rules the ship with the captain].

(Simpitiki-Admin corpus, Tonelli et al., 2016)

– *Voice*: as a result of simplification, a passive sentence may be converted into an active one or vice versa. The former

[4]See Paetzold and Specia (2017) for an up–to–date survey of methods and resources for lexical simplification.

transformation is more expected when humans simplify a text because passive sentences are considered as more complex according to language acquisition data in typical (Maratsos, 1974) and atypical populations, e.g., deaf children (Volpato, 2010). Plain language guidelines also recommend preferring active than passive voice[5]. Yet, the "passivization" rule may still be productive in specific textual typologies like administrative texts, where the author of the simplification can prefer not only to keep but even to insert, a passive, to avoid more unusual syntactic constructs (such as impersonal sentences).

O: **Se trata de un proyecto (…) que coordina el trabajo (…) de las delegaciones municipales de Educacioń y Seguridad.** [lit. **It consists of a project that coordinates the work of the city's education and security delegations**.]
S: *El proyecto está coordinado por las delegaciones municipales de Educacioń y Seguridad*. [lit. *The project is coordinated by the city's education and security delegations*.]

(Spanish corpus, Bott and Saggion, 2014)

O: **Rinvia, quindi, il seguito dell'esame ad altra seduta**. [lit. **He/she postpones, thus, the follow-up examination to other hearing** .]
S: *Il seguito dell'esame viene rinviato ad altra seduta*. [lit. *The follow-up examination is postponed to other hearing*.]

(PaCCSS–IT corpus, Brunato et al., 2016)

– *Verbal features*: The simplification of a text can also alter the distribution of verbal features (such as mood, tense, and person), especially in languages with a rich inflectional paradigm. These features indeed are involved in text complexity as proven by literature on readability assessment, in which verbal features appear upon the variables that discriminate between easy- and difficult-to-read texts (Attardi et al., 2009; Dell'Orletta, 2009; Bautista et al., 2011; Dell'Orletta et al., 2011; François and Fairon, 2012; Narayan and Gardent, 2014). Moreover, for some categories of readers (e.g., second-language learners) some inflections are more difficult to master; thus simplified texts targeting these readers should exhibit more common and less literary tenses than those used in the original texts (Brouwers et al., 2014).

O: Tali elementi **dovranno** supportare e giustificare le scelte progettuali operate. [lit. Such elements **will have** to support and justify project decisions].
S: Tali elementi *devono* supportare e giustificare le scelte progettuali operate. [lit. Such elements *have* to support and justify project decisions.]

(Simpitiki-Admin corpus, Tonelli et al., 2016)

4. A TWO-LEVEL COMPARISON

As described in Section 2, the majority of works on ATS corpora have been devoted to study text simplification with a special interest for the target audience and the expertise of the human simplifier, as well as for the influence of the textual genre. Less attention has been paid to inspecting whether and to what extent the methodology adopted to build ATS resources can affect the structure and the linguistic characteristics of the simplified sentences. To shed light on this under-investigated perspective, in this section we present a methodology based on a fine-grained linguistic analysis aimed at understanding (1) in which respect complex and simple/simplified texts vary and (2) whether and how the observed changes may depend on a manual or a (semi-)automatic approach.

The analysis has been carried out at two levels: concerning the distribution of the simplification operations described in the previous section and of multi-level linguistic features automatically extracted from texts and modeling a wide range of morphosyntactic and syntactic phenomena involved in sentence complexity. In the first case, the aim is to figure out whether some operations are specific only to a given building approach or the same type of simplification operations occurs independently. This would suggest that some sentence transformations should be considered as more "fundamental" to yield a simpler text. Consequently, if an automatic text simplification system learns a model of "simple" language from corpora containing these transformations, we expect that it would apply them to newly generated texts. Secondly, the analysis of the automatically extracted linguistic phenomena is meant to detect similarities or differences between the original and simplified sentences: this type of information may also represent a valuable contribution to evaluate the quality of a resource to be used in real ATS scenarios.

For both levels of analysis, we chose to focus on Italian since this is the only language, except English, for which there are quite large resources derived through the manual and automatic approach. Specifically, we relied on three corpora: the *Terence* and *Teacher* corpora, representative of the manual approach, and more specifically of the "structural" and the "intuitive" approach, and PaCCSS-IT, as representative of the automatic approach. However, the whole methodology is in principle transferable to multiple languages since it relies on sentence transformations shared by many ATS resources, as discussed in Section 3, and on a multi-lingual approach to the automatic extraction of linguistic phenomena, as it is detailed in Section 4.2.

4.1. Distribution of Simplification Operations

As described in Brunato et al. (2016), the approach devised to create PaCCSS-IT was aimed at collecting sentence pairs sharing the same meaning but with a few structural transformations possibly affecting the level of linguistic complexity. To control for meaning preservation, the pairs were selected to share many of their words, namely: the same lemmas tagged as nouns, verbs, and adverbs for what concerns open-class categories, and the same personal pronouns and negative adverbs, for what concerns closed-class categories. Given these strict requirements, some of the simplification operations described in Section 3 are not allowed in PaCCSS-IT by definition: for example, since the alignment between the complex and simple sentence has a

[5]See, for instance, the Wikipedia guidelines for writing articles in Simple English. https://simple.wikipedia.org/wiki/Wikipedia:How_to_write_Simple_English_page.

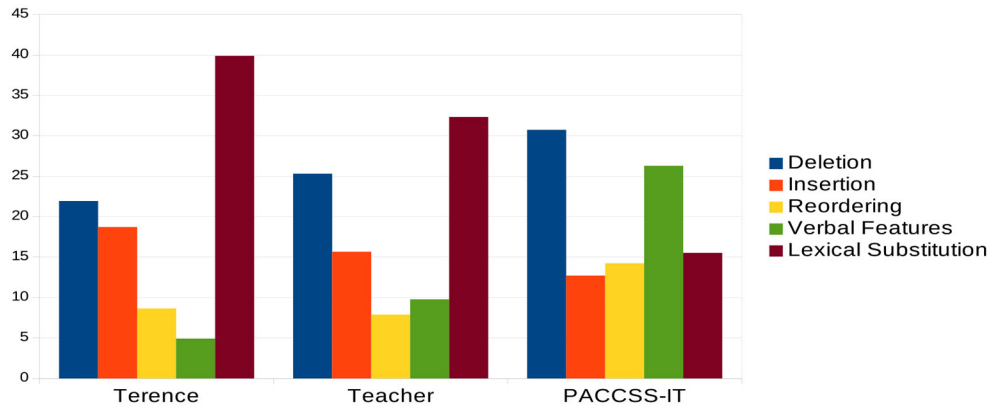

FIGURE 1 | Distribution of simplification operations across corpora.

correspondence 1:1, the operations involving the splitting and merging of sentences are not possible; likewise, transformations involving the replacement of a verb with a deverbal noun (i.e., nominalization) do not occur since the aligned sentences must contain the same nouns and verbs. To allow a comparative analysis between the manually and automatically collected corpora, we thus selected only the subset of rules potentially occurring in all corpora—which correspond to the macro-level operations—and used them to manually annotate the whole *Terence* and *Teacher* corpora and a comparable portion of PaCCSS-IT corresponding to a subset of 921 paired sentences. All corpora were annotated by two undergraduate students in computational linguistics, who received preliminary training lessons on the simplification rules covered by the annotation tagset. All their annotations were verified by two of the authors of the paper.

Figure 1 reports the distribution of simplification operations in the examined corpora. In line with the criteria adopted to build PaCCSS-IT, it can be noted that the automatic approach is mostly characterized by structural transformations (i.e., deletions, insertions, and reordering), which cover almost 60% of the whole amount of operations. On the contrary, operations involving the substitution of words are more frequently exploited in the manual process of sentence simplification (*Terence*: 39.89%; *Teacher*: 32.33%; PaCCSS-IT: 15.52%). Among the operations modifying the syntactic structure, the deletion and the insertion of linguistic material (words or phrases) have a similar frequency across the three corpora. As expected, removing redundant information turned out to be the most frequent sentence transformation (*Terence*: 21.94%; *Teacher*: 25.32%; PaCCSS-IT: 30.74%). Interestingly, the automatic approach intercepts a wider set of simple sentences where phrases and words have been reordered thus possibly showing a more canonical word order which is easier to process (Diessel, 2005; Futrell et al., 2015). A further operation that clearly differentiates the automatic approach from the manual one affects verbal features. Given that the complex and the simple sentences in PaCCSS-IT share the same verb lemmas, the higher percentage of this operation (26.30%) concerns only transformations of the same lemma which changes concerning, e.g., mood, tense, person, and verbal voice. Notably, this is in line with the predominant role played by structural transformations characterizing the automatic approach: for example, the passive/active alternation or the occurrence of implicit vs. explicit moods imply syntactic modifications of the whole sentence.

4.2. Distribution of Linguistic Phenomena

While in the previous section we focused on the comparison between the manual and automatic approach concerning the distribution of simplification rules, here we examine the distribution of a wide set of linguistic phenomena characterizing the complex and simple sentences of each corpus. To conduct this analysis we rely on the methodology of "linguistic profiling," an NLP-based framework of analysis in which a large number of counts of linguistic features extracted from linguistically annotated corpora are used as a text profile and can then be compared to average profiles of texts (or groups of texts) to identify those that are similar in terms of the profiled features (van Halteren, 2004; Montemagni, 2013). This methodology, which is rooted in the seminal works by Douglas Biber who first introduced the multidimensional approach to linguistic analyses of genre variation (Biber, 1993, 1995), has been successfully used in a variety of application scenarios, all focused on the "form" rather than the content of texts: from automatically modeling the developmental patterns in child language acquisition (Lu, 2009; Lubetich and Sagae, 2014) and the evolution of written language competence in school learners' (Weiss and Meurers, 2019; Miaschi et al., 2021), to the prediction of behavioral and cognitive impairments based on the detection of relevant linguistic markers from clinical tests (Roark et al., 2007; Prud'hommeaux et al., 2011); also, in the context of computational sociolinguistics, it has been used for studying variations related to the social dimension of language (Nguyen et al., 2016) or for modeling stylometric characteristics of authors or author groups (Daelemans, 2013).

For our analysis, we relied on Profiling-UD (Brunato et al., 2020), a tool recently introduced that implements the assumptions of linguistic profiling and specifically conceived for corpora annotated according to the Universal Dependencies

TABLE 2 | Overview of the linguistic features used for linguistic profiling.

Level of annotation	Linguistic feature
Raw Text	**Raw text properties**
	Sentence length
	Word length
Vocabulary	**Vocabulary richness**
	Type/Token ratio for words and lemmas
POS tagging	**Morphosyntactic information**
	Distribution of UD and language–specific POS
	Lexical density
	Inflectional morphology
	Inflectional morphology of lexical verbs and auxiliaries
Dependency parsing	**Verbal predicate structure**
	Distribution of verbal heads and verbal roots
	Verb arity and distribution of verbs by arity
	Global and local parsed tree structures
	Depth of the whole syntactic tree
	Average length of dependency links and of the longest link
	Average length of prepositional chains and distribution by depth
	Clause length
	Order of elements
	Relative order of subject and object
	Syntactic relations
	Distribution of dependency relations
	Use of subordination
	Distribution of subordinate and principal clauses
	Average length of subordination chains and distribution by depth
	Relative order of subordinate clauses with respect the main clause

(UD) framework[6]. UD is an ongoing project aimed at developing corpora with a cross-linguistically consistent annotation for many languages, to facilitate multilingual parser development, cross-lingual learning, and parse research from a language typology perspective (De Marneffe et al., 2016). The choice of relying on UD-style annotation makes the process of feature extraction language-independent, as similar phenomena are annotated according to a common annotation scheme at morpho-syntactic and syntactic levels of analysis. The tool performs a two-stage process: linguistic annotation and linguistic profiling. The annotation of the text(s) is carried out by UDPipe (Straka et al., 2016) using the available UD model(s) for the input language. The automatically annotated text(s) are used as input to the further step, performed by the linguistic profiling component defining the rules to extract and quantify the formal properties.

Profiling-UD allows the computation of a wide set of features encoding a variety of morpho-syntactic and syntactic properties of text, which are reliable predictors in a variety of scenarios, from stylometric analyses to genre classification. For our specific purposes, we considered only a subset of them, namely those that have been used in the literature to assess the readability level of texts (Collins-Thompson, 2014) or to investigate which of these features correlate with human judgments on sentence complexity (Brunato et al., 2018). Specifically, they range from superficial ones, such as the average length of words and sentences, to morpho-syntactic information concerning the distribution of parts-of-speech (POS)[7] and the inflectional properties of verbs, to more complex aspects of syntactic structure deriving from the whole parse tree and specific sub-trees (e.g., subordinate clauses). A sub-set of features, which we considered particularly relevant for this study, is reported in **Table 2** where they are grouped into main linguistic phenomena and distinguished according to the level of annotation from which they derive.

As an example, we report in **Figure 2**, a graphical representation of the output of the linguistic annotation in UD format for a sentence of the *Terence* corpus. By applying Profiling-UD on this input sentence, we can observe, for instance, that the sentence contains 16 tokens and these tokens are on average 4.93 characters long. Concerning the distribution of POS, there is 31.25% of nouns, 6.25% of verbs, and 37.5% of determiners, among others. At syntactic level, since we only have one verbal root represented by the main predicate [i.e., *salutarono*, (greeted)], the arity value is 4 corresponding to the four dependents attached to the head [i.e., *giorno* (day) and *colazione* (breakfast), both bearing the role of oblique modifiers, and *bambini* (children) and *genitori* (parents) with the role of subject and object, respectively]. Moreover, the average length of dependency links is 2.38 and the longest link has a value of 7, which corresponds to the number of words separating the dependent *giorno* from its head *salutarono*.

In addition to the set of features extracted by Profiling-UD, we calculate some extra ones characterizing the lexical profile of a sentence, in terms of the percentage distribution of words belonging to the *Basic Italian Vocabulary* (*BIV*) by DeMauro (2000). This is a reference lexical resource for contemporary Italian covering about 7,000 words considered as highly familiar to Italian native speakers. As described in Chiari and De Mauro (2014), VdB derives from a combination of statistical criteria used to select lemmas (both grammatical and content words) mainly based on a frequency list of written Italian, which was subsequently enriched with a frequency list of spoken Italian, and experimental evaluations with primary school pupils. Since its first edition, the final resource is internally subdivided into three usage repertories: "fundamental words" (*FO*), i.e., highest frequency words that cover about 90% of all written and spoken texts), "high usage words" (*HU*), i.e., about 6% of the subsequent high-frequency words) and "high availability words" (*HA*), relatively lower frequency words referring to everyday life whose detection is not based on textual statistical resources but is derived from psycholinguistic insights experimentally verified.

[6]www.universaldependencies.org

[7]Note that for the specific purpose of this study we considered both the UPOS, i.e., the set of Parts-Of-Speech defined by the UD project, and the XPOS, i.e., the set of POS specific for the Italian language that provides a finer-grained morpho-syntactic categorization of tokens.

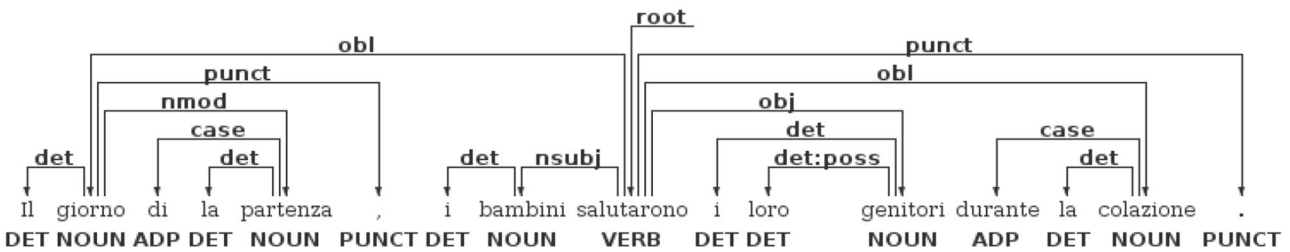

FIGURE 2 | A sentence from the corpus linguistically annotated in the UD-format.

Using Profiling-UD, we thus proceeded to automatically parse all texts of our corpora up to the level of syntactic annotation and to convert them into a rich feature-based representation. For each feature, also including the distribution of lexicon of the *Basic Italian Vocabulary*, we then assessed whether the average distribution in the relative corpus changes significantly between the original and the simplified sentences using a non-parametric statistic text, i.e., the Wilcoxon signed-rank test. **Tables 3**, **4** report the distribution of an excerpt of features whose variation between complex and simple sentences resulted to be statistically significant for at least one of three considered corpora according to the Wilcoxon signed-rank test.

As a general remark, we can observe that *Teacher* is the corpus with the highest difference between the values of features characterizing the original and simplified sentences. Namely, the average difference between the two versions of the corpus is 4.14 considering the features extracted from raw text, vocabulary, and morpho-syntactic level of annotation, and 2.94 for what concerns the features referring to syntactic phenomena. On the contrary, these differences drop to 1.18 and 0.80 for the sentences contained in PaCCSS-IT, and to 0.72 and 0.52 for the *Terence*'s sentences. This represents the first evidence that the intuitive manual approach yields the sharpest linguistic differences among the considered approaches. However, the amount of features whose value varies significantly between the two counterparts is higher in the corpus automatically built, meaning that, differently from the two types of manual approaches, this method intercepts a large variety of linguistic phenomena that make a sentence easier to read.

If we go more into detail, it can be noted that the simple sentences of each corpus are shorter, in terms of the average number of tokens per sentence. This could be expected since sentence length has been considered as a shallow proxy of sentence complexity and is widely used by traditional readability assessment formulas. The different average length between the original and the simple sentence is also influenced by textual genre: while previous studies on genre variation have shown that narrative prose is characterized by longer sentences (see e.g., Biber, 1995 among others), sentences from the web tend instead to be shorter (Santini, 2007). It follows that the sentences contained in *Terence* and *Teacher* were originally longer and thus the effect of simplification is much more evident. Conversely, the original sentences in PaCCSS-IT already had an average length that is much lower than the average sentence length of the Italian language (i.e., 20–25 tokens) and thus were not greatly modified concerning this parameter. Among raw text features, the average length of words appears to be less concerned with sentence simplification. The three corpora do not vary greatly and length variation of *Teacher*'s words results to be not statistically significant.

Focusing on lexical features, we can see that the use of a more frequent lexicon in simple sentences mostly characterizes PaCCSS-IT and *Terence*. In particular, the percentage distribution of all unique words (types) in the *Basic Italian Vocabulary* (*BIV*) increases in the collection of simple sentences except the *Teacher* corpus, for which the distribution does not change significantly. This is specifically the case of the simple sentences collected with the automatic approach adopted in PaCCSS-IT, which have a higher percentage of BIV concerning the corresponding original sentences (almost 5%). Note that according to the strategy devised to automatically build this resource, such an increase of simple lexicon is mainly concerned with the substitution or the insertion of content and functional words annotated with a Parts-Of-Speech not shared by the original/simple pair (which, we recall here, are necessarily nouns, verbs, numerals, personal pronouns, and negative adverbs). For instance, the following pair contains only a minimal variation, affecting the adverb of time ("conclusivamente"), which is substituted with a more frequent one, with the same meaning, but contained in the BIV.

O: Propone **conclusivamente** di esprimere parere favorevole. [lit. He suggests *eventually* giving a favorable opinion]
S: Propone *infine* di esprimere un parere favorevole. [lit. He suggests *lastly* giving a favorable opinion.]

However, if we focus on the internal classification of BIV into the usage repertories of "fundamental" (FO, very frequent words), "high usage" (HU, frequent words) and "high availability" (HA, relatively lower frequency words referring to everyday life), the simplified sentences contained in the *Teacher* corpus report the highest increase of fundamental lexicon.

The approach adopted in the construction of PaCCSS-IT influences also the distribution of morpho-syntactic characteristics deriving from linguistic profiling. Specifically, while the frequency of nouns and verbs is necessarily the same, the way nominal and verbal modification are expressed changes in the complex and simple version of the pairs: simple sentences have more adjectives, articles, and determiners, but fewer

TABLE 3 | Distribution of the raw text, lexical, and morpho-syntactic features in the complex and simple set of sentences for the three corpora.

Feature	Terence			Teacher			PaCCSS–IT		
	Compl	Simp	Diff	Compl	Simp	Diff	Compl	Simp	Diff
Raw text features									
Sentence length	19.92	18.61	**1.31**	21.25	18.56	**2.70**	8.97	8.0	**0.97**
Word length	4.89	4.80	**0.09**	4.74	4.70	0.04	4.70	4.54	**0.16**
Lexical features									
% BIV	75.59	77.31	**−1.72**	78.53	77.77	0.75	72.19	77.08	**−4.88**
% FO	78.14	79.82	**−1.67**	80.21	82.73	**−2.51**	75.03	75.76	**−0.73**
% HU	13.08	12.15	**0.93**	11.98	9.68	**2.30**	20.19	19.82	**0.37**
% HA	8.77	8.03	**0.74**	7.81	7.60	0.21	4.78	4.42	**0.36**
Type/Token ratio	0.942	0.941	-0.001	0.921	0.913	0.008	0.97	0.99	**−0.02**
Morpho–syntactic features									
Morpho–syntactic information									
Adjectives	5.87	5.97	−0.01	5.34	5.11	0.23	5.74	7.90	**−2.15**
Adverbs	6.82	6.97	−0.15	7.62	6.73	0.89	12.26	9.95	**2.31**
Articles	8.79	8.73	0.07	8.24	8.69	−0.45	11.04	12.71	**−1.67**
Conjunctions—coordinating	3.57	3.76	**−0.19**	3.98	4.72	**−0.74**	2.66	3.45	**−0.79**
Conjunctions—subordinating	1.75	2.16	**−0.41**	1.73	1.09	**0.64**	0.32	0.30	0.02
Prepositions	13.31	12.50	**0.81**	10.77	10.51	0.25	5.98	6.21	**−0.23**
Pronouns	5.33	5.04	**0.28**	17.69	17.15	**0.54**	7.23	4.14	**3.09**
Pronouns—relative	0.87	0.81	0.06	0.85	0.28	**0.57**	0.27	0.1	**0.17**
Pronouns—clitic	2.78	2.61	**0.17**	5.25	2.74	**2.51**	2.47	1.60	**0.87**
Punctuation	11.57	11.54	0.03	15.53	15.52	0.01	20.5	15.13	**5.36**
Numbers	1.07	0.91	**0.15**	2.25	2.47	−0.22	2.25	2.47	**−0.22**
Lexical density	0.59	0.60	**−0.00**	0.58	0.62	**−0.04**	0.61	0.60	**0.00**
Inflectional morphology									
Indicative mood	61.23	64.4	**−3.17**	57.14	70.87	**−13.73**	68.14	68.31	**−0.17**
Participial mood	6.95	4.63	**2.32**	3.95	2.84	1.11	3.65	2.42	**1.23**
Gerundive mood	3.44	2.62	**0.83**	1.56	–	**1.56**	0.46	0.04	**0.42**
Infinitive mood	15.98	17.64	**−1.66**	22.1	19.67	**2.43**	12.04	11.65	0.39
Subjunctive mood	1.00	0.57	**0.42**	0.58	–	**0.58**	0.78	0.05	**0.73**
Conditional mood	0.19	0.12	0.07	0.84	0.18	0.66	3.34	0.001	**3.33**
Present tense	6.21	4.74	**1.47**	43.31	90.19	**−46.87**	79.18	80.91	**−1.73**
Imperfect tense	50.66	52.97	**−2.31**	16.39	0.82	**15.57**	2.89	4.29	**−1.40**
Past tense	40.98	39.97	**1.01**	27.45	–	27.45	1.33	1.57	**−0.24**
2 person, singular	0.44	0.51	−0.07	2.77	0.37	**2.4**	0.60	0.44	**0.15**
3 person, singular	64.9	66.09	−1.19	48.59	53.31	−4.72	62.31	58.13	**4.18**
1 person, plural	–	0.09	−0.09	2.95	4.13	−1.18	1.51	1.84	**−0.33**
2 person, plural	–	–	–	0.42	0.32	0.10	0.30	0.19	**0.11**
3 person, plural	18.69	19.14	−0.45	13.86	16.55	−2.69	8.12	7.83	**0.28**

Statistically significant variations with respect to the Wilcoxon signed-rank test at $p < 0.05$ are bold.

adverbs, punctuation marks, and pronouns. Among the latter, clitic pronouns are much more frequent in the original than in the simplified version (2.47 vs. 1.60). A possible explanation, which is consistent with qualitative observations on the corpus is that, in many cases, a sentence with an impersonal verb construction introduced by a clitic pronoun, is paired with a simple one expressing the same meaning but with a personal verb form, as in the following example.

O: Non **si può fare** di ogni erba un fascio. [lit. **It is not possible** to bundle everybody together in one big bunch.]

S: Però non *possiamo fare* di tutta l'erba un fascio. [lit. But *we can't bundle* everybody together in one big bunch.]

On the contrary, at the level of POS distribution, the differences between the original and simplified sentences contained in the corpora manually simplified are less sharp. Interestingly, the main exception is represented by the *Teacher* corpus and it affects the distribution of clitic pronouns. With this respect, this corpus shares a similar tendency with the one automatically derived, that is a very consistent drop in the use of clitic pronouns, which is even sharper. Again, this can be due to an editing operation

TABLE 4 | Distribution of the syntactic features in the complex and simple set of sentences for the three corpora.

Feature	Terence			Teacher			PaCCSS–IT		
	Compl	Simp	Diff	Compl	Simp	Diff	Compl	Simp	Diff
Syntactic relations									
Subjects	6.37	6.87	**−0.50**	5.25	6.71	**−1.46**	7.65	6.94	**0.71**
Objects	4.77	5.12	**−0.34**	4.93	4.92	0.01	1.82	1.90	**−0.07**
Subjects—passive	0.20	0.15	0.04	0.17	0.08	0.09	0.66	0.95	**−0.29**
Use of subordination									
Subordinate clauses	51.86	51.41	**0.45**	53.08	47.35	**5.73**	50.083	50.078	**0.005**
Depth of "chains" of subord.	0.39	0.41	**−0.03**	0.39	0.27	0.12	0.05	0.06	**−0.001**
Post-main subordinates	40.54	43.42	**−2.87**	42.27	27.98	**14.29**	3.28	4.17	**−0.89**
Global and local parsed tree structure									
Parse tree depth	5.80	5.56	**0.24**	5.10	4.46	**0.64**	2.85	2.70	**0.15**
Dependency links length	2.07	2.03	**0.04**	2.29	2.12	**0.17**	1.76	1.63	**0.13**
Length of the longest link	8.01	7.48	**0.53**	9.24	7.32	**1.93**	3.81	3.31	**0.5**
Verbal arity	1.93	1.95	−0.02	1.85	1.91	−0.05	2.09	2.08	**0.01**
Depth of prepositional "chains"	1.06	1	**0.06**	0.90	0.91	−0.01	0.44	0.41	**0.02**
Order of elements									
Pre-verbal subjects	71.07	71.35	−0.28	51.16	61.71	**−10.55**	50.58	43.85	**6.73**
Post-verbal subjects	9.59	11.32	−1.73	16.62	15.51	1.11	15.36	14.37	**0.99**
Pre verbal objects	5.72	5.54	0.17	8.29	3.24	**5.05**	2.03	1.38	**0.65**

Statistically significant variations with respect to the Wilcoxon signed-rank test at $p < 0.05$ are bold.

changing an impersonal with a personal form but also to the insertion of full lexical items that are substitute of clitics as in the following example, where the two instances of the pronoun *ci* (lit. "us") were substituted with two different lexical referents:

O: Non poter mai andar fuori mi opprime, e ho una gran paura che **ci** scoprano e **ci** fucilino. [lit. Never being able to go outside oppresses me, and I'm so afraid that they discover **us** and shot **us**].

S: E' triste non andare fuori. Ho una gran paura delle SS: possono scoprire *il rifugio* e uccidere *tutti noi*. [It's sad not to go outside. I'm so afraid of the SS: they can find out about *the shelter* and kill *all of us*.]

Interestingly, among the characteristics extracted from the morpho-syntactic level of annotation, those concerning the inflectional morphology of verbs undergo the main changes. In this case, the main differences can be observed in the *Teacher* corpus, which in its simplified version contains a higher percentage of verbs at the indicative mood and the present tense, and conversely a lower amount of verbs in their imperfect and past tense. Variations of the verbal morphology also occur at the level of person and number of verbs: *Teacher* simplified sentences contain more verbs at the third singular person, at the first and at the third plural person, than their original counterparts.

When we consider the syntactic features, we can observe that all corpora are characterized by noteworthy changes. Among the considered syntactic relations, subjects are more frequent in the simplified sentences of all corpora even if this is, in particular, the case of the corpus representative of the intuitive simplification. This is in line with what was observed concerning the insertion of explicit arguments that allow reducing the inference load of null-subject sentences. Consider for example the following excerpt where the nominal subject *Ernesta* was inserted in the simple counterpart of the original sentence:

O: Curiosa com'era, si avvicinò per osservarla meglio, prima timidamente, poi con più coraggio. [lit. Curious as she was, (she) moved closer to watch it better, shyly at first, than more courageously.]

S: Curiosa com'era, *Ernesta* si avvicinò per guardarla meglio, prima con paura, poi con più coraggio. [lit. Curious as she was, *Ernestine* moved closer to watch it better, timidly at first, than more courageously.]

Similar observations hold for the distribution of direct objects, even if in this case *Teacher* is the only corpus where they are almost stable. According to Profiling-UD, the distribution of subordinate clauses is calculated as the percentage distribution of main vs subordinate clauses, where the latter are identified based on the UD guidelines that distinguish four different types[8]. We considered this feature since the use of subordination is a broadly studied marker of structural complexity, for example for text simplification purposes (Bott and Saggion, 2014). This is particularly the case of post-verbal subordinate clauses that, according to Miller and Weinert (1998), are easier to read than subordinates preceding the main clause. However, if this is confirmed in *Terence* and PaCCSS-IT, an opposite trend can be observed in the *Teacher* corpus, which are characterized by a lower amount of post-verbal subordinates.

[8]https://universaldependencies.org/u/overview/complex-syntax.html#subordination

The set of features intercepting the global and local syntactic structure of the sentence has been considered since it includes aspects typically related to length factors and correlate with processing difficulty (Frazier, 1985). This is the case of parse tree depth, which can be indicative of increased sentence complexity as stated by, to mention only a few, Yngve (1960) and Gibson (1998). Both Lin (1996) and Gildea and Temperley (2010) showed that the syntactic complexity of sentences can be predicted with measures based on the length of dependency links since long-distance constructions cause cognitive load. As **Table 4** shows, in all three corpora the values of these features tend to decrease. This trend concerns specifically the *Teacher* corpus and the minimization of dependency links, and in particular of the longest one in the sentence. The feature was computed as the linear distance between the head and its dependent in terms of tokens. Simple sentences tend to contain shorter dependency links thus increasing the capacity of working memory (Miller, 1956).

The relative order of subject and direct object with respect to the verb has been shown to be harder to process especially in free word-order languages. The position of these core verb arguments is a language-specific property typically connected with "marked" or "unmarked" word orders and thus highly related with sentence complexity or "abnormality," to put in Haspelmath's words (Haspelmath, 2006). In addition, according to the "adaptability hypothesis," some linguistic properties systematically lead to varying processing behavior in typologically distinct languages (Yadav et al., 2020). For example, if the Subject-Verb-Object (SVO) order is frequent in a language, the sentence processing should become easier when the order is preserved. Accordingly, since Italian is an SVO language, word order variation involving the relative ordering of subjects and objects may yield marked and more difficult to process sentences. This trend is particularly evident in the *Teacher* simplified sentences where the distribution of subjects preceding the verb is higher and the amount of direct objects preceding the verb is lower. We chose to exemplify this phenomenon by showing the following pair of sentences since it represents a quite typical example of how syntactic modifications may be strictly connected with what has been observed at the morpho-syntactic level of analysis:

O: E' un gran miracolo che io non abbia rinunciato a tutte le mie speranze perché esse sembrano assurde. **Le** conservo ancora, nonostante tutto, perché continuo a credere nell'intima bontà dell'uomo. [lit. It is a great miracle that I have not given up all my hopes because they seem absurd. **Them** I still keep, in spite of everything, because I continue to believe in the intimate goodness of man.]

S: Ma io ho fiducia, ho ancora *speranza* perchè credo nella bontà dell'uomo. [lit. But I have faith, I still have *hope* because I believe in the goodness of man.]

The two original sentences were merged into a unique sentence where the clitic pronoun *le* ("them"), operating as a direct object and, according to the Italian grammatical rules, preceding the main verb *conservo* ("keep"), was deleted. This yields a more simple sentence characterized by a canonical order of the verb's core arguments, where the referent of the pronoun, *speranza* ("hope"), serves as a direct object and follows the corresponding verb.

We think that the main results of this detailed analysis are two-fold. On the one hand, they show that the type of approach to construct a text simplification resource has an impact on the amount and strength of linguistic phenomena characterizing the original and simple corpora. Namely, the automatic approach yields a wider range of variations between the two versions, while the manual approach allows generating simple sentences that undergo stronger changes in terms of specific linguistic features. On the other hand, the analysis also highlighted several differences between the structural and intuitive manual approaches, which make the structural one more similar to the automatic one. This is the case for example of the simple lexicon, which increases more in the simple version of Terence and PaCCSS-IT than in *Teacher*, or of specific features related to the verbal morphology, which undergo more changes in the *Teacher*.

5. CONCLUSION

Text simplification is a topic that has received considerable attention in recent years in the computational linguistics community where it is more and more approached as a monolingual machine translation task. In this respect, the availability of large monolingual parallel data is a fundamental requirement to develop systems able to automatically infer the type of transformations that should be applied to "translate" the complex source text into the simple target text. In this article we addressed this topic from a quite less investigated point of view, i.e., the approach adopted in the construction of resources used for Automatic Text Simplification. We identified two main categories of resources, i.e., those simplified by human experts and those obtained through (semi)automatic approaches. We first surveyed existing ATS corpora for multiple languages from this perspective and we then focused on the Italian language for an empirical investigation based on available corpora. This choice was motivated by the fact that this is the only language, among less-resourced ones, for which not only there are resources representative of the "manual" and the "(semi)automatic" approach, but they are also large enough to allow a significant comparison. Comparing three different TS corpora aligned at the sentence level, we thus carried out a deep linguistic analysis of the main sentence phenomena characterizing the original and simple versions to investigate whether and to what extent the approach adopted for their construction can affect the internal composition of the resulting resources both in terms of undergone sentence transformations and linguistic properties.

From the point of view of the distribution of the main sentence transformations detected across paired corpora, we observed that some tendencies are shared in the simplification process, such as the deletion of redundant elements and the insertion of either words or phrases that make explicit missing or implicit information of a sentence. The perspective of linguistic profiling focused on the examination of the distribution of a wide variety of lexical, morpho-syntactic, and syntactic properties of the

sentence has allowed us to better characterize the differences and similarities between automatically and manually derived corpora. In particular, we observe that a human-based simplification affects a relatively less number of linguistic features, yet the values of these features change more from the original to the simplified version. This is particularly the case of the "intuitive" approach, represented in our study by a corpus of texts simplified by teachers according to their feeling of the students' needs and without following predefined simplification rubrics. Instead, the automatically-derived corpus affects the whole structure of the sentence, though it yields less prominent differences for each feature.

Based on the results of both analyses we can also conclude that the automatic approach can deliver resources that not only are large enough to train an ATS system but, more importantly, that exhibit effective transformations toward the use of a simpler language. We believe that this is an important outcome especially if we consider that the method underlying the construction of PaCCSS-IT is language-agnostic and potentially transferable to other languages lacking available resources. However, we are also aware that an analysis of the internal linguistic composition of ATS resources is not sufficient to guarantee their quality. In this sense, a highly important aspect is represented by the human evaluation aimed at testing the effect of simplification in terms of text comprehension. This is also emphasized by the recent efforts of the ATS community toward the development of new metrics, such as SARI (Xu et al., 2016) and ASSET (Alva-Manchego et al., 2020a), which have become new standards for evaluating the quality of automatically simplified texts in light of their high correlation with human judgements of simplicity gain. In this respect, one possible way to expand this study could be to evaluate a subset of the corpus here considered concerning human judgments: this would allow us to assess if the current method of ranking sentence pairs for complexity is in line with the perception of sentence complexity by readers. Moreover, the multilingual perspective on linguistic profiling offered by a tool like Profiling-UD would provide a way to compare the distribution of the same linguistic phenomena within parallel corpora not only derived through different approaches but also across languages.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: http://www.italianlp.it/resources/.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Allen, D. (2009). A study of the role of relative clauses in the simplification of news texts for learners of English. *System* 37, 585–599. doi: 10.1016/j.system.2009.09.004

Aluísio S. M., Specia, L., Pardo, T. A., Maziero, E. G., and de Mattos Fortes, R. P. (2008). "Towards Brazilian Portoguese automatic text simplification systems," in *Proceedings of the Eighth ACM Symposium on Document Engineering* (Sao Paulo: ACM), 240–248.

Alva-Manchego, F., Martin, L., Bordes, A., Scarton, C., Sagot, B., and Specia, L. (2020a). "ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations," in *Proceedings of ACL 2020*, 4668–4679.

Alva-Manchego, F., Scarton, C., and Specia, L. (2020b). Data-driven sentence simplification: survey and benchmark. *Comput. Linguist.* 46, 135–187. doi: 10.1162/coli_a_00370

Aprosio, A. P., Tonelli, S., Turchi, M., Negri, M., and Di Gangi, M. A. (2019). "Neural text simplification in low-resource conditions using weak supervision," in *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation* (Minneapolis, MN), 37–44.

Aranzabe, M. J., De Ilarraza, A. D., and Gonzalez-Dios, I. (2013). Transforming complex sentences using dependency trees for automatic text simplification in Basque. *Proc. Leng. Nat.* 50, 61–68.

Attardi, G., Dell'Orletta, F., Simi, M., and Turian, J. (2009). "Accurate dependency parsing with a stacked multilayer perceptron," in *Proceedings of EVALITA 2009 - Evaluation of NLP and Speech Tools for Italian 2009* (Reggio Emilia).

Bach, N., Gao, Q., Vogel, S., and Waibel, A. (2011). "Tris: A statistical sentence simplifier with loglinear models and margin-based discriminative training," in *Proceedings of 5th International Joint Conference on Natural Language Processing* (Chiang Mai: Asian Federation of Natural Language Processing).

Barlacchi, G., and Tonelli, S. (2013). "ERNESTA: a sentence simplification tool for children's stories in Italian," *Proceedings of the 14th Conference on Intelligent Text Processing and Computational Linguistics* (Samos: ACM), 476–487.

Barzilay, R., and Elhadad, N. (2003). "Sentence alignment for monolingual comparable corpora," in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (Sapporo: ACM), 25–32.

Bautista, S., León, C., Hervás, R., Gervás, P. (2011). "Empirical identification of text simplification strategies for reading-impaired people," in *Proceedings of the European Conference for the Advancement of Assistive Technology* (Maastricht: AAATE), 1–8.

Biber, D. (1993). "The multi-dimensional approach to linguistic analyses of genre variation: an overview of methodology and findings," in *Computers and the Humanities*, Vol. 26, eds N. Ide and D. Walker (Berlin: Kluwer Academic Publishers), 331–345.

Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge, NY: Cambridge University Press.

Botha, J. A., Faruqui, M., Alex, J., Baldridge, J., Das, D. (2018). Learning to split and rephrase from Wikipedia edit history," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (Brussels: Association for Computational Linguistics), 732–737.

Bott, S., and Saggion, H. (2011). "An unsupervised alignment algorithm for text simplification corpus construction," in *Proceedings of the Workshop on Monolingual Text-To-Text Generation, Co-Located With ACL 2011* (Porland, OR).

Bott, S., and Saggion, H. (2014). Text simplification resources for Spanish. *Lang. Resour. Eval.* 48, 93–120. doi: 10.1007/s10579-014-9265-4

Brouwers, L., Bernhard, D., Ligozat, A. L., François, T. (2014). "Syntactic sentence simplification for French," in *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (Gothenburg: Association for Computational Linguistics), 47–56.

Brunato, D., Cimino, A., Dell'Orletta, F., and Venturi, G. (2016). "PaCCSS–IT: a parallel corpus of complex–simple sentences for automatic text simplification," in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)* (Austin, TX), 351–361.

Brunato, D., Cimino, A., Dell'Orletta, F., Venturi, G., and Montemagni, S. (2020). "Profiling-UD: a tool for linguistic profiling of texts," in *Proceedings of The 12th Language Resources and Evaluation Conference* (European Language Resources Association), 7145–7151.

Brunato, D., De Mattei, L., Dell'Orletta, F., Iavarone, B., and Venturi, G. (2018). "Is this sentence difficult? Do you agree?," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (Brussels), 2690–2699.

Brunato, D., Dell'Orletta, F., Venturi, G., and Montemagni, S. (2015). "Design and annotation of the first Italian corpus for text simplification," in *Proceedings of The 9th Linguistic Annotation Workshop* (Denver, CO), 31–41.

Cardon, R., and Grabar, N. (2020). "French biomedical text simplification: when small and precise helps," in *Proceedings of the 28th International Conference on Computational Linguistics* (Barcelona), 710–716.

Carroll, J., Minnen, G., Canning, Y., Devlin, S., and Tait, J., (1998). "Practical simplification of english newspaper text to assist aphasic readers," in *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology, Association for the Advancement of Artificial Intelligence* (Madison WI: AAAI).

Caseli, H., Pereira, T., Specia, L., Pardo, T., Gasperin, C., and Aluísio, S. (2009). "Building a Brazilian Portuguese parallel corpus of original and simplified texts," in *Proceedings of the 10th Conference on Intelligent Text Processing and Computational Linguistics* (Mexico City), 59–70.

Chandrasekar, R., Doran, C., and Srinivas, B. (1996). "Motivations and methods for text simplification," in *Proceedings of the International Conference on Computational Linguistics* (Copenhagen), 1041–1044.

Chiari, I., and De Mauro, T. (2014). "The new basic vocabulary of Italian as a linguistic resource," in *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014* (Pisa), 113–116.

Collados, J. C. (2013). Splitting complex sentences for natural language processing applications: building a simplified Spanish corpus. *Proc. Soc. Behav. Sci.* 95, 464–472. doi: 10.1016/j.sbspro.2013.10.670

Collins-Thompson, K. (2014). "Computational assessment of text readability," in *Recent Advances in Automatic Readability Assessment and Text Simplification Special Issue of International Journal of Applied Linguistics*, Vol. 165, eds T. François and D. Bernhard (Amsterdam: John Benjamin Publishers), 97–135.

Coster, W., and Kauchak, D. (2011). "Learning to simplify sentences using Wikipedia," in *Workshop on Monolingual Text-To-Text Generation, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (Portland, OR), 1–9.

Crossley, S. A., Allen, D., and McNamara, D. S. (2012). Text simplification and comprehensible input: a case for an intuitive approach. *Lang. Teach. Res.* 16, 89–108. doi: 10.1177/1362168811423456

Cutler, A. (1983). "Lexical complexity and sentence processing," in *The Process of Language Understanding*, eds G. Flores d'Arcais and R. Jarvella (Chichester: John Wiley & Sons, Ltd.), 340.

Daelemans, W. (2013). "Explanation in computational stylometry," in *Computational Linguistics and Intelligent Text Processing. CICLing 2013, Lecture Notes in Computer Science, Vol. 7817*, ed A. Gelbukh (Berlin, Heidelberg: Springer), 451–462.

De Belder, J., and Moens, M. F. (2010). "Text simplification for children," in *Proceedings of the SIGIR 2010 Workshop on Accessible Search Systems* (Geneva).

De Marneffe, J., Ginter, M. C., Goldberg, F., Hajic, Y., Manning, J., McDonald, C. D., et al. (2016). "Universal dependencies v1: a multilingual treebank collection," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (Portorož), 1659–1666.

De Mauro, T. (2000). *Il Dizionario della Lingua Italiana*. Turin: Paravia.

Dell'Orletta, F. (2009). "Ensemble system for Part-of-Speech tagging," in *Proceedings of Evalita'09, Evaluation of NLP and Speech Tools for Italian* (Reggio Emilia).

Dell'Orletta, F., Montemagni, S., and Venturi, G. (2011). "READ-IT: assessing readability of Italian texts with a view to text simplification," in *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2011)* (Edinburgh), 73–83.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43, 449–470. doi: 10.1515/ling.2005.43.3.449

Dmitrieva, A., and Tiedemann, J. (2021). "Creating an aligned Russian text simplification dataset from language learner data," in *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing* (Kiyv), 73–79.

Drndarevic, B., Stajner, S., and Saggion, H. (2012). "Reporting simply: a lexical simplification strategy for enhancing text accessibility," in *Proceedings of "Easy to Read on the Web"*.

François, T., and Bernhard, D. (2014). When text readability meets automatic text simplification. *Int. J. Appl. Linguist.* 165, 89–96. doi: 10.1075/itl.165.2.00int

François, T., and Fairon, C. (2012). "An "AI readability" formula for French as a foreign language," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (Jeju), 466–477.

Frazier, L. (1985). "Syntactic complexity," in *Natural Language Parsing*, eds D. R. Dowty, L. Karttunen, and A. M. Zwicky (Cambridge, UK: Cambridge University Press), 129–189.

Futrell, R., Mahowald, K., and Gibson, E. (2015). "Quantifying word order freedom in dependency corpora," in *Proceedings of the Third International Conference on Dependency Linguistics* (Uppsala), 91–100.

Gala, N., Tack, A., Javourey-Drevet, L., François, T., Ziegler, J. C. (2020). "Alector: a parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers," *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)* (Marseille), 1353–1361.

Gardent, C., Shimorina, A., Narayan, S., and Perez-Beltrachini, L. (2017). "Creating training corpora for NLG micro-planning," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)* (Vancouver, BC).

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76.

Gildea, D., and Temperley, D. (2010). Do grammars minimize dependency length? *Cogn. Sci.* 34, 286–310. doi: 10.1111/j.1551-6709.2009.01073.x

Gonzalez-Dios, I., Aranzabe, M. J., and Díaz de Ilarraza, A. (2017). The corpus of Basque simplified texts (CBST). *Lang. Resour. Eval.* 52, 217–247. doi: 10.1007/s10579-017-9407-6

Goto, I., Tanaka, H., and Kumano, T. (2015). "Japanese news simplification: task design, data set construction, and analysis of simplified text," in *Proceedings of MT Summit XV, Vol. 1: MT Researchers' Track Miami* (Miami: FL), 17–31.

Grabar, N., and Cardon, R. (2018). "CLEAR—Simple corpus for medical French," in *Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA)* (Tilburg), 3–9.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *J. Linguist.* 42, 25–70. doi: 10.1017/S0022226705003683

Inui, K., Fujita, A., Takahashi, T., Iida, R., and Iwakura, T. (2003). "Text simplification for reading assistance: a project note," in *Proceedings of the Second International Workshop on Paraphrasing, ACL* (Sapporo: ACL).

Kajiwara, T., and Komachi, M. (2016). "Building a monolingual parallel corpus for text simplification using sentence similarity based on alignment between word embeddings," in *Proceedings of the 26th COLING* (Osaka), 1147–1158.

Kauchak, D. (2013). "Improving text simplification language modeling using unsimplified text data," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Sofia), 1537–1546.

Kintsch, W., and Keenan, J. (1973). Reading rate and retention as a function of the number of prepositions in the base structure of sentences. *Cogn. Psychol.* 5, 257–274.

Klaper, D., Ebling, S., and Volk, M. (2013). "Building a German/simple German parallel corpus for automatic text simplification," *Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (Sofia), 11–19.

Klebanov, B. B., Knight, K., and Marcu, D. (2004). "Text simplification for information-seeking applications", in *On the Move to Meaningful Internet Systems, Lecture Notes in Computer Science*, eds R. Meersman and Z. Tari (Berlin: Springer Verlag), 735–747.

Klerke, S., and Søgaard, A. (2012). "DSim, a Danish parallel corpus for text simplification," in *Proceedings of Language Resources and Evaluation Conference* (Istanbul), 4015–4018.

Kriz, R., Miltsakaki, E., Apidianaki, M., and Callison-Burch, C. (2018). "Simplification using paraphrases and context-based lexical substitution," in *Proceedings of NAACL-HLT 2018* (New Orleans, LA), 207–217.

Lin, D. (1996). "On the structural complexity of natural language sentences," in *Proceedings of the International Conference on Computational Linguistics* (Copenhagen), 729–733.

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *Int. J. Corpus Linguist.* 14, 3–28. doi: 10.1075/ijcl.14.1.02lu

Lubetich, S., and Sagae, K. (2014). "Data-driven measurement of child language development with simple syntactic templates," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (Dublin), 2151–2160.

Maratsos, M. (1974). Children who get worse at understanding the passive: a replication to Bever. *J. Psycholinguist. Res.* 3, 65–74.

Miaschi, A., Brunato, B., Dell'Orletta, F. (2021). A NLP-based stylometric approach for tracking the evolution of L1 written language competence. *J. Writing Res.* 13, 71–105. doi: 10.17239/jowr-2021.13.01.03

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97.

Miller, J., and Weinert, R. (1998). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Clarendon Press.

Montemagni, S. (2013). "Tecnologie linguistico-computazionali e monitoraggio della lingua italiana," in *Studi Italiani di Linguistica Teorica e Applicata (SILTA)*, 145–172.

Narayan, S., and Gardent, C. (2014). "Hybrid simplification using deep semantics and machine translation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Baltimore, MD), 435–445.

Narayan, S., Gardent, C., Cohen, S. B., and Shimorina, A. (2017). "Split and rephrase," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Copenhagen), 606–616.

Nelken, R., and Shieber, S. M. (2006). "Towards robust context-sensitive sentence alignment for monolingual corpora," in *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics* (Trento).

Nguyen, D., Doğruöz, A. S., Rosé, C. P., and de Jong, F. M. G. (2016). Computational sociolinguistics: a survey. *Comput. Linguist.* 42, 537–593. doi: 10.1162/COLI_a_00258

Niklaus, C., Bermeitinger, B., Handschuh, S., and Freitas, A. (2016). "A sentence simplification system for improving relation extraction," in *Proceedings of the 6th International Conference on Computational Linguistics (COLING 2016)* (Osaka).

Nisioi, S., Štajner, S., Ponzetto, S. P., and Dinu, L. P. (2017). "Exploring neural text simplification models," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Vancouver, BC), 85–91.

Orasan, C., Evans, R., and Dornescu, I. (2013). "Text simplification for people with autistic spectrum disorders," in *Towards Multilingual Europe 2020: A Romanian Perspective*, eds D. Tufis, V. Rus, and C. Forascu (Bucharest: Romanian Academy Publishing House), 187–312.

Ozuru, Y., Dempsey, K., and McNamara, D. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learn. Instruct.* 19, 228–242. doi: 10.1016/j.learninstruc.2008.04.003

Paetzold, G. H., and Specia, L. (2017). A survey on lexical simplification. *J. Artif. Intell. Res.* 60, 549–593. doi: 10.1613/jair.5526

Pavlick, E., and Callison-Burch, C. (2016). "Simple PPDB: a paraphrase database for simplification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Berlin), 143–148.

Pellow, D., and Eskenazi, M. (2014). "An open corpus of everyday documents for simplification tasks," in *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* (Gothenburg), 84–93.

Petersen, S. E., and Ostendorf, M. (2007). *Text simplification for language learners: A corpus analysis*. University of Washington, Seattle, United States. Available online at: https://www.isca-speech.org/archive_open/archive_papers/slate_2007/sle7_069.pdf

Prud'hommeaux, E. T., Roark, B., Black, L. M., and van Santen, J. (2011). "Classification of atypical language in autism," in *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (Portland, OR), 88–96.

Rello, L., Baeza-Yates, R., Bott, S., and Saggion, H. (2013). "Simplify or help? Text simplification strategies for people with dyslexia," in *W4A '13 Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (Rio de Janeiro).

Roark, B., Mitchell, M., and Hollingshead, K. (2007). "Syntactic complexity measures for detecting mild cognitive impairment," in *Proceedings of the Workshop on Biological, Translational, and Clinical Language Processing* (Czech Republic: ACL), 1–8.

Santini, M. (2007). "Characterizing genres of web pages: genre hybridism and individualization," in *40th Annual Hawaii International Conference on Conference on System Sciences* (Big Island), 71–78.

Säuberli, A., Ebling, S., and Volk, M. (2020). "Benchmarking data-driven automatic text simplification for German," in *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI2020)*, 41–48.

Scarton, C., Paetzold, G. H., and Specia, L. (2018). "Text simplification from professionally produced corpora," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (Miyazaki), 3504–3510.

Shardlow, M. (2014). "A survey of automated text simplification," in *International Journal of Advanced Computer Science and Applications (IJACSA)*, Special Issue on "Natural Language Processing" (West Yorkshire), 58–70.

Siddharthan, A. (2002). "An architecture for a text simplification system," in *Proceedings of the Language Engineering Conference (LEC 2002)* (Hyderabad).

Siddharthan, A. (2010). "Complex lexico-syntactic reformulation of sentences using typed dependency representations," in *Proceedings of the 6th International Natural Language Generation Conference* (Trim), 125–133.

Siddharthan, A. (2014). "A survey of research on text simplification," in *Recent Advances in Automatic Readability Assessment and Text Simplification*, eds T. Francois and D. Bernhard (Amsterdam: John Benjamins Publishing Company), 259–298.

Slobin, D. I., and Bever, R. G. (1982). Children use canonical sentence schemas. A cross-linguistic study of word order and inflections. *Cognition* 12, 229–265.

Specia, L. (2010). "Translating from complex to simplified sentences," in *Computational Processing of the Portuguese Language. Proceedings of the 9th International Conference (PROPOR 2010)* (Porto Alegre), 30–39.

Štajner, S., and Popović, M. (2016). Can text simplification help machine translation? *Balt. J. Modern Comput.* 4, 23–242.

Straka, M., Hajic, J., and Strakova, J. (2016). "UD-Pipe: trainable pipeline for processing CoNLL-U files per-forming tokenization, morphological analysis, pos tag-ging and parsing," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)* (Potroroz), 4290–4297.

Suter, J., Ebling, S., Volk, M. (2016). "Rule-based automatic text simplification for German," in *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)* (Bochum), 279–287.

Tonelli, S., Palmero Aprosio, A., and Saltori, F. (2016). "SIMPITIKI: a Simplification corpus for Italian," in *Proceedings of Third Italian Conference on Computational Linguistics* (Napoli: Academic University Press), 1–6.

van Halteren, H. (2004). "Linguistic profiling for author recognition and verification," in *Proceedings of the Association for Computational Linguistics* (Barcelona: ACM), 200–207.

Vickrey, D., and Koller, D. (2008). "Sentence simplification for semantic role labeling," in *Proceedings of ACL-08: HLT* (Columbus), 344–352.

Volpato, F. (2010). *The acquisition of relative clauses and phi-features: evidence from hearing and hearing-impaired populations* (Ph.D. dissertation). Ca' Foscari University of Venice, Venice, Italy.

Weiss, Z., and Meurers, D. (2019). "Analyzing linguistic complexity and accuracy in academic language development of German across elementary and secondary school," in *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications* (Florence: Association for Computational Linguistics).

Wilson, M. (1988). MRC psycholinguistic database: machine-usable dictionary, version 2.00. *Behav. Res. Methods Instrum. Comput.* 20, 6–10.

Woodsend, K., and Lapata, M. (2011). "Learning to simplify sentences with quasi-synchronous grammar and integer programming," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Edinburgh: ACM), 409–420.

Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: new data can help. *Trans. Assoc. Comput. Linguist.* 3, 283–297. doi: 10.1162/tacl_a_00139

Xu, W., Napoles, C., Pavlick, E., Chen, Q., and Callison-Burch, C. (2016). Optimizing statistical machine translation for text simplification. *Trans. Assoc. Comput. Linguist.* 4, 401–415. doi: 10.1162/tacl_a_00107

Yadav, H., Vaidya, A., Shukla, V., and Husain, S. (2020). Word order typology interacts with linguistic complexity: a cross-linguistic corpus study. *Cogn. Sci.* 44:e12822. doi: 10.1111/cogs.12822

Yasseri, T., Kornai, A., and Kertész, J. (2012). A practical approach to language complexity: a Wikipedia case study. *PLoS ONE* 7:e48386. doi: 10.1371/journal.pone.0048386

Yngve, V. H. A. (1960). "A model and an hypothesis for language structure," in *Proceedings of the American Philosophical Society* (Philadelphia, PA: American Philosophical Society), 444–466.

Zhu, Z., Bernhard, D., and Gurevych, I. (2010). "A monolingual tree-based translation model for sentence simplification," in *Proceedings of the 23rd International Conference on Computational Linguistics* (Beijing), 1353–1361.

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.

Publisher's Note: All claims expressed in this article are solely those of the author and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Brunato, Dell'Orletta and Venturi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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