THE ADAPTIVE VALUE OF LANGUAGES: NON-LINGUISTIC CAUSES OF LANGUAGE DIVERSITY

EDITED BY: Antonio Benítez-Burraco and Steven Moran

frontiers Research Topics

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: researchtopics@frontiersin.org

THE ADAPTIVE VALUE OF LANGUAGES: NON-LINGUISTIC CAUSES OF LANGUAGE DIVERSITY

Topic Editors:
Antonio Benítez-Burraco, University of Seville, Spain
Steven Moran, University of Zurich, Switzerland

The goal of this eBook is to shed light on the non-linguistic causes of language diversity, and in particular, to explore the possibility that some aspects of the structure of languages may result from an adaptation to the natural and/or human-made environment. Traditionally, language diversity has been claimed to result from random, internally-motivated changes in language structure. However, ongoing research suggests instead that different factors that are external to language can promote language change and ultimately account for aspects of language diversity, specifically features of the social and physical environments. The contributions in this eBook discuss whether some aspects of languages are an adaptation to ecological, social, or even technological niches.

Table of Contents

Editorial: The Adaptive Value of Languages: Non-linguistic Causes of Language Diversity

Antonio Benítez-Burraco [1]* and Steven Moran [2]

[1] Department of Spanish, Linguistics, and Theory of Literature, University of Seville, Seville, Spain, [2] Department of Comparative Linguistics, University of Zurich, Zürich, Switzerland

Editorial on the Research Topic

The Adaptive Value of Languages: Non-linguistic Causes of Language Diversity

The goal of this volume is to shed light on the non-linguistic causes of language diversity, and particularly, to explore the possibility that some aspects of the structure of languages result from an adaptation to the natural and/or human-made environment. Variation is pervasive in language. The languages we speak are not homogeneous. They change, both structurally and functionally, from one social group to another, from children to adults, from men to women, from one ethnic group to another, not to mention through historical, and evolutionary time. Moreover, the context in which conversational exchanges take place also affects the structure and the pattern of usage across languages. Besides social variation, geography also accounts for aspects of the variation observed within languages. The differential dispersal of linguistic features across geographically-defined areas usually results in different dialects of one language spoken across the whole distribution area of the language. Ultimately, each person acquires and makes use of a subtly different version of their mother tongue. All of this is very familiar, and over the years, linguists have learnt that these aspects of linguistic variation result from linguistic and extralinguistic factors are constrained in systematic ways, to the extent that they can be described by the right mixture of general principles and statistical biases (e.g. Labov, 2001).

In this Research Topic, we have put the focus on macrovariation across languages from a typological perspective, instead of microvariation within languages, because this aspect of language diversity has been quite satisfactorily characterized by sociolinguists, dialectologists and experts in discourse analysis. When we examine variation at this macro level, we soon realize that thousands of languages are spoken across the world and that they are endowed with distinctive, sometimes idiosyncratic, phonologies, morphologies, and grammars. These aspects of linguistic variation seem to be constrained as well, and we have equally learnt to characterize them in terms of a mixture of common principles and dimensions where languages can differ one from another (e.g., Baker, 2001). Nonetheless, it is not clear what are the causes of this variation. If we put aside the lexicon, which is generally acknowledged to serve as a reservoir for relevant cultural features of the society speaking the language, the twentieth century consensus has been that all languages are roughly equal in terms of overall complexity and that aspects of languages known to vary result from random drift or internally-motivated changes in language structure (Fromkin and Rodman, 1983; Dixon, 1997). To a great extent, this consensus is based on the assumption that human cognition is similarly configured in all human beings, and therefore, that the human faculty for language is uniform within the species (Chomsky, 1965, 1980; Moro, 2008). In the sixties, this assumption crystallized in the Chomskyan hypothesis of the "Universal Grammar."

This is not exact. In truth, there is also a high degree of variation "at the bottom," namely, regarding the biological underpinnings of the faculty that enables us to acquire and use languages (let's call it, more neutrally, our *language-readiness*). Accordingly, different language modalities (signed vs. spoken) do exist and can co-exist in the mind of the same user (Emmorey and McCullough, 2009). Additionally, the scores obtained in psycholinguistic tasks change from one person to another across the normal population (Fenson et al., 2000). Language disorders are the extreme of this kind of variation (Benítez-Burraco, 2016). Likewise, language developmental milestones are achieved at different times by children, relying on cognitive abilities that also vary from one to another (Bates et al., 1988; Dehaene et al., 1997). Additionally, the brain areas involved in language processing change, to some extent, from one individual to another (Fedorenko and Kanwisher, 2009; Prat and Just, 2011). Finally, many different genes (not just one or a handful) regulate the development of the brain areas important for language and many of them have functional variants that affect language processing in the neurotypical population (see Benítez-Burraco, 2009 for an overview). Surely, robust biological mechanisms exist as well that channel all this variation, to the extent that a similar faculty of language emerges in all human beings at the end of development, pathologies aside). Although the factors involved are different by nature, this does not differ from the convergence of all speakers of a particular language on a similar interiorized grammar in spite of having being reared in linguistic environments that are not identical.

Likewise, it seems now that languages also differ regarding their global complexity. The complexity of languages can increase, for instance, as a result of specific linguistic processes, like grammaticalization, which increases the number of categories or the number of irregularities (Givón, 1979). More importantly, the overall language complexity, as well as the complexity of specific components of the languages' grammars, can perhaps be explained by extralinguistic factors as well. Accordingly, language complexity has been found to correlate with features of the social environment impacting on language contact and language acquisition. For example, it seems to be greater when the language has more native speakers, when speakers are not involved in frequent cross-cultural exchanges, and when they are isolated (Bolender, 2007; McWhorter, 2007; Wray and Grace, 2007; Lupyan and Dale, 2010; Trudgill, 2011). As for another example, it has been claimed that a positive correlation exists between population size and phoneme inventory size (Hay and Bauer, 2007, but see Moran et al., 2012 for an opposite view). Eventually, core properties of human languages, like duality of patterning, have been argued to emerge as a result of iterative learning and cultural evolution, as nicely illustrated by research in village sign languages (Sandler et al., 2005) or in language evolution (Fleming, 2017). In a similar vein, language structure is also thought to be influenced on a long timescale by the physical environment, either directly or indirectly, via its effect on social structures. Familiar examples are the negative effect of dry climates on tone usage and the number of vowels (Everett et al., 2015), or of dense vegetation on sounds characterized by lower frequencies (Maddieson,

2011; Maddieson and Coupé, 2015). More generally, global language diversity has been claimed to negatively correlate with the ecological risk, that is, the amount of variation which people face in their food supply over time (Nettle, 1998). Similarly, the number of phyla or stocks has been suggested to negatively correlates with the time of occupancy of a territory (Nettle, 1998). Overall, it seems desirable to have a better knowledge of current patterns of linguistic diversity across the world, and particularly, of the ecological and socio-cultural factors that correlate with (and ideally, explain) aspects of this diversity. From an evolutionary perspective, we wish to know more about the adaptive value of language diversity and how it emerges over time as the physical, social, and cultural environment becomes modified. Several of the papers of this Research Topic explore this kind of correlation (and causation). Ultimately, we expect that these and other similar studies cast light as well onto some aspects of the deep evolution of language (and languages), provided that niche construction (perhaps via human self-domestication) has proven to account for aspects of language complexity via cultural evolution (Benítez-Burraco et al., 2016) and because some aspects of languages seem to be an adaptation to ecological, social, or cultural niches.

Finally, language complexity is also expected to be influenced by cognitive patterns, for instance, if some kind of processing preference biases language learning and use, and ultimately, what becomes grammaticalized (Bornkessel-Schlesewsky and Schlesewsky, 2009). [Note the other way around is also true, because aspects of language that are more costly to process and learn might favor the creation of "cognitive gadgets" through modifications in learning and data-acquisition mechanisms (Clarke and Heyes, 2017)]. More generally, recent research has concluded that cognitive differences among human populations do exist and are in part due to genetic changes in response to environmental factors, and not only to cultural or sociological forces (Winegard et al., 2017). Similarly, our "language genotype" (that is, the set of genes involved in the development and functioning of brain areas recruited for language processing) is not homogeneous either, with variants of specific genes contributing to normal variation in speech and language abilities (Deriziotis and Fisher, 2017). Accordingly, we could speculate about certain gene alleles influencing on aspects of languages that are known to vary, like phonology or morphosyntax. Again, this effect might be direct, if the involved genes contribute, for instance, to aspects of our vocal behavior. But most plausibly, we should expect that the effect is indirect, if specific alleles bias language acquisition or processing in some subtle ways, ultimately impacting on language change through iterated cultural transmission (Dediu, 2008, 2011). It is clear then that it seems desirable to better understand the complex interaction between genes, cognition, and the environment, and its effects on language diversity, both in the present-day populations and in the remote prehistory. In this sense, gene-culture co-evolution is expected to account for crucial aspects of language diversity too.

In this volume we bring together 12 contributions from 25 leading scholars in different research areas of interest for the questions we have highlighted above. Three of the

papers discuss important theoretical and methodological issues. Mendívil-Giró adds a note of caution regarding the sources of language variability. According to his view, it is the structure of the brain/mind that mostly affects language structure and we should make dependent of this circumstance any putative effect of the environment on how languages are built. Roberts presents a *maximum robustness* approach for studying adaptation in language. The method is a causal, incremental and robust approach aimed at testing hypotheses and identifying linguistic adaptation patterns in a world of increasing data, methods, and computational power. He addresses how to formalize a theory and how to identify criteria for integrating results from different approaches and methods into clear hypothesis testing and results assessment. Finally, the paper by Coupé focuses on optimal statistical tools for analyzing potential correlations between linguistic and extralinguistic variables. In particular, he discusses several techniques that help modeling data that are not analyzable with simpler linear regression models, including linear mixed-effects regression models (LMM), generalized linear mixed-effects models (GLMM), generalized additive models (GAM), and generalized additive models for location, scale, and shape (GAMLSS), which allow one to circumvent the limitations of commons distributions.

Turning to the papers exploring correlations between linguistic and extralinguistic variables, two of them address potential links between aspects of the physical environment and features of languages. Maddieson has found that the proportion of sonority vs. obstruency is higher in languages spoken in warmer climes. Interestingly, he suggests that given the highly malleable nature of the phonological structure of human languages, the time scale in which environmental factors influence the phonological make up of languages is acting at a scale faster than previously put forward in the literature. Likewise, Everettshows evidence for a positive association between reduced ambient humidity and reduced vowel-usage rates in a large sample of the world's languages. Importantly, some physiological evidence, involving larynx behavior, is presented to account for the observed correlation. Overall, the effect of the environment on languages' phonologies is controversial and we should be cautious with such approaches and scrupulous of the results, as stressed by Roberts and Maddieson.

Four other papers focus on the links between language diversity and sociological features. Nichols examines the effect of language mixing on the emergence of what she calls "linguistic attractors," that is, linguistic items, and features that are preferred by languages in their evolution. As she highlights, the emergence of linguistic attractors is linked to specific demographic, sociological, cultural, and environmental factors. Greenhill et al. contribute to the long and ongoing debate of whether population size has an observable effect on language change. In particular, they ask whether rates of lexical replacement in three large language families (Austronesian, Indo-European and the Bantu subfamily of Niger-Congo) are affected by speaker population size. Their results show an effect that does not generalize across

families. Greenhill et al.'s paper is also important as well because it highlights the differences between historical transmission of languages and the evolution of biological organisms. Whereas evolutionary theory makes clear predictions of rates and patterns of genetic change in regard to population size, it seems that language change may be driven by different mechanisms. Sinnemäki and Di Garbo focus on a related effect of the sociolinguistic environment on language structure, namely, the effect of the number of native speakers and the proportion of adult second language learners, which have been claimed to have an impact on language complexity, and particularly, on morphological complexity (Lupyan and Dale, 2010). Their data suggest that different sociolinguistic variables might affect different grammatical features differently. Importantly, they argue that modeling together several sociolinguistic features favors detecting possible adaptation of linguistic structure to the sociolinguistic environment. Lastly, Schembri et al. explore the links between the social environment and language structure sign languages. This is important provided that sign languages are endowed with the same structural features and properties as oral languages. What Schembri et al. have found is that sign languages change might support the view that morphological complexity depends on social features of the speech community. Nonetheless, they warn against a direct effect of population size and network density on language complexity, which seems to depend as well on how and when the language is acquired and its degree of contact with other language modalities.

Finally, three papers deal with the cognitive aspects of language variation. González-Perilli et al. study color object perception in two different Spanish-speaking populations, and show that Uruguayans, who use single words for two shades of blue, are more accurate at distinguishing between light blue and dark blue in a color stimuli perception task than are Spaniards, who use compound terms. These findings add to the ongoing debate of whether language and culture affect how individuals organize and process information from their world experience. Linguistic relativity effects are disputed by researchers, but there is much evidence for them across different cognitive domains and languages, including spatial cognition, and color recognition. Kempe and Brooks raise two important points of caution regarding the finding by Lupyan and Dale (2010) that morphological complexity is negatively correlated with population size. First is the need to improve our characterization (and understanding) of language complexity, if we want to properly address the questions of whether languages are equally complex and whether languages remain so by compensating for complexity in different subsystems of grammar [see (Moran and Blasi, 2014), and inter alia, for an overview]. Regarding morphological complexity, which is the focus of Kempe and Brooks' paper, the authors suggest that operationalizing morphological complexity based on combined informational value of morphological cues in the languages might be the best choice to capture the links between language processing and language change. Second, Kempe and Brooks also warn against the view that the cognitive limitations of children support mechanisms beneficial for learning of

complex morphology relative to adults. The authors argue convincingly that the difference in learning strategies by child and adult learners needs to have a more solid empirical foundation in which it is crucial to define morphological complexity with operationalizations that are cognitively-based. Lastly, the paper by Toya and Hashimoto aims to identify the environmental triggers and the evolutionary path of recursive combination, thought to be a human-specific ability and a core operation in human languages. They rely on a learning game approach. Their results suggest that recursive combination is adaptive because it results in more robust production mechanisms and more diversified products, a lesson that can be extended to material culture, human cognition, and language.

This volume contributes to the exciting challenges of disentangling the effect of the environment on language structure and complexity, and ultimately, helps us to form a better understanding of the nature and evolution of human language.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Baker, M. (2001). *The Atoms of Language*. New York, NY: Basic Books.

Bates, E., Bretherton, I., and Snyder, L. (1988). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. Cambridge: Cambridge University Press.

Benítez-Burraco, A. (2009). *Genes y lenguaje [Genes and Language]*. Barcelona: Reverté.

Benítez-Burraco, A. (2016). "A biolinguistic approach to language disorders: towards a paradigm shift in clinical linguistics," in *Advances in Biolinguistics: The Human Language Faculty and its Biological Basis*, eds K. Fujita and C. Boeckx (London: Routledge), 256–272.

Benítez-Burraco, A., Theofanopoulou, C., and Boeckx, C. (2016). Globularization and domestication. *Topoi* 37, 265–278. doi: 10.1007/s11245-016-9399-7

Bolender, J. (2007). Prehistoric cognition by description: a Russellian approach to the upper paleolithic. *Biol. Philos.* 22, 383–399. doi: 10.1007/s10539-006-9058-2

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Lang. Linguist. Compass* 3, 19–58. doi: 10.1111/j.1749-818X.2008.00099.x

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. (1980). *Rules and Representations*. New York, NY: Columbia University Press.

Clarke, E., and Heyes, C. (2017). The swashbuckling anthropologist: henrich on the secret of our success. *Biol. Philos.* 32, 289–305. doi: 10.1007/s10539-016-9554-y

Dediu, D. (2008). The role of genetic biases in shaping the correlations between languages and genes. *J. Theor. Biol.* 254:400–407. doi: 10.1016/j.jtbi.2008.05.028

Dediu, D. (2011). Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? *Hum. Biol.* 83, 279–296. doi: 10.3378/027.083.0208

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., et al. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport* 8, 3809–3815. doi: 10.1097/00001756-199712010-00030

Deriziotis, P., and Fisher, S. E. (2017). Speech and language: translating the genome. *Trends Genet.* 33, 642–656. doi: 10.1016/j.tig.2017.07.002

Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Emmorey, K., and McCullough, S. (2009). The bimodal bilingual brain: effects of sign language experience. *Brain Lang.* 109, 124–132. doi: 10.1016/j.bandl.2008.03.005

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci. U.S.A.* 112: 1322–1327. doi: 10.1073/pnas.1417413112

Fedorenko, E., and Kanwisher, N. (2009). Neuroimaging of language: why hasn't a clearer picture emerged? *Lang. Linguist. Compass* 3, 839–865. doi: 10.1111/j.1749-818X.2009.00143.x

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., and Thal, D. (2000). Measuring variability in early child language: don't shoot the messenger. *Child Dev.* 71, 323–328. doi: 10.1111/1467-8624.00147

Fleming, L. (2017). Phoneme inventory size and the transition from monoplanar to dually patterned speech. *J. Lang. Evol.* 2, 52–66. doi: 10.1093/jole/lzx010

Fromkin, V., and Rodman, R. (1983). *An Introduction to Language*. New York, NY: Holt, Rinehart & Winston.

Givón, T. (1979). *On Understanding Grammar*. New York, NY: Academic Press.

Hay, J., and Bauer, L. (2007). Phoneme inventory size and population size. *Language* 83, 388–400. doi: 10.1353/lan.2007.0071

Labov, W. (2001). *Principles of Linguistic Change*. Oxford: Blackwell.

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Maddieson, I. (2011). "Phonological complexity in linguistic patterning," in *Proceedings of the 17th International Congress of Phonetic Sciences,* eds W. S. Lee and E. Zee (Hong Kong: City University of Hong Kong*)*, 28–34.

Maddieson, I., and Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation hypothesis. *J. Acoust. Soc. Am.* 138:1838–1838. doi: 10.1121/1.4933848

McWhorter, J. (2007). *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. Oxford: OUP. Miestamo.

Moran, S., and Blasi, D. (2014). "Cross-linguistic comparison of complexity measures in phonological systems," in *Measuring Grammatical Complexity,* eds F. J. Newmeyer and L. Preston (Oxford: Oxford University Press), 217–240.

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting the population vs phoneme-inventory correlation. *LSA Ann. Meet. Extend. Abst.* [S.l.], 3, 29. doi: 10.3765/exabs.v0i0.601

Moro, A. (2008). *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. Cambridge: MIT Press.

Nettle, D. (1998). Explaining global patterns of language diversity. *J. Anthropol. Archaeol.* 17, 354–374.

Prat, C. S., and Just, M. A. (2011). Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cereb. Cortex* 21, 1747–1760. doi: 10.1093/cercor/bhq241

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2661–2665. doi: 10.1073/pnas.0405448102

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: OUP.

Winegard, B., Winegard, B., and Boutwell, B. (2017). Human biological and psychological diversity. *Evol. Psychol. Sci.* 3, 159–180. doi: 10.1007/s40806-016-0081-5

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Why Don't Languages Adapt to Their Environment?

José-Luis Mendívil-Giró*

Department of General and Hispanic Linguistics, University of Zaragoza, Zaragoza, Spain

The issue of whether languages adapt to their environment depends on our understanding of *language*, *adaptation,* and *environment*. I consider these three concepts from an internalist or biolinguistic point of view. If *adaptation* is defined as the result of the differential transmission of phenotypic traits by means of natural selection, then both natural species and languages are adapted. Recall that according to Darwin's own insight, the evolutionary mechanisms for species and languages are "curiously the same" (or "curiously parallel"). However, if the concept of adaptation entails that the environment is the essential source of the structure of evolving objects, then neither natural species nor languages can be said to be adapted to their environment. In the case of languages, I will argue that much of their structure is insensitive to historical change and, therefore, incapable of adaptation to the external environment. The immediate environment of languages is in fact internal to the mind/brain and is thus less variable than the social and physical environment in which people live. On the other hand, the dimensions of languages that are variable have such an indirect relation with the physical and social environment that the notion of adaptation to extra-linguistic reality can only be applied weakly, and then it is unable to explain the main patterns of linguistic structural diversity.

Keywords: language change, language evolution, adaptation, language typology, evolutionary theory, language diversity, faculty of language, I-language

INTRODUCTION: REASONS FOR SKEPTICISM

My aim here is to consider proposals that seek to explain the structure of languages in terms of adaptation to their physical and cultural environment, and to do so with a degree of skepticism. Ladd et al. characterize these proposals as "attempts to relate facts about language structure to facts about speakers and their environment—variables such as group size, geographical location, genetic makeup, and cultural expectations" (Ladd et al., 2015, p. 227). This is not, of course, to deny the inherent interest or value of such work (see current syntheses in Ladd et al., 2015; Lupyan and Dale, 2016). Actually, my critical position toward the claim that there is an influence of extralinguistic factors in the structure of languages is based on a restrictive conception of what is *the structure* of languages. Then, I neither reject nor question the works that detect (more or less robust) correlations between certain external factors and certain aspects of languages, but I argue that if we understand the structure of languages as it is done in the context of current syntactic theory (especially in the generativist domain), then the claim that the structure of languages can be explained as the result of an adaptation to environmental factors (social, physical, or otherwise) is misleading and inadequately simplifying. This is so because in the aforementioned tradition, the notion of "the structure of languages" transcends relatively superficial aspects (such as the

morphological manifestation of certain grammatical categories or the variation in word order) and focuses on (essentially syntactic) formal structural aspects that underlie all languages and that, ultimately, define what is a possible human language.

Therefore, my skepticism arises from two principal claims: (i) the influence of the physical and cultural environment in which languages are developed has a limited scope for explaining the structure of languages, including their main patterns of typological variation, and (ii) such studies do not lead to a satisfying account of what a human language is, from a cognitive and biological perspective, but rather, they take us back to a traditional (and incomplete) view of language as a purely cultural phenomenon.

COMPARING LANGUAGES AND SPECIES

Following August Schleicher, the first major linguist to address the analogy between languages and species suggested by Darwin, I will assume that "not a word of Darwin's need be changed here if we wish to apply this reasoning to languages" [Schleicher, 1863, p. 64 (I quote from the English translation included in Koerner, 1983)]. The reason for my assumption is that in both cases the evolving objects are *historically modified natural objects*. This identification allows us to say that the process of linguistic change and that of natural evolution are *formally* alike, although *substantially* different (for a review of different interpretations of the analogy between languages and species, see Mendívil-Giró, 2006, 2014).

Although various proposals for establishing the specific terms of the comparison have been suggested (e.g., Croft, 2000), the most appropriate one for my purpose is that formulated by Schleicher himself, in his review of the German edition of the *Origin of Species*:

> "The species of a genus are what we call the languages of a family, the races of a species are with us the dialects of a language; the sub-dialects or patois correspond with the varieties of the species, and that which is characteristic of a person's mode of speaking corresponds with the individual" (Schleicher, 1863, p. 32).

What Schleicher calls "that which is characteristic of a person's mode of speaking" is the closest concept to the Chomskyan notion of I-language that could be formulated at that time. Chomsky's (1985) distinction between I-language and E-language was formulated to make clear that the object of study of linguistics as part of cognitive science is not an external object, a shared code or a social institution, but a property of a speaker's mind/brain. Adopting this point of view, I argue that in the comparison between linguistic change and natural evolution the appropriate terms for comparison are as follows: the equivalent of the natural organism (the individual) is the I-language, while the equivalent of the species is a set of similar I-languages (what is usually called *a language*). Thus, in this context, a language such as Spanish is simply the set of I-languages of Spanish-speaking people (i.e., of the people we identify as users of this way of speaking that we call Spanish), just as the natural species of

tigers is nothing other than the set of organisms that we identify as tigers. In both cases the criterion of delimitation, based on similarity, is diffuse and somewhat arbitrary: the criterion of fertile breeding in natural species (Mayr, 1942), and the criterion of mutual intelligibility in languages (Dixon, 1997).

Central to this comparison is that both natural species and natural languages are groups of similar individuals. A natural species is made up of "sufficiently similar" individuals. An orangutan and a human being have more in common than an orangutan and a cow, but all three belong to different species. We know that the greater similarity between an orang-utan and a human is due to the fact that their common ancestor is far more recent (about 6 million years) than in the case of humans and cows, which goes back hundreds of millions of years. A "linguistic species" (i.e., a language in the normal use of the term) consists of "sufficiently similar" individuals (I-languages). Thus, the linguistic equivalent of the natural organism (e.g., a tiger) is each person's language organ (the I-language). The linguistic equivalent of the natural species (e.g., *Panthera tigris*) is the grouping of such language organs. And likewise Spanish and French are more alike than French and Russian, but all three are different languages. We know that the greater similarity between Spanish and French is due to the fact that their common ancestor is much more recent (about 1,500 years) than the ancestor they share with Russian (about 6,000 years).

If an I-language is a person's language organ (his/her faculty of language), there are not around 6,000 languages in the world, but billions, as many as there are people (in fact many more, given that bilingual people have more than one I-language). The only thing that can be said to exist, from an internalist, cognitive, point of view, are those billions of I-languages. All else (varieties, dialects, languages, families, etc.) are abstract constructs that we make by grouping I-languages according to their resemblances or their historical origins. The same is true in the biological realm: what exist are the emerging states of matter that we call life forms, the organisms (the billions of animals, plants, fungi, etc., living on the planet), whereas varieties, species, families, kingdoms, etc., are abstract constructs that we make on the basis of genetic and morphological similarity and historical origins.

And just as we would not say that tigers are manifestations or realizations of the species of tigers (which would have an independent existence), it is not appropriate to say that I-languages are manifestations or realizations of the Spanish or the Russian language (which would have an independent existence in grammars, in dictionaries or in social communities). The Chomskyan cognitive shift had as a central tenet the assertion that languages are not exclusively external, social objects that humans learn, use and transmit from generation to generation, but are in fact different (historically modified) states of the same language faculty, a specific attribute of human cognition. Similarly, natural organisms are different (historically modified) states of the same biochemical phenomenon: life (see Moreno and Mendívil-Giró, 2014 for a development of these ideas).

Comparable to natural evolution in biological organisms, then, is the process of linguistic change in human languages. The assumption that follows, hence, is that the process of language evolution (as a human faculty) is part of natural evolution, and

not part of linguistic change. In other words, the process of linguistic change is one that affects (in historical time) the systems of knowledge we call I-languages, and has no relation to the evolutionary processes that could give rise (in geological time) to the faculty of language. To avoid the "unfortunate ambiguity" (cf. Hurford, 1992, p. 273) that expressions like *language evolution* have in English, I use the term *linguistic change* to refer to the process of historical change in languages, and I will reserve the term *evolution* for biological changes, including the evolutionary emergence of the language faculty (an issue that I will not discuss here). In this sense it is possible to affirm, following Berwick and Chomsky (2016, p. 92), that "languages change, but they do not evolve." For arguments against the assumption that the process of linguistic change is part of the process of language evolution, see Mendívil-Giró (2016) and Longobardi (2003), who clearly distinguishes between historical adequacy and evolutionary adequacy in language sciences.

The parallelism between natural evolution and linguistic change in fact goes beyond the interesting similarities that Darwin (1871) observed, and persists in the relevant spheres of scholarship. Gould (2002) analyses in detail the controversy between adaptationist, externalist, and functionalist evolutionary theorists (using Gould's, 1996 characterization of neo-Darwinism) and, on the other hand, anti-neo-Darwinist theorists (such as Brian Goodwin, Stuart Kauffman, and Gould himself). In linguistics too there is also a parallel controversy, revolving around functionalist and non-functionalist theorists of language change (see Lass, 1997 for a detailed critical review, and for an argument against functional/adaptive models of linguistic change).

The impetus in the functionalist, adaptive approach to linguistic change is contemporary to the emergence and development of the Prague School of Linguistics (see Cercle Linguistique de Prague, 1929). I refer mainly to the conception of language as a social institution in the service of communication and to the preference for teleological explanations of linguistic change. It is relevant noting that the revival of teleological tendencies in the explanation of language change coincides in time and in orientation with the emergence in the twenties and thirties in the twentieth century of the *Modern Synthesis* of evolutionary theory. The new synthesis implies an inclination to consider natural selection as the only motive power of natural evolution, which implies the idea that every change must be adaptive. In my view, this trend corresponds to functionalist approaches to linguistic change and to the more recent tendency to consider languages as complex adaptive systems (Kirby, 1999).

Gould (1996) has described the fundamental difference between the neo-Darwinist model and its alternatives making use of the metaphor of the billiard ball against Galton's polyhedron. According to the neo-Darwinist point of view, an organism could be represented as a billiard ball in motion. Each time the cue hits the ball there is a variable movement. There is a free variation that goes in all directions. The cue hitting the ball would be natural selection, and the ball goes where selection drives it. This constitutes, in terms of Gould, an *externalist, functionalist,* and *adaptationist* evolutionary theory. By contrast, the anti-neo-Darwinist point of view presents the metaphor differently.

The organism would be as a polyhedron resting on one of its facets. Once the cue hits it, the prospects for change are very constrained: it is a polyhedron, which has a certain internal structure that limits variation, so that certain options are more likely than others and some are impossible, however interesting that might be from an adaptive point of view.

Of course, this is not the place to review the long dispute over the meaning and implications of the term *adaptation* in evolutionary theory, nor to reiterate the debate on the channeling of previous history and the laws of nature "on which natural selection was privileged to work" (Kauffman, 1993, p. 643). However, it is important to note that by adopting a cognitive point of view in the study of languages one cannot ignore the strict restrictions that the human brain and cognition impose on the structural design of languages, independently of those aspects susceptible to historical change (and, therefore, candidates for possible processes of adaptation to the environment).

Gould characterized the controversy in evolutionary theory as follows:

> *"In what ways does the skewed and partial occupancy of the attainable morphospace of adaptive design record the operation of internal constraints (both negative limitations and positive channels), and not only the simple failure of unlimited number of unconstrained lineages to reach all possible position in the allotted time?"* (Gould, 2002, p. 1053).

And both options have an equivalent view in current linguistic theory. The internist and formalist approach (characteristic of generative linguistics) conceives languages as systems of knowledge restricted in their range of variation by the structure of the human faculty of language (i.e., as Galton's polyhedrons). This view correlates with a uniformitarian conception of language diversity and with a restrictive conception of linguistic change. The externalist and functionalist approach (represented by cognitive-functional linguistics) conceives languages as external cultural objects that owe their structure to the adaptation to speakers' cognitive and communicative requirements (i.e., as billiard balls). This view correlates with a less constrained conception of linguistic change and with an emphasis on the diversity of languages (see Mendívil-Giró, 2012 for a review of this controversy).

I will argue that what we know about how, and how much, languages can change in time and in relation to the environment places us in the first scenario: i.e., one in which the human faculty of language strictly channels the aspects and components of languages that can vary in time and space.

BUT WHAT CHANGES WHEN LANGUAGES CHANGE?

According to Hauser et al. (2002) influential model, the human language faculty could be conceived of as a complex system minimally integrated by three components: a conceptual-intentional (CI) system (related to meaning and interpretation), a sensory-motor (SM) system (related to the perception and production of linguistic signals), and a computational system

(Narrow Syntax, responsible for the creation of the syntactic structure that underlies linguistic expressions, and ultimately for the compositionality and productivity of human language).

Following later developments of this model (Chomsky, 2007; Berwick and Chomsky, 2011, 2016), I will assume that the computational system has an asymmetrical relationship with the two "external" components (CI and SM), such that the computational system would be optimized for its interaction with the CI system, while the relationship with the SM system would be ancillary or secondary. See **Figure 1**.

It is then implied that the computational system is coupled with the CI system to form an *internal language of thought* (ILOT), one that would be essentially homogeneous within the species, and the evolutionary design of which would not be for communication but for thought. Chomsky has suggested that from an evolutionary point of view "the earliest stage of language would have been just that: a language of thought, used internally" (Chomsky, 2007, p. 13).

The connection of the ILOT with the SM system is what would allow the "externalization" of language for interaction and communication with others. Since the connection of the ILOT with the externalization systems is posterior or secondary, it would be precisely within this process that the principal source of the structural diversity among human languages would emerge:

> "*Parameterization and diversity, then, would be mostly – possibly entirely – restricted to externalization. That is pretty much what we seem to find: a computational system efficiently generating expressions interpretable at the semantic/pragmatic interface, with Diversity resulting from complex and highly varied modes of externalization, which, furthermore, are readily susceptible to historical change*" (Berwick and Chomsky, 2011, pp. 37–38).

The connection of the ILOT with the SM system is what allows the externalization of language and, incidentally, what causes the existence of different I-languages. The essential hypothesis is that the same ILOT underlies all languages, so that differences between them are not caused by differences in the CI, the computational, or even the SM systems (which would be biologically conditioned), but follow from differences in how the ILOT is connected to the SM system. Let us suppose, to simplify, that the interface between the ILOT and the sensorimotor system is a kind of "lexicon," i.e., a repertoire of morpho-phonological formants that allow the externalization of the hierarchical syntactic-semantic representations (produced by the computational system in its interaction with the CI system) in the form of chains of morphemes and phonemes (or, if applicable, visual signs). The role of the lexical interface, then, is to transform abstract hierarchical structures into sequential structures legible at the sensorimotor system. A possible way to understand the format of this lexical interface would be in terms of the type of lexical entries postulated in so-called nanosyntax (Starke, 2009).

Such a model predicts that the diversity in I-languages is the result of variations in externalization, i.e., variations in the configuration of the *lexical interface* represented in **Figure 1**. As shown in the diagram, the development of language in an individual implies the learning (the internalization) of the

"lexical" material necessary for communication, and it is exactly during this process that reanalyses can occur. A reanalysis is a mismatch in the grammar of two speakers between an internal representation and the linguistic expression produced by the SM system. It can be seen as the equivalent of genetic mutations in organisms.

Let us consider a simplified example: in present-day English the future is expressed as a phrase (*I will love*) whereas in Spanish it is expressed as a single word (*Amaré*). According to the model presented, the underlying syntactic structures of the two expressions are very similar (as well as their meaning), while the morphological (and phonological) structures are very different. However, what is now a bound morpheme in the Spanish future (-*é*) was an auxiliary verb in earlier stages of this language (derived from the vulgar Latin phrase *amare habeo* "I have to love," an alternative to the classic Latin synthetic form *amabo* "I will love"). The transition from a phrase (main verb + auxiliary) to a word (root + affix) at some point in the historical evolution of Romance necessarily implied a process of reanalysis (a mutation). Hence, and again to simplify, we could say that for speaker S_1 expression E has the underlying structure Verb+Aux, whereas for speaker S_2 the *same* expression E has the underlying structure Root+Affix, i.e., speaker S_2 *reanalyses* expression E, conferring on it a different underlying structure (Root+Affix) than that of speaker S_1 (Verb+Aux). In a sense, then, the I-language of speaker S_2 has a mutation, because the relationship between the elements of expression E and its underlying structure is different from that in the I-language of speaker S_1. The listener (or the child acquiring a language) does not have immediate access to the syntactic structure or to the semantic representation underlying a given expression, but only to the sound wave that externalizes it. The task of the listeners (or learners) is to use their I-language (including their own lexical interface) to discover this structure by analyzing the sound wave received. In the ideal case, the structure that they get is identical to what the speaker had in mind. When this is not the case, we can say that reanalysis has occurred. So reanalysis is basically a decoding (or acquisition) error, and when this error (this "mutation") is stabilized in the listener's I-language and is extended to other speakers, we say that there has been a linguistic change. The model predicts that changes happen in the lexical interface that materializes syntactic structures, not in the computational system itself. This view is coherent with the *inertial* theory of syntactic change (see Longobardi, 2001; Keenan, 2002).

On the other hand, in linguistic change, as in the case with natural evolution, one has to clearly differentiate the reasons why an innovation arises and the reasons why this innovation extends over a population over time. There are many factors that might lead, for example, to the introduction or elimination of a particular acoustic feature in a phonetic segment (from climatic conditions to the presence of speakers of other languages), but a linguistic change will only occur if that mutation extends to other individuals (I-languages), and this itself will only happen if the speakers imitate the speech of the innovators, and the innovations pass these on to subsequent generations. As Labov (1963) showed, the crucial factor in the selection of innovative variants, whether phonetic, morphological, lexical,
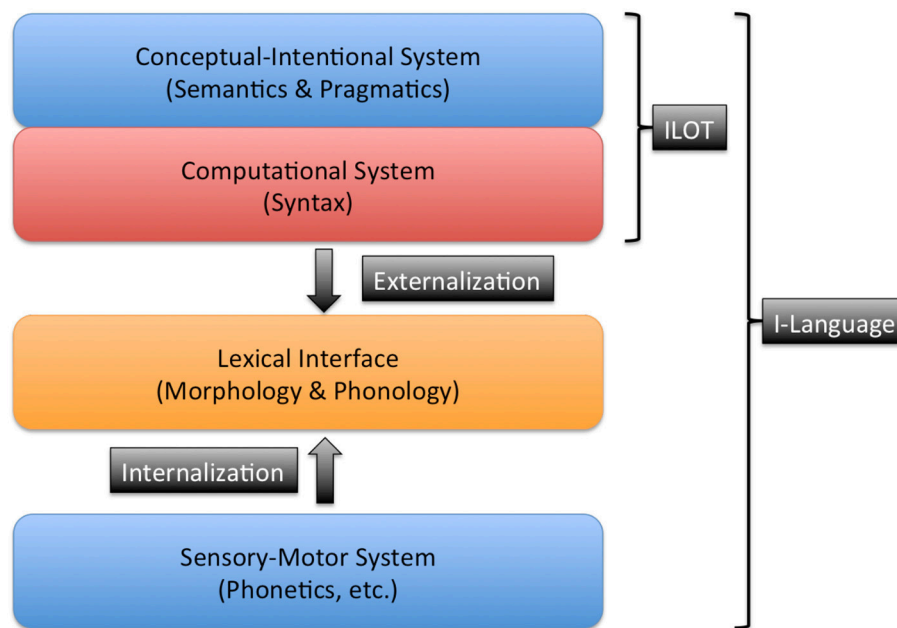
FIGURE 1 | The structure of an I-language. For each component the traditional area of research is indicated. The main components of language are asymmetrically related. The conceptual-intentional (CI) system has a direct connection with the computational system and they form an internal language of thought (ILOT). This ILOT is connected with the sensory-motor (SM) system for language externalization. This connection is established through a lexical interface. The lexical interface changes historically during the process of transmission from generation to generation.

or syntactic, is not functional efficiency or cost of execution, but social prestige. Some authors (e.g., Croft, 2000) argue that innovations are functional/adaptive, i.e., they have a teleological motivation. But as Lass notes, "unless a motivation is arbitrary, its implementation ought not to subject to contingent factors like age, sex, prestige, etc." (Lass, 1997, p. 364).

Differences between languages (such as differences between natural species) are the result of change, but linguistic changes only occur in the most superficial dimension of languages, those that are exposed to learning from the environment and are susceptible to historical reanalysis. In the same way, biological evolution significantly alters the form and structure of organisms, but does not modify the biochemistry on which they are built, this remaining unchanged since the emergence of the first forms of life.

THE STRUCTURAL TYPOLOGY OF LANGUAGES DOES NOT CORRELATE WITH THE CULTURAL DIVERSITY OF SPEAKERS

Even assuming that externalization patterns are the only thing that changes historically in languages, it could still be argued that there is a great deal of room for variation and that, therefore, the structural diversity of languages could reflect processes of adaptation to the environment. Indeed, we know that notable variation in the structure of languages does exist, although the model proposed in **Figure 1** would rule out the kind of weakly

restricted variation which some authors continue to advocate (see Evans and Levinson, 2009; Mendívil-Giró, 2012 for a critique).

The lack of correlation between different linguistic types and different aspects of human cultures is a strong argument in favor of a restrictive vision of the notion of adaptation applied to human languages, and in favor of a non-exclusively cultural vision of what a language is.

The parameters of linguistic structural variation that have always caught the attention of typologists are those of a morphosyntactic nature (i.e., related to how the morphology of languages reflects the syntactic structure). There are languages with case marking morphemes, and languages without them; there are languages in which verbs are conjugated and agree with several arguments, and languages in which they do not; there are languages in which heads precede complements, and languages in which this happens in reverse; and there are languages in which interrogative words move to the front of sentences, and languages in which they do not (see Dryer and Haspelmath, 2013 for a general survey). Between each of the mentioned options there is a complex range of intermediate steps. For example, among the languages that morphologically mark grammatical relations between verbs and arguments (either with cases or with agreement), some follow the nominative-accusative pattern (formally grouping the subject and differentiating the direct object) and others the ergative-absolutive pattern (formally grouping the subject of the intransitive verb and the object, and differentiating the subject of the transitive verb). Yet there are also languages that are accusative in certain tenses/aspects and ergative in others (see Dixon, 1994). All such variation is

compatible with the model set out in **Figure 1**, and a number of research programs are currently addressing the issues of structural typology based on differences in the externalization component (e.g. Richards, 2016).

What is relevant to us here is that, as Pinker (2007) has pointed out, "the non-universal, learned, variable aspects of language don't fit into any meaningful purposive narrative about the surrounding culture." The causes of the changes that produce such variation are inherent to linguistic structure itself, and to the mechanism of change (reanalysis). To quote Pinker once more, these changes "aren't part of any symbolic or teleological plan of the culture." Adapting Pinker's words to our example above, we can say that there are ergative languages and accusative languages, but there are no ergative cultures and accusative cultures. As Baker suggests, "indeed, there is no ecological regularity in how the major linguistic types are distributed around the world" (Baker, 2003, p. 350).

The assumption that there is a correlation between culture or worldview and the grammatical structure of languages is as old as reflections on language typology. In the past it was assumed that the degree of "cultural evolution" determined the degree of "linguistic evolution." Thus, if we turn again to the case of ergativity, it was claimed that ergativity correlated with a lack of rationality: "What for us is a true cause is for primitive man merely an event involving mystical forces" or "savage man apparently feels that most events are not due to his own volition" (quoted by Seely, 1977, *apud* Dixon, 1994, p. 214). Dixon argues that by using the same data we could conclude that only speakers of ergative languages have a true notion of agency, since only these speakers formally identify the agentive argument; he concludes that, "in fact, there is no one-to-one correspondence between grammatical marking and mental view of the world" (Dixon, 1994, p. 214).

Even in more recent times, there is no shortage of (more sophisticated and reasonable) proposals about the existence of covariation between culture and grammar, especially relating grammatical complexity with cultural complexity, such as Swadesh (1971), Perkins (1988), or Everett (2005). Swadesh (1971) p. 49) mentions a correlation between inflectional categories and languages' geographical and social extension. But this correlation, if it really exists, does not reveal an adaptation of grammar to culture, but is probably a consequence of morphological simplification, typical of many so-called "world languages" (see section The Brain Internal Environment: Language Learning and Language Processing for discussion). Perkins (1988) proposes a correlation between grammatical complexity and cultural complexity. He surveys in 50 languages several morphological deictic features (tense, person, deictic affixes), syntactic devices related to the coding of reference (determiners, relatives, conjunctions), as well as a measure of cultural complexity (based on the size of settlements, the number of types of craft specialists, and social and political hierarchy depth). Perkins finds a strong correlation that would imply a kind of "linguistic evolution": languages of complex cultures have few deictic affixes and many syntactic devices. However, Nichols applies her methods to these data and points out that these correlations "may actually reflect only

accidentally coincident macroareal linguistic distributions and have no ultimate connection to cultural complexity" (Nichols, 1992, p. 317). (Everett, 2005) proposal on the cultural constraints in Pirahã's grammar is not statistically significant, and the proposed correlation itself has been questioned (see Nevins et al., 2009).

The most reasonable conclusion, therefore, is that there is no correlation between the structural diversity of languages and the cultural diversity of speakers. The fact that one language, for example Mohawk, has more morphological complexity than another, for example English, has no relation to the complexity of the culture in which those languages are spoken, or to the sophistication of its literary tradition, but simply depends on a chain of previous historical facts. The bound morphemes that characterize the complex morphology of many languages are the result of the historical reanalysis of ancient free words. Yet the almost invariable, morphologically simple words that characterize other languages are often the result of the loss of morphological complexity, also resulting from historical reanalysis. In both cases reanalyses, like genetic mutations, are blind and random processes, and Darwin's conclusions can be applied to them: "There seems to be no more design in the variability of organic beings, and in the action of natural selection, than in the course the wind blows" (Darwin, 1893/2000, p. 63).

This conclusion has a solid empirical support. Both Nichols (1992) and Nettle (1999) quantitatively analyse linguistic diversity in time and space and, although with different samples and methodologies, they reach similar conclusions: although there are social and geographical factors that correlate with linguistic diversity and with the density of languages, there is no correlation between typological structural diversity and external factors. As Nettle points out: "Structural diversity [...] shows no overall pattern and no correlation with other types of diversity" (Nettle, 1999, p. 137).

Nettle suggests that some extralinguistic factors, such as the size of the speech community, could be related to the preservation of less frequent typological configurations (for example, OS word order, with the object preceding the subject). The argument is based on the assumption that infrequent types are less optimal in functional terms. This assertion is doubtful, because functional optimality is defined in relation to the greater or lesser frequency (I consider the relation between processing and grammar in section The Brain Internal Environment: Language Learning and Language Processing). If we ignore that problem, Nettle's suggestion is interesting. In this case the idea is that, as it happens in population genetics, the effects of random drift are greater when the population is small. But even in this case, it cannot be said that there is a correlation between linguistic types and extralinguistic factors, i.e., it cannot be said that small groups of speakers favor the evolution of certain linguistic types, nor that there is a causal relationship between a small group of speakers and the subject position in the sentence. Note that it could also be argued (what seems more likely) that the possible cause of the maintenance of an infrequent structural type in a given place is the isolation that defines small groups of speakers, isolation that would protect that group from the influence of

speakers from other languages (word order is a grammatical feature very prone to diffusion; see Dixon, 1997). What this case shows is that the size of groups of speakers can influence the dynamics of linguistic changes, something perfectly coherent with the model presented here, but that does not allow to affirm that a certain structural feature (the OS order) is an adaptation to a certain type of linguistic context (the size of the community of speakers).

Nichols' (1992) conclusions on the historical evolution of linguistic diversity are also very relevant in this context:

> "*This survey has uncovered no evidence that human language in general has changed since the earliest stage recoverable by the method used here. There is simply diversity, distributed geographically. The only thing that has demonstrably changed since the first stage of humanity is the geographical distribution of diversity*" (Nichols, 1992, p. 277).

If the generation of the structural diversity of languages were the result of adaptive processes to non-linguistic aspects (and not a continuous drift within a restricted design space) we should expect some kind of progression in the historical change of languages, such as we observe in other cultural institutions (politics, art, science, or technology), but this is not the case.

Although structural types of languages do not correlate with the types of societies and cultures that populate our planet, it is still possible to see how certain formal aspects of languages can be explained as processes of adaptation to the environment within the process of linguistic change. However, prior to this we need to determine what is understood by environment and what aspects of a language are sensitive to it.

WHAT IS THE ENVIRONMENT TO WHICH THE VARIABLE PARTS OF LANGUAGES WOULD ADAPT?

So far I have assumed a generic notion of environment, as formulated in the *leit Motiv* of the *Research Topic* in which this contribution is included ("to explore the possibility that some aspects of the structure of languages may result from an adaptation to the natural and/or human-made environment"). I have shown that the claim that there is covariation between morphosyntactic typology and aspects of the environment (so defined) is empirically weak, something that is consistent with the prediction made by the presented model of what I-languages are, and what their margin of variation is.

The diagram in **Figure 1** represents any I-language (i.e., the equivalent of a natural organism). As I have pointed out, it is obvious that every I-language has a variable component (the externalization component), therefore susceptible of adaptation to the environment (although to a lesser degree than it is assumed in models that conceive languages as purely cultural objects). But from this point of view, the notion of environment cannot be the same I have been using. What is the language external medium to which these variable parts could have adapted?

It is not a simple question. The structure of **Figure 1** may be interpreted as a sandwich, so that only the outer layers would be susceptible to contact with the environment. Thus, we could

consider that the CI and SM systems are "more external" than the computational system. The CI part of any language may be in contact with the rest of the conceptual system of people, so that it would then be expected that certain aspects of the physical, social, and cultural environment in which people develop and live can have an influence on the range of available concepts and notions. This would explain a relatively trivial aspect of the adaptation of languages to the environment, that of the substantive lexicon (Regier et al., 2016). In a culture with highly developed technology there will be words and phrases to denote scientific instruments, techniques, and concepts not found in languages spoken by hunter-gatherer communities, which, on the other hand, would have areas of the lexicon relating to wildly occurring food, animals, and methods of survival unrecognized in the languages of modern urban communities. Changes in culture, technology, and lifestyle often lead to changes in the lexical inventory that we require in everyday life. When a society moves from a rural to an industrialized life, the most widely used lexical inventory also changes. In this area, as pointed out by Ladd et al. (2015), several quantitative studies have shown that there is a correlation between environmental factors (latitude, ultra violet radiation) and the size of the lexical repertoire of color terms. But the differences in the type of conceptual elements that have specific lexical expression are not related to the morphosyntactic structure of languages. Indeed, languages spoken by supposedly simpler societies, hunter-gatherer societies, often have greater morphosyntactic complexity (greater "maturity" in the sense used by Dahl, 2004) than many European languages such as English or Romance languages.

On the other side of the sandwich, we have a sensory-motor system, which in oral languages corresponds to the vocal-auditory system. It is conceivable that certain aspects of the physical environment may bias the kind of sounds most used in some languages (see Everett et al., 2016), but again there would be very limited effects on the morphosyntactic structure of languages.

So, which environmental factors could have molded the historical drift of the morphosyntactic systems of languages? It is quite possible that such factors do not exist or have a weak effect, since the structural typology seems to be relatively isolated from the semantic and material dimension of languages and does not seem to fit them. But if we were to look for them, the place to start is within the brain.

THE BRAIN INTERNAL ENVIRONMENT: LANGUAGE LEARNING AND LANGUAGE PROCESSING

According to the model I have described here, the object of study, from a cognitive perspective, is not that of languages understood as social institutions, but the I-languages that reside in the minds/brains of individuals. In this context it is imperative that we recall that the only environment with which "mental organs" are in direct contact is the brain itself. If there is an "external" medium to which I-languages can adapt, it must be internal to the mind/brain.

It may be argued that many of the most notable changes that have been documented in the history of languages have contact with other languages as a crucial factor. And, indeed, it is indisputable that language contact has much more effect on linguistic phenotypes than the social or physical environment in which people live. But languages do not come into direct contact within the physical environment or in society, but only in the brains of speakers. Language A can only have influence on language B if the speaker of B has some kind of knowledge of language A. In our terms we could say that the development of a new lexical interface can affect the previous lexical interface, which can alter the linguistic emissions that the new generation of speakers will use to develop their own lexical interface.

Natural evolution is only possible thanks to the reproduction of organisms, and linguistic change is only possible thanks to the transmission of languages from generation to generation. Much of the structure of an I-language is transmitted from parents to children along with the rest of their biological endowment, but obviously the variable parts of language are learned (internalized) from environmental linguistic stimuli. As I have already noted, this is the phase in which mutations in the lexical interface can occur. These mutations, depending on their range of transmission, can give rise to linguistic changes and, ultimately, to what we see as a different language. The task of the child who learns a language is to reproduce in her mind/brain the lexical interface of her interlocutors, a typically insecure ("abductive," cf. Andersen, 1973) procedure that is at the basis of linguistic change.

As Dahl (2004) has shown, the usual dynamics of linguistic change produce an increase in morphosyntactic complexity (*maturity*) up to a certain limit, and thereafter such complexity tends to be maintained. The degree of maturity of a language is measured in terms of the quantity of structures involving a previous derivational history, i.e., non-universal processes that can only be explained by long previous evolutionary chains, such as inflectional and derivative morphology, incorporation, the existence of phonological tone, case marking, or ergativity. However, we might note that according to the model presented in **Figure 1** this natural increase in linguistic complexity actually amounts to an increase in the complexity of the lexical interface, not the whole language itself. In this sense, no languages are more complex than others, but there are languages with more complex lexical interfaces than others. This is an important difference. The notable grammatical differences between, on one extreme, Georgian and, on the other, Tok Pisin, do not imply differences in the deep layers of structure (basically the CI system and the computational system), but rather differences in the historical evolution of their externalization components. The proof of this is that the two languages serve their users in carrying out the same cognitive and communicative functions.

The initial intuition here is simple: the more prior uninterrupted history, the greater morphosyntactic complexity, and vice versa. In fact, McWhorter (2011) argues that the natural state of a language, i.e., when no drastic disturbances in its transmission from generation to generation have occurred, is "highly complex, to an extent that seems extreme to speakers of languages like English" (2011, p. 1). It seems clear that the brain of human children is able to internalize lexical interfaces as complex

as those of Native American languages or Caucasus languages, typical examples of "mature" systems in Dahl's sense. Neither the brains of other organisms nor the brains of the majority of human adults are as efficient in the internalization of arbitrary systems of gender and noun classifiers, agreement patterns, or quirky cases (not to mention phonological systems). Consequently, McWhorter hypothesizes that whenever we find languages with low degrees of morphosyntactic complexity it is because such languages have been interrupted in their normal accumulation of complexity; i.e., languages with relatively low degrees of complexity "owe this state to second-language acquisition in the past" (McWhorter, 2011, p. 2). In this category we could include languages like English, Romance languages, Persian, Mandarin Chinese, and Indonesian. Compared to other, related languages (such as Sanskrit, Latin, Greek, or Baltic) these languages (which McWhorter calls *Non-Hybrid Conventionalized Second-Language Varieties*) are characterized by a loss of complexity that reveals evidence of widespread second-language learning in the past. In fact, Lupyan and Dale (2010) and Bentz and Winter (2013) present quantitative evidence showing that languages spoken by many second language speakers tend to have relatively small nominal case systems compared with languages with low proportions of L2 speakers. According to this model, creoles are extreme cases of the same phenomenon: "where complexity has been lost to a radical degree, we can assume that the language was born in a situation in which adult acquisition was universal" (McWhorter, 2011, p. 2). These cases of suboptimal transmission would therefore be clear examples in which the brains of adult learners have operated as an environmental factor to which some parts of languages have adapted.

Another brain internal potential source of modeling forces for morphosyntactic systems can be found in language use in real time (see Newmeyer, 2005, for a conciliatory synthesis on the division of labor between linguistic and processing principles in grammar development). The model I have presented stipulates that only the externalization component is subject to change and, therefore, to variation. It is therefore expected that processing principles (both in speech production and perception) have a remarkable role in the structure and dynamics of externalization systems (i.e., in the morphological mechanisms of syntax realization), precisely because these systems are relevant to the use of language for communication. In fact, language processing principles (see Hawkins, 2004, for a very explicit model) play their role by relating these two components (the computational system and the lexical interface of **Figure 1**).

Just by way of illustration, I will consider Bickel et al. (2015) regarding the development and persistence of ergative systems in relation to universal processing preferences. Using experimental evidence, Bickel et al. (2015) propose that there is a universal principle that favors the processing of an initial unmarked NP (in nominative or absolutive case) as an agent (as in *John sold a car*). When the rest of the sentence shows that this unmarked NP is not an agentive subject (as it would be in an ergative language, which marks the subjects of the transitive verbs), they observed an event-related potential (N-400) signaling a reanalysis of the role of the first NP (for example, as a patient argument). Bickel et al. hypothesize that this principle is "species-wide and independent of the structural affordances of specific languages"

(Bickel et al., 2015, p. 2) and that, as such, "the principle favors the development and maintenance of case-marking systems that equate base-form cases with agents rather than with patients" (Bickel et al., 2015, p. 2), i.e., nominative-accusative systems over ergative-absolutive ones. Using a large database of linguistic changes in various language families (617 languages in total) they note that of the two possible historical changes, ergative > accusative or accusative > ergative, languages show a clear bias toward the former:

> "Languages tend to avoid ergatives when they evolve over time: if a language has ergative case marking, it is more likely to lose than to keep it, and if a language lacks ergative case marking, it is unlikely to develop it. To be sure, ergative cases can arise and be maintained for a while, but the probabilities of this are always lower than the probabilities of avoiding ergatives" (Bickel et al., 2015, p. 18).

If Bickel et al.'s conclusions are correct, we would again have a clear example of how a language-external (but mind-internal) factor can condition the adaptation of languages in their processes of change. However, this also leads us to an important conclusion, one at the heart of our present discussion: even though a general principle of processing exerts a measurable pressure on linguistic systems, the inertia of the language's previous history is capable of overcoming it, showing that morphosyntactic structure is stubbornly resistant to external adaptive pressures, even though they are internal to the mind/brain and supposedly universal.

It is important to note that ergative systems are mature systems in Dahl's sense, which would also explain, at least in part, both the unequal statistical distribution of the two types of languages, and the historical bias documented by Bickel et al. The relevant fact for us here is that a language like Basque, which is fully ergative, shows no symptoms of maladjustment and remains fully functional for its users. More relevant still, there are processes of historical development of ergativity (otherwise, ergative languages would never have existed), which show that grammatical structure is largely immune to the influence of external (i.e., non-grammatical) factors. Actually, a recent synthesis of the research on processing costs of ergativity in Basque (Zawiszewski, 2017) concludes that there are no profound differences in the mechanisms underlying processing in languages with different case marking systems:

> "In general, the electrophysiological pattern found when processing ergative case violations corresponds to that revealed during similar case violations in accusative languages (…) and thus indicate that the mechanisms underlying language comprehension are comparable across languages with a different case morphology." (Zawiszewski, 2017, p. 706).

CONCLUSIONS

If we adopt McWhorter's theory, we could say that adult brains have influenced the historical development of some human languages to a decisive extent. From an externalist view of languages, it could be said that some languages have adapted to (non-flexible) mature brains, simplifying their historical accretions and rendering themselves easier to be learned. But from the internalist point of view, this statement is unsatisfactory. The externalist approach tends to identify languages with their lexical interfaces, and this identification, at least in part, is behind the different appreciation of the degree of adaptation of languages to the environment. From an internalist point of view, the notion of adaptation of languages to their environment is only acceptable in a weak sense. According to my argument, *weak* means that only relatively superficial aspects of languages can be explained as adaptations to extralinguistic reality. I do not intend to conclude that statements such as the following are incorrect:

> "[L]inguistic differences, from sounds to grammars, may also reflect adaptations to different environments in which the languages are learned and used. The aspects of the environment that could shape language include the social, the physical, and the technological" (Lupyan and Dale, 2016, p. 1).

Of course, as reflected in the model of **Figure 1**, every language has a cultural component (internalized from the environment) that is susceptible to change and, therefore, to vary in relation to external factors (i.e., adaptation). However, statements like the previous one suggest that this process of adaptation is sufficient to explain the structure of languages and their typology, and that conclusion is what I have tried to put into question in this contribution.

Many and diverse external and internal factors have left their mark on languages, especially in their systems of externalization, but I do not believe that this in itself allows us to claim that the structure of languages is essentially a matter of adaptation to the environment.

AUTHOR CONTRIBUTIONS

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

FUNDING

REFERENCES

Andersen, H. (1973). Abductive and deductive change. *Language* 49, 567–595. doi: 10.2307/412063

Baker, M. C. (2003). Linguistic differences and language design. *Trends Cogn. Sci.* 7, 349–353. doi: 10.1016/S1364-6613(03)00157-8

Bentz, C., and Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Lang. Dyn. Change* 3, 1–27. doi: 10.1163/22105832-13030105

Berwick, R. C., and Chomsky, N. (2011). "The Biolinguistic program: the current state of its development," in *The Biolinguistic Enterprise*, eds A. M. Di Sciullo and C. Boeckx, (Oxford: Oxford University Press), 19–41.

Berwick, R. C., and Chomsky, N. (2016). *Why Only Us*. Boston, MA: The MIT Press. doi: 10.7551/mitpress/9780262034241.001.0001

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: evidence from case marking. *PLoS ONE* 10:e0132819. doi: 10.1371/journal.pone.0132819

Cercle Linguistique de Prague (1929). Thèses présentées au Premier Congrès des philologues slaves. *Travaux Linguist. Prague* 1, 5–29.

Chomsky, N. (1985). *Knowledge of Language. Its Nature, Origins and Use*. New York, NY: Praeger.

Chomsky, N. (2007). "Approaching UG from below," in *Interfaces + Recursion = language?" Chomsky's Minimalism and the View from Semantics*, eds U. Sauerland, and H.-M. Gärtner (Berlin: Mouton de Gruyter), 1–30.

Croft, W. (2000). *Explaining Language Change. An Evolutionary Approach*. London: Longman.

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam; Philadelphia, PA: John Benjamins.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1893/2000). *The Autobiography of Charles Darwin*. Amherst, MA: Prometheus Books.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511611896

Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511612060

Dryer, M. S., and Haspelmath, M. (eds.). (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info (accessed September 25 2017).

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–448. doi: 10.1017/S0140525X0999094X

Everett, C., Blasí, D. E., and Roberts, S. G. (2016). Language evolution and climate: the case of desiccation and tone. *J. Lang. Evol.* 1, 33–46. doi: 10.1093/jole/lzv004

Everett, D. L. (2005). Cultural Constraints on grammar and cognition in Pirahã. *Curr. Anthropol.* 46, 621–646. doi: 10.1086/431525

Gould, S. J. (1996). "The pattern of life's history," in *The Third Culture. Beyond the Scientific Revolution*, ed J. Brockman (New York, NY: Simon & Schuster), 51–73.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge MA: Harvard University Press.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how it evolved? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hurford, J. R. (1992). "An approach to the phylogeny of the language faculty," in *The Evolution of Human Languages*, eds J. A. Hawkins and M. Gell-Mann (Reading MA: Addison-Wesley), 273–303.

Kauffman, S. A. (1993). *The Origins of Order. Self-organization and Selection in Evolution*. London: Oxford University Press.

Keenan, E. (2002). "Explaining the creation of reflexive pronouns in English," in *Studies in the History of the English Language: A Millennial Perspective*, eds D. Minkova and R. Stockwell (Berlin; New York, NY: Mouton de Gruyter), 325–354. doi: 10.1515/9783110197143.3.325

Kirby, S. (1999). *Function, Selection, and Innateness. The Emergence of Language Universals*. Oxford: Oxford University Press.

Koerner, K. (ed.). (1983). *Linguistics and Evolutionary Theory. Three Essays. Edited by A. Schleicher, E. Haeckel, and W. Bleek*. Amsterdam; Philadelphia, PA: John Benjamins.

Labov, W. (1963). The social motivation of a sound change. *Word* 19, 273–309. doi: 10.1080/00437956.1963.11659799

Ladd, R. D., Roberts, S. G., and Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annu. Rev. Linguist.* 1, 221–241. doi: 10.1146/annurev-linguist-030514-124819

Lass, R. (1997). *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.

Longobardi, G. (2001). Formal syntax, diachronic Minimalism, and etymology: the history of French *chez*. *Linguist. Inq.* 32, 275–302. doi: 10.1162/00243890152001771

Longobardi, G. (2003). Methods in parametric linguistics and cognitive history. *Linguist. Variat. Yearbook* 3, 101–138. doi: 10.1075/livy.3.06lon

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Lupyan, G., and Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cognit. Sci.* 20, 649–660. doi: 10.1016/j.tics.2016.07.005

Mayr, E. (1942). *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York, NY: Columbia University Press.

McWhorter, J. H. (2011). *Linguistic Simplicity and Complexity. Why do Languages Undress*? Boston, MA; Berlin: The Gruyter Mouton.

Mendívil-Giró, J.-L. (2006). "Languages and species: limits and scope of a venerable comparison," in *The Biolinguistic Turn. Issues on Language and Biology*, eds J. Rosselló, and J. Martín (Barcelona: PPU), 82–118.

Mendívil-Giró, J.-L. (2012). "The myth of language diversity, in language, from a biological point of view," eds C. Boeckx, M. C. Horno, and J. L. Mendívil-Giró (Newcastle: Cambridge Scholars), 85–133.

Mendívil-Giró, J.-L. (2014). What are languages? A biolinguistic perspective. *Open Linguist.* 1, 71–95. doi: 10.2478/opli-2014-0005

Mendívil-Giró, J.-L. (2016). *El cambio lingüístico. Sus Causas, Mecanismos y Consecuencias*. Madrid: Síntesis.

Moreno, J.-C., and Mendívil-Giró, J.-L. (2014). *On Biology, History and Culture in Human Language. A Critical Overview*. Sheffield: Equinox.

Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

Nevins, A., Pesetsky, D., and Rodrigues, C. (2009). Pirahã exceptionality: a reassessment. *Language* 85, 355–404. doi: 10.1353/lan.0.0107

Newmeyer, F. J. (2005). *Possible and Probable Languages. A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago, IL: University of Chicago Press.

Perkins, R. D. (1988). "The covariation of culture and grammar," in *Studies in Syntactic Typology*, eds. M. Hammond, E. Moravcsik, and J. Wirth (Amsterdam; Philadelphia, PA: Benjamins), 359–378. doi: 10.1075/tsl.17.22per

Pinker, S. (2007). *"Response to Recursion and Human Thought: Why the Pirahã Don't Have Numbers", A Talk with Daniel L. Everett*. Available online at: https://www.edge.org/conversation/recursion-and-human-thought (accessed September 25, 2017).

Regier, T., Carstensen, A., and Kemp, C. (2016). Languages support efficient communication about the environment: words for snow revisited. *PLoS ONE* 11:e0151138. doi: 10.1371/journal.pone.0151138

Richards, N. (2016). *Contiguity Theory*. Cambridge MA: The MIT Press. doi: 10.7551/mitpress/9780262034425.001.0001

Schleicher, A. (1863). *Die Darwinsche Theorie und die Sprachwissenschaft*. Weimar, H. Böhlau (Quoted from the 1869 English version *Darwinism Tested by the Science of Language* by J. C. Hotten included in Koerner, ed., 1–71).

Seely, J. (1977). An ergative historiography. *Historiogr. Linguist.* 4, 191–206. doi: 10.1075/hl.4.2.04see

Starke, M. (2009). Nanosyntax: a short primer to a new approach to language. *Nordlyd* 36, 1–6. doi: 10.7557/12.213

Swadesh, M. (1971). *The Origin and Diversification of Language*. London: Routledge and Kegan Paul.

Zawiszewski, A. (2017). "Processing ergativity: behavioral and electrophysiological evidence," in *The Oxford Handbook of Ergativity*, eds J. Coon, D. Massam, and L. Travis (Oxford: Oxford University Press), 693–708.

frontiers
in Psychology

Robust, Causal, and Incremental Approaches to Investigating Linguistic Adaptation

Seán G. Roberts*

EXCD Lab, Department of Anthropology and Archaeology, University of Bristol, Bristol, United Kingdom

This paper discusses the maximum robustness approach for studying cases of adaptation in language. We live in an age where we have more data on more languages than ever before, and more data to link it with from other domains. This should make it easier to test hypotheses involving adaptation, and also to spot new patterns that might be explained by adaptation. However, there is not much discussion of the overall approach to research in this area. There are outstanding questions about how to formalize theories, what the criteria are for directing research and how to integrate results from different methods into a clear assessment of a hypothesis. This paper addresses some of those issues by suggesting an approach which is causal, incremental and robust. It illustrates the approach with reference to a recent claim that dry environments select against the use of precise contrasts in pitch. Study 1 replicates a previous analysis of the link between humidity and lexical tone with an alternative dataset and finds that it is not robust. Study 2 performs an analysis with a continuous measure of tone and finds no significant correlation. Study 3 addresses a more recent analysis of the link between humidity and vowel use and finds that it is robust, though the effect size is small and the robustness of the measurement of vowel use is low. Methodological robustness of the general theory is addressed by suggesting additional approaches including iterated learning, a historical case study, corpus studies, and studying individual speech.

Keywords: adaptation, humidity, tone, vowels, robustness, causal graph

1. INTRODUCTION

The goal of evolutionary approaches to linguistics is to explain similarities and differences between languages. As Bickel (2015) might put it, "what's where why?." The final part of this question—why—is crucial. It requires the demonstration of causal effects including how languages adapt to functional pressures. This is not an easy task. It involves dealing with long causal chains stretching from biology, cognition, and interaction to many different areas of language. It also involves dealing with many possible alternative explanations and the complexities of linguistic history. Some parts of adaptational explanations can be addressed directly with controlled experiments. However, because of the range of timescales involved it is inevitable that some of the steps are addressed with more abstract methods such as modeling, artificial language learning, or historical reconstruction. How can we combine results from such different approaches into coherent evidence for or against a particular theory? Many studies seeking to show adaptation in human language also rely on large-scale global databases. Indeed, we are experiencing a kind of gold rush of cross-cultural statistical studies, where it feels like anyone with a laptop and access to the internet could find the next big

discovery in cultural evolution (Ladd et al., 2015). However, there is not much discussion, within the field of evolutionary linguistics at least, of the general strategy for using these new data and methods to address questions of adaptation. This paper presents one general strategy and discusses its advantages. It outlines concrete, explicit steps which help formulate and communicate questions more clearly and arrive at a clearer understanding of the answers. The main point I would like to make is that, when dealing with cross-cultural statistical methods, there is no single smoking gun that will definitively prove a theory correct, nor a single magic bullet that will disprove it entirely.

The strategy that I will advocate has several features. It is causal, incremental and robust. I will call this the *maximum robustness method*. It is useful to contrast this with what might be called the *maximum validity method*. Briefly, the maximum validity method proceeds as follows:

- Set out a series of specific assumptions and claims that your hypothesis makes.
- Collect and code your data according to your assumptions.
- Run the most valid test given your data and assumptions.
- The outcome is the best answer given the assumptions.

That is, it attempts to perform the most relevant and valid test of a specific hypothesis, then accepts the single result as the best possible evaluation. This may be a caricature of a possible approach to science, but I suspect it is probably the default approach in most individual studies in linguistics. One study that exemplifies this is Hammarstrom (2010), which tests whether the adoption of farming practices leads to higher rates of dispersal and so language families with greater numbers of languages. Hammarström collected data on "ALL attested language families" (about 7,000 languages, capitalisation in the original) and quantifies the number of languages within each given explicit assumptions. Each language was classified as having either an agricultural or hunter-gatherer subsistence type. Then a single, bespoke independent samples test was run to determine the result.

Of course, all statistical analyses should aim to be valid. However, given the range of methods and possible measures available now, it is often difficult to identify the single most valid approach. Indeed, Silberzahn and Uhlmann (2015) gave the same dataset and research question to several researchers to analyse independently. The results varied widely in terms of effect size due to differences in the statistical approaches, yet all of them were defensible. Trying to argue for the most valid approach may lead to arguments from authority rather than logic, so another approach is the maximum robustness method:

- Set out a series of general assumptions and claims that your hypothesis makes.
- Run tests with as many specific assumptions and sources of data as possible
- Demonstrate that all tests give qualitatively the same answer, **or** identify similarities in approaches which lead to negative results.
- The outcome is a space of results that suggests how robust the hypothesis is.

As Levins (1966) put it, "Our truth is the intersection of independent lies." An example of the maximum robustness approach is found in Roberts et al. (2015), which reconsidered the link between future tense and economic variables first proposed by Chen (2013). The correlation in the initial study was strong, but did not account for the effect of shared linguistic history, leaving open the possibility that the correlation was an artefact of Galton's problem. Instead of presenting one methodology which would have provided a single answer (the maximum validity method), Roberts et al. (2015) used nine different statistical methods and two datasets to address the question. They produced a space of results and linked each one to the assumptions of its method. They found that the correlation did appear to be robust, except when the method allowed four key factors: using individual level data as opposed to collapsing within languages; controlling for local economic effects within countries; controlling for cultural descent within language families; and controlling for areal contact. When all these controls were applied the correlation was not significant.

The maximum robustness method also discourages the idea that there is a single best analysis which definitively proves or disproves a theory. The space of results should tell us more than simply that the first paper was flawed: it suggests that collapsing information within languages loses some important aspects of the data, and that all three of the historical processes are at play in human cultural evolution (see also Moran et al., 2012). Furthermore, the ultimate suggestion of the paper was that large-scale, cross-cultural statistics was not the best approach for addressing this question due to the complexities of the confounding factors, and instead future research should concentrate on localized experiments, which are quite feasible in this case (Thoma and Tytus, 2017).

Aspects of both approaches are, of course, part of the ideal scientific method, particularly the careful expression of assumptions from the maximum validity method, and the repeated testing of the maximum robustness method. However, due to limited resources or data, most studies tend to gravitate toward maximum validity. In particular, the complexity of running many different tests in the robustness approach and the difficulty of reconciling conflicting results makes the maximum robustness method difficult to conduct in a single paper. I will argue that the robust approach is worth it, but also probably needs to be combined with an incremental approach in order to be effective. The combination of robust and incremental approaches is particularly useful for studies of cultural evolution where there may be a long chain of causal connections that span many disciplines and a large range of appropriate methodologies to address each link.

The paper is organized as follows. The rest of this section summarizes the hypothesized adaptive link between tone and humidity which will be used as a case study in the rest of the paper. In sections 2–4, the features of the maximum robustness approach are presented. Section 2 shows how causal graphs can be used in a six-step process to map out an explicit expression of a hypothesis, its implications and potential confounds. Section 3 discusses the idea of incremental research and how gradual progress toward support for a hypothesis

is the most pragmatic approach. Section 4 discusses the final part of the approach, which is robustness: converging evidence from many angles provides the best way of supporting adaptive hypotheses.

Sections 5–7 apply some of the ideas from the maximum robustness approach to the hypothesis linking tone and humidity. Section 5 attempts to replicate previous analyses with an alternative source of data to ensure their robustness. Section 6 suggests some hypothetical ways that other links in the causal chain linking humidity and tones could be tested. Section 7 summarizes the current state of the hypothesis given the evidence presented here. Section 8 provides a brief conclusion.

1.1. Case Study: Linguistic Adaptations to Humidity

Everett et al. (2015) suggested that the distribution of languages that use lexical tone across the world could be predicted by humidity. This work will be used as a case study to illustrate the maximum robustness method. The choice is not intended to suggest that it is robust. Indeed, it is because this is a controversial idea that it serves as a good example and would benefit from the maximum robustness approach.

The idea of trying to explain properties of language as being adapted to external climatic influences goes back a long way. As far back as the late eighteenth century, de Rivarol suggested that languages are "melodious and voluptuous in mild climes, harsh and dull under a sad sky" (de Rivarol, 1784), and Lord Monboddo hypothesized a link between laryngeal desiccation, production, and the distribution of particular sounds in language:

> "But the total want of P and W may be looked on as the grand literal distinction, between the Scandinavian and the German dialects of the Gothic. And this seems a remarkable instance of the effect of climate upon language; for P and W are the most open of the labial letters; and V is the most shut. The former requires an open mouth: the later may be pronounced with mouth almost closed, which rendered it an acceptable substitute in the cold climate of Scandinavia, where the people delighted as they will delight, in gutturals and dentals. The climate rendered their organs rigid and contracted; and cold made them keep their mouths as much shut as possible."
> (Pinkerton, 1789, p. 19)

These are, of course, too limited (and poetical) to constitute substantial, rigorous evidence for the proposed link, but modern databases and statistical methods allow us to test these hypotheses quantitatively. This has been done for links between climate and phonetics (see Munroe et al., 2009; Maddieson and Coupé, 2015). More recently, Everett et al. (2015, 2016a) reviewed research from laryngology showing that dry air affects the vocal folds, making careful control of pitch difficult. There are many cases of animal signals adapting to environmental conditions such as humidity (e.g., Snell-Rood, 2012). This suggests that human languages would also adapt to the local humidity over long periods of time so that careful control of pitch (e.g., complex lexical tone systems) would be rarer in drier regions. This was tested in a sample of around 3,000 languages and moderate support was found for

the hypothesis that complex tone languages were rarer in drier regions.

The more general theory that desiccation affects the vocal folds was recently extended in a paper in this issue to predictions about vowels. Since vowels require more precise control of vocal folds than consonants, they should also be relied upon less in drier regions. Accordingly, Everett (2017) shows that speakers in drier regions use vowels less frequently in their basic vocabulary.

This theory fits within the "distributional typology" approach, which attempts to explain patterns in typological variables as causal effects from functional pressures or historical events using statistical analyses (Bickel, 2015). However, most previous analyses in this vein have concentrated on functional pressures from cognition (e.g., Bickel et al., 2015), rather than physical pressures from the ambient climate. Accordingly, the link between tone and humidity has been criticized on many grounds, both methodological and theoretical (see Everett et al., 2016a and responses: Collins, 2016; de Boer, 2016; Donohue, 2016; Ember, 2016a; Gussenhoven, 2016; Hammarström, 2016; Ladd, 2016; Moran, 2016; Progovac and Ratliff, 2016; Winter and Wedel, 2016). The aim of this paper is not to address those criticisms, but to attempt to use this research to illustrate the maximum robustness method. Section 5 tests the robustness of the claim about tones and section 6 tests the robustness of the claim about vowels.

2. CAUSAL GRAPHS

This section presents the six steps for using causal graphs in a maximum robustness approach to research. The first step of this approach (and many others) is to be explicit about the causality of the claims in the hypothesis. This seems like a trivial requirement for any investigation, but is a subtly difficult challenge (defining causality itself is tricky, and I avoid doing so here, but see Blasi and Roberts, 2017 for a discussion relating to humidity and tone). As every researcher knows, discovering a simple correlation is not the same as proving a causal link. The gold standard for demonstrating causality is a controlled experiment, but often correlations are the first step toward this ultimate goal. More importantly for this paper, behind each study there should be at least a hypothesis about a causal effect, and that is what this section is interested in capturing. One of the clearest methods for defining causal relationships is by using causal graphs (e.g., Pearl, 2000). Nodes represent variables and edges represent casual processes. As well as an investigative methodology in its own right, causal graphs can be used as a tool for helping researchers to think about their hypotheses and to guide the direction of research. For example, the basic causal claim in Everett et al. (2015) is that ambient humidity causes a change in the number of distinctions in tones. However, this leaves out many processes in between and many other possible explanations of a statistical link. Here, I will suggest a number of steps to help arrive at a full causal picture of the domain of the hypothesis.

Step (1) Draw the main causal link between the elements of the hypothesis.

These are usually the measurable variables mentioned in the prose formulation of the hypothesis. In this case:

Ambient Humidity → Fewer tones

This is depicted in **Figure 1**.

Step (2a)   Break down the main causal link into more fine-grained causal links.

The goal is to identify more local links to spell out a more detailed description of the causal process. This involves being more explicit about the physical causality, and often in the case of global statistical studies involving adaptation, about the mechanisms of propagation and diffusion. The result is a chain of causal links. For the case of humidity and tone, one could imagine the chain of causal effects based on production effort:

Ambient Humidity → Laryngeal desiccation → Production costs→ Frequency of tokens → Cultural diffusion → Fewer tones

That is, ambient humidity causes desiccation of the larynx and vocal folds, which affects production (finer control of fundamental frequencies requires more production effort). This leads to a change in the frequency of tokens (fewer tokens involving complex pitch). Through cultural diffusion, this could lead to a change in the linguistic system as a whole so that there were fewer distinctions in lexical tone.

Step (2b)   Consider alternative causal pathways between the elements of the chain that would also support the hypothesis.

The goal here is to imagine alternative pathways between the main causal variables, or any of the other links already described. In the case of tone and humidity, there is an alternative pathway involving *interaction*: The affects of desiccation on production leads to weaker distinctions being transmitted, this influences perception in the listener, leading to miscommunication, and eventually to a selection pressure against fine tonal distinctions (**Figure 1**).

Step (3)   Asses the current evidence for each causal link.

For each causal link, is there causal evidence that supports it? The best evidence might come from controlled experiments, but may also include causally informed statistical work or theoretical work. For example, there are several experiments that relate to the causal links between humidity and tone (see Everett et al., 2015, 2016a). For example, Hemler et al. (1997) demonstrate that humidity affects vocal fold vibration accuracy. Leydon et al. (2009) and Sundarrajan et al. (2017) demonstrate that vocal fold vibration causes changes in production. At the same time, criticisms of each causal link can be added to suggest negative evidence (**Figure 1**).

Step (4)   Place the causal graph in a wider context.

The next step is more challenging. It requires thinking outside of the narrow focus of the hypothesis into any other possible causal links that might interfere with the main causal pathway. Two types of link in particular should be sought. The first is anything that directly affects the final variable (i.e., number of tones). The second is any series of links that provide an alternative causal link

between the two main variables that are not part of the general hypothesis. The goal is to find any causal link going from new variables to the final causal variable that, to put it technically, are not d-separated by the first causal variable. That is, there are plenty of things that affect humidity, but what we are interested in is things that affect humidity and also the final variable, possibly with intermediate steps. It is likely that these links will come from outside of the field of linguistics (Bickel and Nichols, 2006).

One possible alternative pathway is a direct effect of humidity on sound transmission via sound absorption. This link is well-understood at a basic physical level (humid air conducts higher frequencies better, Bass et al., 1984) and many animal communication systems show adaptation to this constraint (Snell-Rood, 2012). It is unclear whether this would cause the same selection pressure as the production effort caused by laryngeal desiccation, but at this point it is worth considering. Another possible pathway is a direct effect of humidity on perception. There is some weak evidence that repeated exposure to dry, cold environments damages the ear in a way that could influence perception (Morgan, 1954). This is an unlikely explanation, but at this point the goal is to list possible causal links, not to evaluate them.

An example of a wider context was suggested in Everett et al. (2015), and is redrawn in **Figure 2**. It includes links from the literature on climate, disease and migration (Michaelowa, 2001; Ember, 2016a). The climate affects various demographic and disease-related variables which contribute to the likelihood of contact between languages and so possibly the eventual borrowing of tones in some climatic regions but not others. This could explain a statistical link between the climate and the distribution of tone that is not part of the core claim of the original hypothesis.

Other alternative pathways include a link between the ecology (density of foliage in the environment) and acoustic transmission. This has been explored as an alternative hypothesis linking the environment and linguistic sounds (Morton, 1975; Fought et al., 2004; Ey and Fischer, 2009; Munroe et al., 2009; Maddieson and Coupé, 2015; Coupé, 2017; Maddieson, under review). The ecology may also influence the kinds of meanings that speakers need to talk about and the semantic, pragmatic and social distinctions that are important to them (Regier et al., 2016), which may affect the frequency of tokens.

Step (4)   Identify possible confounds.

Given this wider picture, it should now be possible to identify causal factors that provide alternative causal explanations for a correlation between the two main variables. The causal graph may now be quite complicated, but we can use tools from causal graph theory to focus our attention on relevant potential confounds. For example, the wider causal graph above includes a large number of variables to do with demography and disease. However, the only place where this influences the main causal pathway is through contact. Therefore, if we can somehow control for the influence of contact on diffusion, then it follows that controlling for the demographic and disease variables is redundant. This is a Markov causal condition which is one of the fundamental parts of causal graph theory: variables can only

FIGURE 1 | The first three steps in the causal approach. Left: Step 1, starting with the main causal claim. Middle: Step 2, breaking the main claim down into a chain of links. Right: Step 3, assessing criticism and support. Criticisms are listed on the left and supporting evidence is listed on the right. Question marks represent no supporting evidence for the particular link.

FIGURE 2 | Steps 4 and 5 of the causal approach. Additional nodes represent the wider causal context. Nodes and arrows in black indicate potential confounds that need to be addressed.

be influenced by directly connected ancestors (this assumes that the causal graph drawn by the researcher is correct). This is an important point when considering control variables. Not every variable which is *correlated* with the main dependent needs to be controlled for, only those with a plausible direct causal influence.

Figure 2 shows the alternative pathways with the variables that interact with the main causal pathway highlighted. There are potentially many more confounding factors (e.g., having the right conditions for tonogenesis), but the point here is to demonstrate how thinking with causal graphs helps to make concrete the claims of a hypothesis and identify possible confounds.

Step (6)   Choose the next link to research.

Given the final causal graph, it should now be possible to identify the next best step in the research programme. In the current example, it is clear that the question of historical diffusion and the confound of borrowing needs to be addressed. Beyond that, other suggestions are presented. For example, the interaction, sound absorption and perception pathways all rely on creating problems in miscommunication. Therefore, investigation into those mechanisms might begin with that link. More generally, the production effort pathway requires fewer causal steps, and so might be easier to investigate first. It is also possible that the evaluation of evidence and confounds will suggest that the hypothesis is not worth pursuing at all.

2.1. Advantages of the Causal Approach
Producing a causal graph such as the one above has several advantages for the large-scale statistical studies in linguistics.

2.1.1. Clear Communication of the Hypothesis
Expressing hypotheses as detailed causal graphs forces researchers to be explicit about their claims. This avoids confusion and focuses criticism on specific issues. Together with an empirical approach, this should lead to more productive debate between researchers, because criticisms can address assumptions and data on particular points, rather than criticising a whole approach or the author themselves. One of the weaknesses of the maximum validity approach is that it relies on the judgement of the authors about what the most valid approach is. If a critic disagrees on the choice of a particular step in the analysis, it is difficult to interpret the value of the result. **Figure 1** links some of the criticisms to particular links in the causal chain, indicating where improvement needs to be made.

2.1.2. Identification of Strong and Weak Links in the Causal Chain
By linking evidence to particular causal links, it should become clear which parts of a hypothesis are well supported and which require more investigation. Regarding humidity and tone, there is already experimental evidence for many of the early steps in the causal chain. There are three broad regions that remain to be tested. The first is the link between production costs and frequency of tokens, either directly or through interaction. The second is the link between frequency and the current distribution of tone systems in the world through cultural diffusion. The third is the potential confounding influence of other factors, particularly borrowing.

2.1.3. Identification of Possible Confounds
The procedure above encourages an attempt to think of possible confounds and identify where in the causal chain they might apply. In the section above, the Markov causal condition was discussed which means that not all variables involved in alternative accounts necessarily need to be controlled for. This saves time and focusses research on relevant issues. It is worth noting that accounting for alternative influences on the key variables does not always reduce statistical power. In some cases, it may account for other noisy processes and reveal a causal effect in the main causal chain.

2.1.4. Deconstruction of the Problem into Sub-hypotheses That Can Be Addressed Separately with Different Methods
The first 2 weak areas of the causal chain above may not be amenable to strict experimental control. In particular, the diffusion of linguistic variants is hard to study directly because of the timescales involved. However, the advantage of creating this causal graph is that it breaks the investigation down into smaller links, and each of these links can be investigated in its own right with the most appropriate methods and data. While physical acoustics and laboratory phonetics methods can be applied to the initial parts of the chain, there are more appropriate methods for the later parts including computational modeling (Kandler and Steele, 2008; Gavin et al., 2017), artificial language learning experiments (Tamariz, 2017b), historical corpus analyses and historical computational techniques such as phylogenetic ancestral state reconstruction (Gavin et al., 2013; Honkola et al., 2013).

One clear example of this modular approach is in the recent research into the link between genetics, vocal tract morphology, sound production, and global distributions of sound inventories (Dediu et al., 2017). The hypothesis was expressed as a chain of individual links, where each link was addressed with the most relevant method. For example, the first causal link is between genetic differences and individual differences in vocal tract anatomy, such as the shape of the hard palette. This was investigated with clinical measurements and backed up by evidence form developmental biology (Dediu and Moisik, 2016). Those physical differences have small effects on the effort required to produce particular sounds, causing biases in speech production. This was tested with a computational model of biomechanics and a cross-cultural phonetic learning experiment (Moisik and Dediu, 2017). The biases are amplified by cultural evolution into phonetic differences at the population level. This was tested by using the biomechanical model as an agent in an iterated agent based model and testing the effect of multiple generations of diffusion (Janssen et al., 2016). This predicts that physical differences cause the patterns of phonological inventories that we see in the world, which was tested on a database of worldwide phonology (Dediu et al., 2017).

2.1.5. Guidance for Incremental Approaches

The causal graph, together with the evaluation of current evidence and potential confounds, should suggest the next steps for testing the hypothesis. Researcher resources are limited, and not all avenues can be explored. This method helps identify the most pragmatic way forward. This aids an incremental approach, which is discussed below.

3. INCREMENTAL RESEARCH

I argue that research into cultural evolution should be incremental in three senses. First, it should build upon existing theories, typologies and knowledge from linguistics and other fields, rather than use new approximations that fit the data or model. This is not entirely straightforward to assess. For example, for many historical and descriptive linguists, the link between the physical climate and phonology was new and apparently motivated by spotting a pattern in the world. However, from a background in laryngology, acoustic physics or animal communication, the theory is a logical progression of some well-known phenomena.

Secondly, there is no need for every paper to prove the theory in its entirety. Instead, it is best to see a theory as a causal chain with many links, and researchers can investigate one link at a time. Each link may be best addressed by different methods and data (see above). Indeed, with recent advances in digital data curation, it is now possible to constantly update data and analyses. For example, PHOIBLE (Moran et al., 2014) and Glottolog (Hammarström et al., 2017) are constantly updated through github (see https://github.com/phoible/dev and https://github.com/clld/glottolog). We need no longer see a paper as the definitive last word on a dataset.

Finally, research might move from correlational to causal evidence in stages. Realistically, researchers will start with links that are easier to demonstrate given current data and advance toward more definitive, carefully controlled evidence. For example:

1. Demonstrate a synchronic relationship.
2. Demonstrate a diachronic relationship.
3. Demonstrate experimental evidence.

In parallel, researchers should attempt to elicit and disprove alternative explanations. Given the complexity of working between multiple fields, this will also be an incremental and interactive task. For example, based on criticisms and suggestions by Hammarström (2016) of the statistical methodology in the original paper on humidity and tone, Everett et al. (2016b) improved the method and re-ran the statistics (see below). It may be much easier to demonstrate confounds in a study than to correct for those confounds, which might mean that the possible criticisms of a hypothesis might develop much more quickly than the positive evidence for the hypothesis. One example of this comes from work in Collins (2016), which includes a computational simulation of a confounding mechanism (the diffusion of tone through local borrowing) before a simulation of the climatic hypothesis was developed. Given the slow progress

of studies with new methods, it would be rash to dismiss (or fully accept) the original idea on the basis of a single study, and the incremental method advises patience on the part of researchers.

Indeed, one way to see early correlational studies is as "feasibility studies." Everett et al. (2015) do not actually test any of the intermediate causal steps outlined in the causal graph above, but instead simply show a correlation between the variables at either end of the chain. This kind of study may be still be worthwhile, in particular for new avenues of research, in order to establish basic plausibility. If all the causal links suggested by the hypothesis hold, then we should expect to see a correlation between the two main variables. Finding such a correlation provides a motivation (to researchers and their funders) to investigate further with potentially more costly or more time-consuming methods. Of course, a key question for studies that play this kind of role is the robustness of their claims, an issue to which we now turn.

4. ROBUSTNESS IN CROSS-CULTURAL STATISTICAL RESEARCH

This section discusses different kinds of robustness and how they relate to cross-linguistic analyses. Types of robustness discussed include measurement robustness, structural robustness, representational robustness, methodological robustness, estimation robustness, and robustness against *ad hoc* hypotheses. The section ends with a short summary of how the causal thinking, incrementality, and robustness can be combined to form the maximum robustness approach.

Robustness is a term used in many areas of research, but particularly in the use of computational and statistical modeling (Levins, 1966; Weisberg, 2006; Weisberg and Reisman, 2008; Wimsatt, 2012). Robust results are ones that hold under a range of assumptions. Seeking robustness is desirable when a models makes assumptions about various processes and quantities that cannot be confirmed in the real world. Weisberg and Reisman (2008) discuss different kinds of robustness based on different kinds of assumptions: structural robustness relates to assumptions about the causal structure of a model and parameter robustness relates to the range of model parameters under which a result holds. Macro economics studies often test the stability of an estimate of the strength of a relationship between two variables in a regression when adding a range of alternative control variables (Leamer, 1985). There is also some discussion about whether robustness provides proof of a causal relationship as opposed to a mere correlation, though Woodward (2006) is doubtful that this is logically sound.

In cross-cultural statistical analyses, there are many different kinds of assumptions that could affect a result, and so many types of robustness which might be desirable. The sections below discuss some of them, moving from well-established types such as measurement robustness, structural robustness, and representational robustness to a discussion of some types of robustness that apply particularly to theories of cultural adaptation in linguistics (methodological robustness, estimation robustness, robustness against *ad hoc* hypotheses).

4.1. Measurement Robustness

Woodward (2006) discusses *measurement robustness*: the measurement of a variable is robust if different independent methods or measurement events agree. Most psycholinguistic studies that involve manual coding often assess reliability by comparing the judgements of multiple independent coders (e.g., using Cohen's κ, Cohen, 1968). Often in large-scale statistical analyses, we assume that the measurement of the variables is accurate and unbiased, but without trying to confirm this. Because quantifying aspects of a linguistic system is not an entirely objective process, it is likely that there are biases in measurement based on factors such as the theoretical background of the linguist (Moran, 2012, 2016; Easton et al., 2015). However, testing reliability is difficult due to the scarcity of multiple, independent sources for global linguistic data and the difficulty of finding proficient coders for some languages (though fluency is not always needed, see Dingemanse and Enfield, 2015). Additional independent measures are not possible for extinct languages with only one source. However, for some variables there are independent measures. For example, there are at least two databases counting the number of tones in a language (PHOIBLE, Moran et al., 2014, and the ANU phonotactics database, Donohue et al., 2013, see also Allison et al., 2006). Section 5.2.1 tests the robustness across these databases.

Beyond checking that the measures correlate, it is also important to test whether the measures are systematically biased for a particular language family or area, or according to the main dependent variable, which is also done below. This issue also applies to typological interpretation of primary sources. For example, the Glottobank project (http://glottobank.org/) is constructing a typological database of language structures based on primary materials such as grammar descriptions. The reliability of codings from multiple coders was measured.

It is likely that measurement robustness will be better for more concrete, lower-level features than for high-level categories. The "multivariate typology" approach suggests that high-level typological categories often do not capture the full similarities and differences between languages, and instead encourages linguists to break down abstract distinctions into "maximally fine-grained features" (see Bickel, 2010, 2011; Bickel et al., 2014). For example, instead of classifying a language as having "SVO" word order, that category can be broken down into different features that encode the word order in different contexts. Studies have shown that it is possible to do this to distinguish between dialects (Spruit, 2006) and for other domains such as for phonology (Macklin-Cordes and Round, 2015). Probabilistic typologies go one step further by coding the probability or frequency with which a particular construction is observed, building in an inherent measure of uncertainty (Bickel et al., 2009).

The maximum robustness approach differs from the maximum validity approach with regards to the importance of measurement robustness. The maximum validity aims to cover as many languages as possible with a particular typology, prioritising collecting data on currently uncoded languages. The maximum robustness approach instead advocates obtaining independent measures of currently coded languages.

It also makes sense to test whether results are robust when using alternative datasets. That is, does the correlation between humidity and tone hold in both the ANU database and the PHOIBLE database? Since the datasets might not overlap entirely, this is not exactly measurement robustness, but the same principles apply—the more often a correlation is replicated over different sources of data, the more certain we can be that the correlation is meaningful. The sections below test whether this is the case for humidity and tone. The measurements of humidity are also not guaranteed to be totally valid, since they are based on climate models for which there are alternatives, but we do not address this in this paper.

Studies that control for linguistic history also make assumptions about the historical relatedness of languages, most basically which language family a language belongs to. The Glottolog database (Hammarström et al., 2017) is emerging as the leading authority on this, and is particularly useful because it has an explicit set of assumptions behind its classification. However, other classifications exist, and some studies run statistical tests using alternative classifications to check that the result remains similar (e.g., Torreira et al., 2014). Cross-linguistic analyses also make assumptions about the identity of languages. For example, identifier codes are used to link data between databases (ISO code, Glottocodes). Different sources can disagree about the identification of a particular variety, or have errors in matching. Identifying errors and robustness is difficult, but one approach might be to cross-reference the identifier codes with independent measures of their geolocation. Some languages are spoken over large areas and there are justified disagreements on where to place point locations, but the majority of languages are small and well represented by a point. For example, when matching up languages in the ANU phonotactics database with those in the PHOIBLE database, the distance between the stated geographic coordinates is below 500 km for over 95% of languages (2% of languages differed by more than 1,000 km, which might represent problems).

4.2. Structural Robustness

Most of the robustness tests in macro economics papers relate to whether the main result of interest still holds under a range of controls for potential confounding factors, what is referred to as a *sensitivity analysis*. This kind of robustness is most closely related to structural robustness (Weisberg and Reisman, 2008), since it relates to the structure of the statistical model. Identifying the relevant control variables is not easy. Procedures such as systematic literature reviews help to identify potential confounding variables in a systematic manner (see Bero et al., 1998; Khan et al., 2001; Liberati et al., 2009). The causal graph approach above aims to help this process, particularly in identifying variables that do not need to be controlled for. This process is also becoming easier with the rise of meta-databases of statistical results such as *Metalab* (Lewis et al., 2015) and the *Explaining Human Cultures* database (Ember, 2016b). An database of causal links in evolutionary linguistics is currently in development (Roberts, 2018).

Minimally in cross-linguistic research a control for historical influence is needed. For example, it is now standard to use a

language's family as a random effect in regression models, and many papers use geographic areas as a control for horizontal contact. In order to apply certain controls, it may be necessary to implement different methods. **Table 1** summarizes the tests done on the correlation between future tense and economic decisions in Roberts et al. (2015). Inclusion of different control variables affects whether the correlation is significant, suggesting that it is an artefact of historical processes (see also Mavisakalyan and Weber, 2017 for a wider review of studies).

Further options in statistical analyses could affect a result. For example, in mixed effects modeling there are different approaches to testing for significance, including comparing the overall fit of nested models (with either "forwards" or "backwards" comparison) or looking at the estimations of a coefficient within a full model (see Roberts et al., 2015 for a comparison). There is little agreement on these, and best practices appear to differ by discipline. Even the most sensible random effects structure is often debated (Barr et al., 2013; Bates et al., 2015).

Woodward (2006) and Hoover and Perez (2004) note that ensuring structural robustness is a hard problem, especially since results may be sensitive not just to the set of control variables, but to the particular combination of control variables, causing an exponential explosion of possible control models. Barth and Kapatsinski (2017) suggest that this is a real problem for linguistics because aspects of language are highly redundant and inter-related. Instead of committing to one "best" model for the final results, Barth and Kapatsinski (2017) suggest a "multimodel inference" approach, which assesses the hypothesized relationship in a wide range of models.

Another option which is becoming more tractable is to give an unbiased statistical model free reign to pick and choose the particular variables that it tests in order to explain the variation in a target variable. There are some methods from machine learning that provide this kind of option. For example, Slonimska and Roberts (2016) predicted that /w/ and /h/ sounds at the start of a turn would be a good predictor that the next turn would be a question in a corpus of English conversation (because many interrogative words start with /w,h/). Instead of testing the proportion of questions that begin with /w,h/ vs. ones that do not, Slonimska and Roberts (2016) allowed a decision tree algorithm to divide the full set of phonemes in English in any combination that best predicted the distribution of questions. In line with the author's predictions, and in spite of a large number of other possibilities, the tree found that separating turns beginning with /w,h/ from the rest was an efficient way of identifying questions. This essentially provides an unbiased (or at least sociologically unbiased) approach to the hypothesis and expands the space of alternative hypotheses considered, without facing a combinatorial explosion.

4.3. Methodological Robustness

Weisberg and Reisman (2008) discuss the notion of *structural robustness* in modeling: if the same core components cause the same result across a range of alternative models, then the results are robustly due to those core components. Irvine et al. (2013) extend this notion to include the ability to compare results from models and lab experiments: abstract computational models allow precise specification and transparency, but the representation of cognition may not be realistic. In contrast, lab experiments with human participants use realistic cognition (real human brains), but the precise mechanisms are not transparent. However, if the same results are observed across the two methods, then we can conclude that the core causal components that are shared between the models are robustly responsible for the result. We can extend this further to apply to a wider range of methods, what might be called *methodological robustness*: if the same result is obtained from a wide range of methodological approaches (models, lab experiments, corpus studies, etc.), then we can be increasingly certain that the result is not due to the particular assumptions of a given method. For example, the cultural evolution of compositional structure in language as a product of pressures for compression and expression in iterated learning has been demonstrated in computational models and lab experiments (see Irvine et al., 2013; Kirby et al., 2015). In another example, Slonimska and Roberts (2016, 2017) use cross-cultural typology, corpus analyses, and psycholinguistic experiments to provide robust evidence for the idea that forms of interrogative words adapt to the pragmatic requirements of conversation.

Methodological robustness depends on there being a general theory that can produce hypotheses for many particular cases. For example, the hypothesis regarding humidity and tone may derive from a more general theory about how humidity affects vocal production. The more general theory can produce a range of hypotheses such as communities in drier climates using fewer vowels, or individual speakers using different tones in different parts of the year. Later sections of this paper discuss some concrete ways to test the hypothesized link between humidity and language using a wider range of methods.

4.4. Representational Robustness

Weisberg and Reisman (2008) also discuss *representational robustness*: whether a result holds when a computational model represents particular aspects using a different representational schema. The most obvious application is to test whether the same conceptual model provides the same results when implemented in two different programming languages. If so, we can be more confident that the result is not due to a particular intricacies of a particular programming language. Roberts et al. (2015) found that results could differ substantially between running the stats on different operating systems, due to small bugs in the code for the *lme4* package (since fixed, see Roberts et al., 2015). Representational robustness is often sought when the methods become complicated in order to ensure that the procedures are correct. For example, Everett et al. (2016b) implemented the statistical tests in both R and Python. Similar results suggest that there were no procedural errors in either. However, this may be better thought of as kind of check on the validity of an analysis in these cases, rather than a check of robustness.

4.5. Estimation Robustness

All statistical analyses make some assumptions about the statistical procedure. However, robustness does not relate to

TABLE 1 | Summary of the statistical tests in Roberts et al. (2015) relating future tense to economic choices.

Test	Data aggregated?	Controls			Correlation significant?
		Language family	Geographic area	Country	
Mixed effects model	No	Yes	Yes	Yes	No
Regression on matched samples	No	Yes	No	Yes	Yes
Serendipity test	No	Yes	No	Yes	Yes
Independent samples	Yes	Yes	No	No	Yes
Partial Mantel test	Yes	Yes	Yes	No	Yes
Partial Stratified Mantel test	Yes	Yes	Yes	No	No
Geographic autocorrelation	Yes	No	Yes	No	Yes
Phylogenetic Generalized Least Squares	Yes	Yes	No	No	Yes
PGLS within families	Yes	Yes	No	No	No

many assumptions such as the normality of the data because that can be verified for a particular dataset. Instead, robustness relates to assumptions which we cannot verify or for which decisions are somewhat arbitrary. This relates to choices like the statistical framework that is chosen and the particular optimizer used to estimate the coefficients in a regression model. These issues are most similar to the concept of parameter robustness, though that relates to parameters of the mechanistic model. Therefore, this kind of robustness may be termed *estimation robustness*: invariance of the results to assumptions about the statistical estimation. For example, Roberts et al. (2015) compared the results of the same model structure with different kinds of assumptions in the estimators (linear mixed effect models in *lme4*, Bates et al., 2011; Bayesian mixed effect models in *blme*, Dorie, 2011) and demonstrated that results differed considerably, suggesting that the correlation was not robust. Little attention is paid to whether results are robust to changes in the optimizer algorithm within a particular framework. This is mostly justified, since it is unlikely to make a difference, but explicit testing is also possible. The analyses in study 3 below use two different mixed effects modeling frameworks and seven different optimizer algorithms to demonstrate the robustness of the result.

4.6. Robustness against *ad hoc* Hypotheses

The age of large-scale databases and cheap computation has some dangers: it is easy to test a wide range of relationships between variables without having an *a priori* theory which would predict it. It would be possible to search for strong correlations and then invent an *ad hoc* hypothesis to suit them. Alternatively, researchers may come across a strong correlation by chance and focus their research on it, when a wider view of the domain would have lead them to test different hypotheses. The origin story of the link between lexical tone and a particular genotype may be such a case (Dediu and Ladd, 2007). How can we make sure that a result is robust to this fallacy? Obviously, transparency and honesty apply, but these are not exact methodologies. One approach taken by Dediu and Ladd (2007) and also by Roberts et al. (2015) is to assume that the relationship between the two variables of interest should be stronger than the relationship between one of them

and a set of other variables that could have been considered (a "serendipity test"). Roberts et al. (2015) tested whether economic decisions were more strongly correlated with future tense than with any of the other variables in the World Atlas of Language Structures. This is kind of the opposite of controlling for multiple comparisons: controlling for the tests *which could have been done*. In both of the publications, other correlations were reliably weaker, providing evidence that pursuing the hypothesis may be productive.

4.7. Combining the Causal, Incremental, and Robust Approaches

Combining the approaches from the sections above provides the specification for a maximum robustness approach. The causal structure of hypotheses should be explicitly defined using causal graphs. This should point the way to the next most useful analysis. This analysis should tackle a sub-part of the causal graph with the most appropriate method. The analysis should not aim to definitively prove or disprove the hypothesis, but provide incremental evidence for or against it. Individual studies should attempt to demonstrate at least structural robustness and estimation robustness. Reviewing evidence from multiple studies and multiple methodologies will contribute toward methodological robustness (and maybe measurement and *ad hoc* robustness).

The disadvantage of this approach is that it is unclear how to assess theories when evidence from different studies does not agree. For example, when discussing sensitivity (structural robustness), Leamer (1985) suggests an "extreme bounds" approach: the correlation should be considered non-significant if it is not significant in any single test. Sala-i Martin (1997) points out that, given the massive number of possible control tests, this is too strict, and suggests a threshold for significance such as 95% of tests being significant. The aim is not to try to break the correlation in order to disprove it, but to break the correlation in order to learn more about why it is observed.

In the next sections, I try to apply this approach to the link between humidity and tone, particularly regarding measurement, methodological, and estimation robustness.

5. TESTING THE ROBUSTNESS OF THE LINK BETWEEN HUMIDITY AND TONE

Given the discussion above, one of the most pressing issues for the link between humidity and tone is the robustness of the initial statistical correlation. In the next few sections, I present some replication studies with an alternative dataset, and also some hypothetical future studies that could address some of the specific causal links. Study 1 replicates the initial correlation between humidity and tone from Everett et al. (2015) using an alternative dataset. Study 2 looks at a continuous measure of tone. Study 3 extends Everett (2017)'s study of humidity and vowel use using two phonological datasets and one phonetic dataset. All data, analysis scripts and results are available in an online repository: https://github.com/seannyD/HumidityToneReplication.

5.1. Study 1: Replication of Percentile Test with Alternative Dataset

The statistical tests from Everett et al. (2015) (and further refined in Everett et al., 2016b) used linguistic data from the ANU phonotactics database. These tests can be replicated using measures of tone from the PHOIBLE database (Moran et al., 2014). Data on the number of tones for 1,100 languages was obtained from PHOIBLE and linked to the humidity data from Everett et al. (2015) (several sources in PHOIBLE such as UPSID do not code tone, and these were excluded). As in Everett et al. (2015) languages were divided into complex (three or more tones) and non-complex (two or fewer tones) languages.

Figure 3 shows the data from the two linguistic databases side-by-side. There are some differences, but the main pattern is the same: languages with no tones are more frequent at lower humidities than languages languages without tone, and the distributions are more similar in the more humid region. This difference in dry regions only presents a problem for statistical methods which test for a difference in means (see Blasi and

Roberts, 2017 for a discussion), which is why an alternative test was formulated.

The procedure for what was called "test 3" in Everett et al. (2016b) tests whether the size of the difference in the 25th percentile of humidity between a sample of complex tone languages and non-complex tone languages is greater than a baseline sample of languages. Sampling is done so that languages are independent in both language family and area. This is an improved version of the test in Everett et al. (2015) based on suggestions by Hammarström (2016).

1. Sample humidity measures from all languages such that no language overlaps in language family nor area. Call this group *R*.
2. From all languages with complex tone, sample humidity measures from languages such that no language overlaps in language family nor area. Call this group *C*.
3. From all languages with non-complex tone, sample humidity measures from languages such that no language overlaps in language family nor area. Call this group *N*.
4. Make sure the groups are the same size (randomly throw out items from the larger groups).
5. For a given percentile *x*, work out the humidity percentile for each group (the humidity below which *x*% of the data lie).

TABLE 2 | Results from the percentile test (test 3) using data from the ANU phonotactics database (from Everett et al., 2016b) and from PHOIBLE (this paper).

Source	Percentile			
	15th	**25th**	**50th**	**75th**
ANU Phonotactics database	0.977	0.818	0.0368	0.113
PHOIBLE	0.855	0.906	0.866	0.715

Numbers represent proportion of samples in which the size of the difference in humidity quantiles between complex and non-complex tone languages was bigger than between complex languages and a random sample of languages.
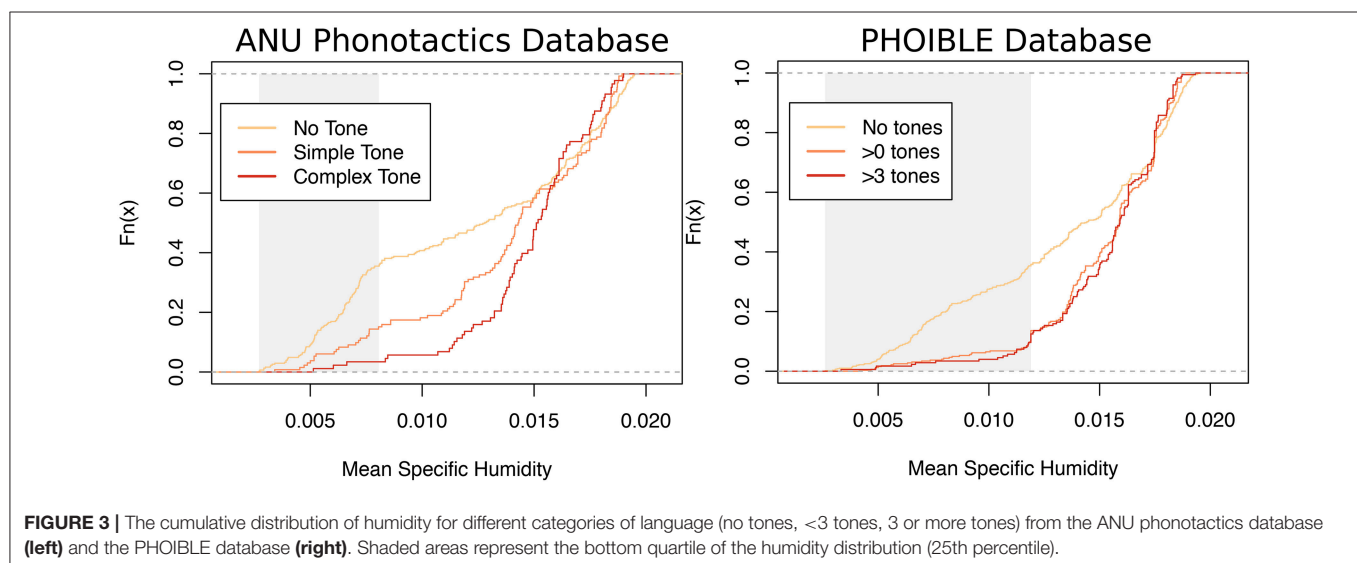

FIGURE 3 | The cumulative distribution of humidity for different categories of language (no tones, <3 tones, 3 or more tones) from the ANU phonotactics database **(left)** and the PHOIBLE database **(right)**. Shaded areas represent the bottom quartile of the humidity distribution (25th percentile).

6. Return the difference in percentiles: $C - N$ and $C - R$.
7. Repeat steps 1–6 for 5,000 times.
8. Test the proportion of times $(C - N) > (C - R)$.

Table 2 shows the results. In the original study, there were two crucial results: first, that the difference between humidity percentiles was larger than the baseline in more than 95% of samples for the lowest humidity percentile (15th percentile). Secondly, that this value was much lower for higher percentiles (50th and 75th). Neither of these results holds when using the PHOIBLE data.

5.2. Study 2: Using Continuous Measures of Tone

The original tests split the data into complex and non-complex tone languages. A continuous variable allows an analysis to predict the number of tones directly. **Figure 4** shows the distribution of humidity by continuous tones.

Mixed effects models in the *lme4* package (Bates et al., 2011) for R (R Core Team, 2011) were used to predict the raw number of tones (see Supplementary Material 1). A poisson distribution was used to capture the discrete and skewed nature of the data. The model had random intercepts for language family and geographic area and random slopes for the effect of humidity for both family and area. Including humidity as a fixed effect in the model did not significantly improve the fit of the model ($\beta = 0.19$, log likelihood difference = 0.22, df = 1 , $\chi^2 = 0.45$, $p = 0.50$). The same model was tested using the *MCMCglmm* package for R (Hadfield, 2010), which converges on estimates using a Bayesian Monte Carlo Markov chain. This approach is better able to detect multiple conflicting solutions to the fixed effect estimates. The results broadly agreed with those from the *lme4* model ($\beta = 0.20 \; [-0.04, 0.44]$, $p = 0.11$).

We can dig deeper into the model to try to understand why this relationship is not significant. **Table 3** shows how the estimate of the coefficient for humidity changes when removing particular parts of the random effect structure. The estimate is similar when removing the random slope for humidity by language family, suggesting that the effect of humidity does not differ much between families. On the other hand, the estimate is more significant when leaving out the either the intercept or slope by geographic area. Rather than arguing about which result is more valid, we should instead see these differences as suggesting something about the structure of the data. In this case, it suggests that the relationship between tone and humidity is not robust to controls for historical relationships, and in particular confounded by areal effects. This would fit with criticisms which suggest that borrowing is an important confound (Collins, 2016; Winter and Wedel, 2016). In robustness terms, the result is not structurally robust: the correlation does not survive controlling for the confound of contact.

5.2.1. Measurement Robustness for Tone
The results differ considerably in the alternative dataset, mainly because of measurement disagreements. The two sources overlap on 667 languages (Glottolog codes). The correlation in number of tones is only moderate (Cohen's weighted $\kappa = 0.61$, $r = 0.62$,

FIGURE 4 | Mean specific humidity by the number of tones in a language from the PHOIBLE database.
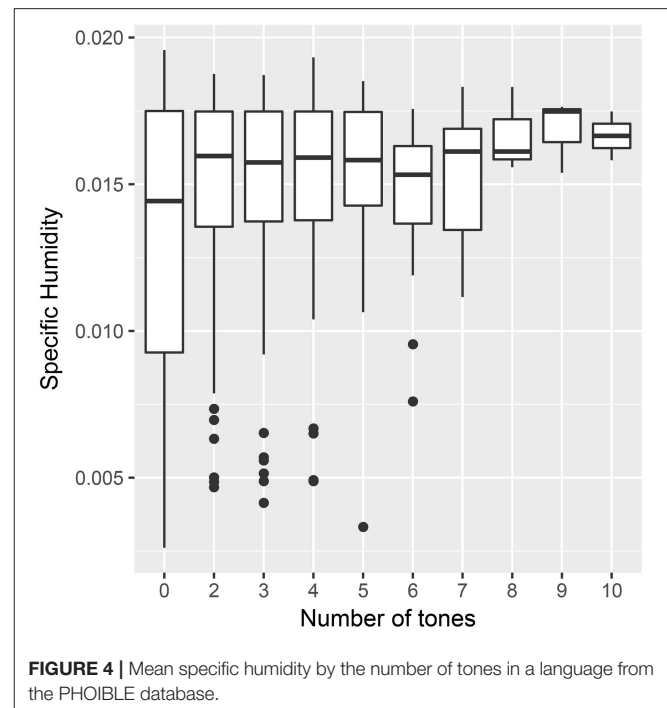
TABLE 3 | How the estimate for the coefficient for the effect of humidity on the number of tones changes when altering the random effects structure (*lme4* model with PHOIBLE data).

Model	Estimate	Std. Error	z-value	Pr(> |z|)
Full model	0.19	0.28	0.68	0.495
No family intercept	0.75	0.41	1.83	0.067
No family slope	0.19	0.28	0.68	0.495
No area intercept	0.52	0.19	2.70	0.007
No area slope	0.16	0.05	3.43	0.001

see Supplementary Material 2). When categorising languages into those having tones and those having no tones, the databases agree 82% of the time (Cohen's $\kappa = 0.64$, "moderate" agreement according to Landis and Koch, 1977, similar results comparing two or fewer tones to three or more). On average, the ANU database predicts a greater number of tones than in PHOIBLE.

PHOIBLE has many sources, but few languages are coded in more than one, making measurement robustness difficult to assess. Where it is possible to measure agreement between these sources, in one case it is very low (AA vs. GM: Cohen's weighted $\kappa = 0.08$, $r = 0.05$, $n = 36$) and in another it is very high (SPA vs. UZ: Cohen's weighted $\kappa = 0.95$, $r = 0.95$, $n = 26$). We can compare this to the agreement between vowels, which sits between these two extremes (AA vs. GM: Cohen's weighted $\kappa = 0.56$, $r = 0.52$, $n = 36$; SPA vs. UZ: Cohen's weighted $\kappa = 0.53$, $r = 0.53$, $n = 26$). The differences might be due to differences in methodological approaches, theoretical background or errors in data entry or coding of languages, but are most likely to be due to the inherently difficult nature of quantifying a phonetic system (e.g., dealing with length,

nasalisation, diphthongs, see Maddieson, 2013; Moran, 2016 or a specific case e.g., Montes Rodriguez, 2004, p. 111). It is worth noting that PHOIBLE and many other recent databases provide features for continuous and centralized updating and refining of the data by members of the research community through github (e.g., see PHOIBLE's issue tracker: https://github.com/phoible/dev/issues), so these problems will hopefully decrease as time goes on.

The low measurement robustness for the number of tones is concerning, especially since another independent source is hard to produce. However, it is only problematic for the statistical inquiries of this paper if the differences are *biased according to humidity*. This was tested by trying to predict the difference between the two estimates using a mixed effects model with random intercepts for language family and geographic area. If the differences are completely unbiased, then the random effects should not account for a significant proportion of the variance. This was the case (for family $p = 0.09$; for area $p = 1$, see Supplementary Material 2). If estimates differ in particular humidity conditions, then a fixed effect of humidity should improve the fit of the model. This was not the case ($p = 0.42$). Therefore, the differences between the sources are not biased with regards to language, area, or humidity. Given the results above, however, it is still clear that the different sources lead to different results regarding the link between tone and humidity. This is another reason to take a maximum robustness approach.

6. STUDY 3: HUMIDITY AND PROPORTION OF VOWELS

This section tests the robustness of the link between humidity and vowels. Everett (2017) looked at the proportion of vowels vs. consonants in basic wordlists from the ASJP database (Wichmann et al., 2013). This was used as a measure of the relative frequency of vowels and consonants during speech, and it was shown that this correlated with the specific humidity of the areas in which the languages were spoken. In this section, a different approach is taken: to try and predict the proportion of vowels in a language's phoneme inventory by humidity. The relative frequency of phonemes is a more ideal measure (indeed, Everett argues that phoneme inventories are misleading since it is habitual use that is more important). However, the basic word lists used in the study are relatively restricted, and the theory could extend to affecting the number of distinctions in the phoneme system. In any case, the study here is an illustrative example of expanding the range of analyses.

Data on phoneme inventories was taken from the PHOIBLE database (Moran et al., 2014). As above, a linear mixed effects model (in package *lme4*) was used to predict the ratio of vowels to consonants within a language's inventory by humidity (see Supplementary Material 3). Since the vowel ratio may be affected by the total phoneme inventory size, it was added as a fixed effect. Adding humidity as an additional fixed effect significantly improved the fit of the model ($\beta = 0.17$, log likelihood difference $= 3.9$, df $= 1$, $\chi^2 = 7.77$, $p = 0.005$), indicating that higher humidity was associated with a greater proportion of

vowels. Interestingly, there was a significant interaction between humidity and inventory size ($\beta = 0.10$, log likelihood difference $= 9.3$, df $= 1$, $\chi^2 = 18.57$, $p < 0.001$), such that the correlation between proportion of vowels and humidity is stronger for languages with larger inventories.

We can test the estimation robustness of the finding. The estimates did not change much when using 6 alternative optimizers, providing at least some estimation robustness (see Supplementary Material 2). The coefficient estimates are also very similar when using the *MCMCglmm* package. There was a significant effect of humidity ($\beta = 0.16 [0.05, 0.28]$, $p = 0.004$) and a significant interaction between humidity and inventory size is also significant ($\beta = 0.10 [0.05, 0.17]$, $p < 0.001$). In this model, there is also a significant main effect of inventory size ($0.09 [0.03, 0.16]$, $p = 0.003$).

The effect size is very small. The model predicts that when comparing the language in the driest environment to the language in the most humid environment that the proportion of vowels should increase from about 25% to about 35%. In a language with an average phoneme inventory size, that's a difference of about 3 vowels.

6.1. Measurement Robustness for the Relative Frequency of Vowels

Everett (2017) used the proportion of vowels in the basic vocabulary lists of the ASJP database as a proxy for the relative frequency of vowels in general speech. The ASJP database contains a large number of languages (over 7,000 varieties), but a small number of concepts (most languages have 40, some have 100). Are the estimates robust when increasing the number of concepts? An alternative could be the database of lexical items compiled by Slonimska and Roberts (2017) from sources in Haspelmath and Tadmor (2009); Key and Comrie (2015) and Borin et al. (2016). It has 999 concepts in 226 languages (about 10 times more concepts but 10 times fewer languages compared to the ASJP). The correlation between the proportion of vowels in the ASJP and the alternative dataset is reasonably good ($r = 0.65$). However, the magnitude of the differences between the two measures varies significantly between language families, geographic areas and (weakly) according to humidity. That is, the ASJP estimates of vowel frequency are biased (unlike the estimates for tones). It is not clear what the next course of action here is. The alternative dataset does not have enough languages to reliably detect the original correlation, but a larger database with many more concepts is unlikely to appear soon (though see the upcoming Lexibank database, http://glottobank.org/#lexibank). In this case, it may be best to turn to other measures. For example, phonetic measures may be more reliable because there are objective, repeatable methods.

6.2. Using Phonetic Measurements
It is also possible to use phonetic measurements to test the hypothesis (see also Maddieson, under review in this volume). Indeed, it might be easier to automatically extract and replicate phonetic measurements (Ennever et al., 2017). The hypothesis predicts that speakers in drier climates would use a more

FIGURE 5 | An example of the range of vowel systems from three Austronesian languages. The humidity in which the languages are spoken are shown in parentheses next to the language name in the legend.

FIGURE 6 | The relationship between the range of frequencies a vowel space covers and the specific humidity in which it is used. The regression line is drawn according to the mixed effects model estimates.

restricted range of frequencies in vowels. Becker-Kristal (2010) provides a database of phonetic measurements of vowels for many languages based on a meta-analysis. Following Weirich and Simpson (2014), for each language, the F1 and F2 measures of all vowels in a language were taken, then the area of the convex hull of the points was calculated. This represents the range that a vowel system takes up (see **Figure 5**).

The area of 219 vowel systems was calculated. A mixed effects model was used to predict vowel area with random intercepts for language family and geographic area. Even when controlling for the number of vowels in a system, adding specific humidity as a fixed effect significantly improved the fit of the model (β = 0.16, log likelihood difference = 2.5, χ^2 = 5.01, df = 1, p = 0.025, see Supplementary Material 4). The effect size was small (see **Figure 6**), and there was not enough data to include random slopes for specific humidity, so this result is probably not robust. The main point here is that it is possible to use more fine-grained measures from alternative data sources to test large-scale statistical claims and contribute to the methodological robustness of the result.

7. METHODOLOGICAL ROBUSTNESS FOR THE EFFECTS OF HUMIDITY ON LANGUAGE

The studies above looked at the correlation between humidity and tone or between humidity and vowel use, but these are just the end-points of a more detailed causal chain drawn up in section 2. The maximum robustness approach suggests that each of these links can be addressed with different methods and data. The following subsections suggest some ideas for how this might be done. The point here is not to test each link, but to demonstrate that methods from many different areas of linguistics can be brought to bear on them.
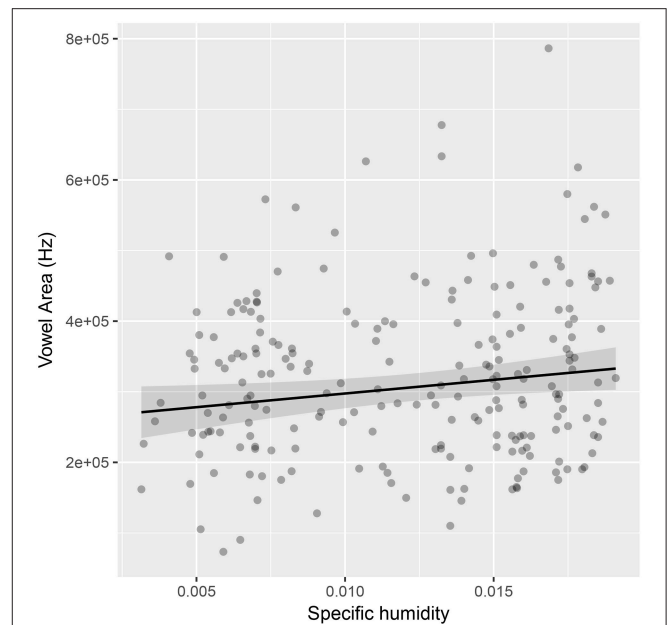
7.1. Iterated Learning

One could imagine an iterated learning study to address the link between humidity, desiccation, production, and the loss of tones. A participant would learn an artificial language where the labels were auditory words with distinctions in tone and non-tone segments. They would be asked to reproduce the correct labels, and their productions would be given as the input language to a new participant. This process would be repeated to create a chain of generations in which the labels would change gradually. Chains would be run in specially controlled rooms with two conditions: dry air and humid air. The prediction is that distinctions in tone would survive in the humid condition, but be more likely to disappear from the dry condition (**Figure 7**). Alternative conditions could be tested such as having the participants communicate with a partner using the language, to test the role of miscommunication over and above production error.

7.2. Historical Case Study

We can use the cross-linguistic data to find promising case-studies for more detailed historical linguistic work. Data on tones and humidity were used together with the historical tree suggested by Glottolog to infer the likely ancestral states in the Atlantic Congo family. The most interesting section is the Narrow Bantu clade, where two sub-groups (Eastern and Central-Western Bantu) enter drier climates. This clade is also known to have generally fewer tonal contrasts and simpler tone systems (e.g., only high vs. low, Güldemann, 2011, p. 115). Crucially, some languages within the sub-groups re-enter humid zones, for instance the languages which border Rwanda and Tanzania. **Figure 8** shows the tone and humidity measures for some of these languages, linked by the phylogenetic tree inferred by Currie et al.

FIGURE 7 | A hypothetical iterated learning experiment to investigate the link between humidity, desiccation, production, and the loss of tone distinctions. The desiccation hypothesis predicts that languages transmitted in a humid condition would develop a (possibly compositional) system that preserved distinctions in tone, while languages in a dry condition would develop a compositional system that relied on non-tonal distinctions.

FIGURE 8 | Phylomorpho space plot of Northeast Savannah Bantu languages. Red dots are the actual attested languages, and these are joined with lines to black dots representing their ancestors reflecting the consensus phylogenetic tree from Currie et al. (2013), where the position of ancestor points regarding both number of tonal contrasts and mean humidity are calculated via continuous ancestral state reconstruction. Some of the branches of the tree are altered for clarity, line lengths are not meaningful.

(2013) (group J, also closely related in Glottolog). The trend is the predicted one—fewer tones occur in drier climates. For example, Jita and Gwere split up, one heading into a humid region, and

one heading into a dry region, with the predicted change in tones. The points are also not very clustered by historical relatedness— Yaka and Gwere are similar in tones and in humidity, but actually not close on the phylogenetic tree. We suggest that this group of languages provides an excellent candidate for a more detailed case study of historical changes to tone.

7.3. Corpus Study: Production

Croft's approach to language change is that the locus of change is individual utterances (Croft, 2000, perhaps improved by Tamariz's focus on the reproductions themselves separate from their meanings, Tamariz, 2017a). That is, selection operates on variation in productions turn-by-turn, and not just in generations. The hypothesis linking tone and humidity would predict that, in order for change to happen, there would have to be underlying variation on which selection could operate, and that this should be visible within speech communities. For example, do users of tone languages show variation in the proportion of different tone types that they use, and do they vary systematically with humidity. That is, a language offers a speaker multiple choices about how to express a meaning. These options may differ in the demands they make on vocal fold control, and so may be more or less difficult to produce in different locations or at different times of the year. This could be tested in two ways. First, do speakers of a language such as Cantonese produce tones differently or use a different proportion of tones in the humid parts of China compared to the colder, more arid parts? Secondly, do speakers' productions differ according to the seasonal change in humidity?

Databases with geolocations and dates of production are not common, but the CHILDES database does include the dates of recordings. Data from 189 recordings of Cantonese were obtained from CHILDES (Fletcher et al., 1996, 2000; Lee et al., 1996; Weizman and Fletcher, 2000) and productions by children were removed. The number of each type of tone was calculated for each recording, then linked to the month that the recording was taken. The following prediction was made: contour tones would require more precise control of vocal fold vibration, so would be avoided during the drier months. Mean monthly specific humidity was collected around Hong Kong for the years spanning the corpus collection (Kalnay et al., 1996). Mixed effects modeling with random intercepts for source corpus was used to test whether the contribution of humidity significantly predicted the proportion of use of contour tones (see Supplementary Material 5). There was no significant effect ($\chi^2 = 0.28$, $p = 0.59$), and in fact the use of contour tones does not vary over the year.

7.4. Corpus Study: Miscommunication

One link in the causal chain relating to interaction predicts that more complex tones are more difficult to produce and therefore more likely to cause problems of understanding. For example, Mandarin has 4 tones, with the 3rd tone being a contour tone with a wide range, possibly requiring more precise control of the vocal folds (though often reduced in speech). One might predict that turns in conversation including 3rd tones would be more susceptible to errors in production and perception, and therefore more likely to elicit repair from interlocutors.

A corpus of repair sequences in Mandarin conversation was obtained from Dingemanse et al. (2015), collected and transcribed by Kobin Kendrick. The proportion of each type of tone was counted in trouble sources (turns followed by open other-initiated repair, indicating a problem of hearing or understanding) and compared to the baseline proportion of each tone type in a wider corpus (Wan and Jaeger, 1998). The 3rd tone was significantly more likely in trouble sources (using a χ^2 test on the tone counts in **Table 4**, $\chi^2 = 9.89$, df = 3, $p = 0.02$). This is in line with the hypothesis, but much more could be done to check the robustness of this claim. In particular, it should be possible to look at tone type usage in the actual source of the problem for restricted repair initiations, rather than the whole prior turn in general. Again, the point here is that specific data and analyses can be brought to bear on particular links in the causal chain.

TABLE 4 | Counts (and percentages) of different tonemes in Mandarin in general (baseline from Wan and Jaeger, 1998), and in conversational turns that lead to open repair (from Dingemanse et al., 2015).

Tone	1	2	3	4
Baseline	46 (23%)	44 (22%)	43 (21%)	71 (34%)
Trouble sources	56 (24%)	34 (15%)	75 (33%)	64 (28%)
Percentage difference	+2	−7	+12	−7

7.5. Individual Speech

It is also possible to look at whether individual speakers shift the way they speak due to the climate, though a large sample of recordings would be needed. The ideal database would be a few minutes recording every day over the course of several years. This kind of database is rare, but they do exist. For example, Larry King has recorded a show almost every day for over a decade. CNN provides transcripts of these shows from 2000 to 2011 (http://transcripts.cnn.com/TRANSCRIPTS/lkl.html), a total of around 3,500 recordings. These transcripts were downloaded and King's turns were extracted. Personal names and locations were removed and the text was transcribed to a phonological representation using the CMU pronouncing dictionary (Weide, 2005, on average 95% of tokens were transcribable, 91% of types). Daily specific humidity estimates are also available for each show's recording date and location (assuming Los Angeles, Kalnay et al., 1996). We can then test whether King uses a smaller proportion of vowels compared to consonants during drier weather.

King's vowel ratio is very stable. It ranges between 0.63 and 0.70 (more consonants than vowels, sd = 0.008, see **Figure 9**). A general additive model was used to test the relationship between the vowel ratio and humidity. There was a significant relationship [$F_{(2.85, 3.61)} = 4.95$, $p = 0.001$, see Supplementary Material 6], but higher humidity was associated with the use of proportionally fewer vowels, going against the prediction.

There are, of course, many problems with this study. The recordings are mostly done in air-conditioned studios (in fact, the results are consistent with drier air due to air conditioning during summer), and occasionally broadcast from other cities. It
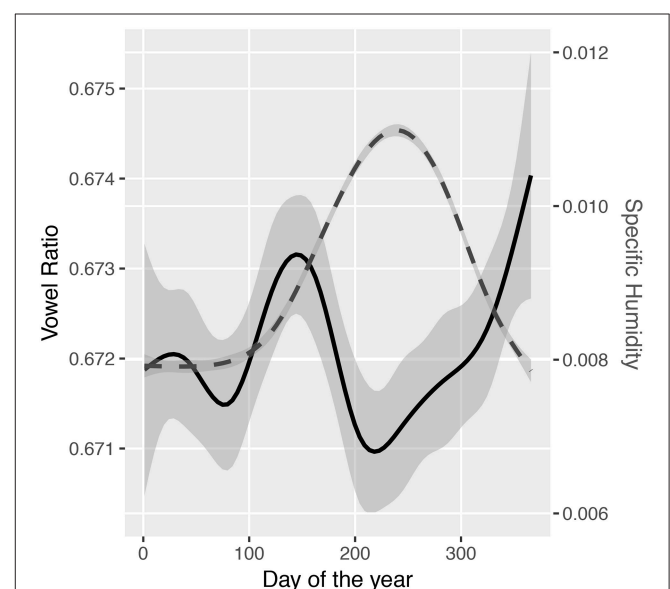

FIGURE 9 | A graph showing the average proportion of vowels used by Larry King (solid line, left axis) and the average specific humidity of Los Angeles (dashed line, right axis) over days of the year. Lines are smoothed using a general additive model with 95% confidence intervals.

TABLE 5 | A summary of the previous and current results relating humidity to tone and vowels.

Variable	Test	Source		
		ANU	**PHOIBLE**	
Tones	Percentile test	Yes	No	
	LME	No	No	
	MCMC	?	No	
	Measurement robustness	Medium	-	
		ANU	**PHOIBLE**	**ASJP**
Vowels: lexicon	LME	–	–	Yes
	Measurement robustness	–	–	Low
Vowels: inventory	LME	?	Yes	
	MCMC	?	Yes	
	Alternative methods	**Various sources**		
Tones	Iterated learning	?		
	Historical case study	?		
	Corpus study: production	No		
	Corpus study: miscommunication	Yes		
Vowels	Corpus study: individual speech	No		
	Phonetic measurements	Yes		

Cells represent whether there was a significant relationship.

is also probable that seasonal topics contribute to the variation in vowel ratio. However, the point here is that this question is at least empirically approachable.

8. SUMMARY OF THE STATE OF THE HYPOTHESIS LINKING HUMIDITY AND LANGUAGE

The maximum robustness method identified several weak parts in the causal chain relating humidity and tone. These are mostly regarding whether the effect on production is large enough to affect linguistic systems in the long term. **Table 5** summarizes the robustness analyses above. The original relationship between tone and humidity was not replicated in an alternative database. Consistent with criticisms that the borrowing is a confounding factor, the strength of the relationship is mostly accounted for by differences in particular geographic areas. The alternative methods had mixed results, but at least demonstrated that the hypothesis can be approached from many different angles.

The relationship between humidity and vowels is more robust, though the effect size is small. There is positive evidence from relative frequency of vowels in the basic vocabulary, relative frequency in phonemic inventories (with some estimation robustness) and phonetic measures. However, the measurement robustness of the frequency of vowels in the lexicon may be low. It should be noted that Everett (2017) suggests that looking at phoneme inventories is not a valid test of the hypothesis, since usage

frequency is more important (and see also Maddieson, under review).

In summary, the effect of humidity on language is an intriguing frontier in accounts of linguistic adaptation. However, the basic correlation between humidity and tone is not robust. It is unlikely that new independent global data on tones will become available soon, so the best next step for this line of research is to diversify the methodological approaches and reach toward experimental and diachronic studies. Having said this, there are other lines of research that are better grounded in linguistic theory and are more likely to represent substantial effect sizes. What is needed is a more detailed mechanistic model of how production is affected by the ecological conditions, and how these effects put a pressure for whole linguistic systems to change.

9. CONCLUSION

This paper discussed the approach to studying how languages adapt to external selective pressures by looking at patterns in large-scale, cross-cultural databases. It advocated a maximum robustness approach, which is empirical, causal, incremental, and robust. Each of these aspects feeds into the others to provide an increasingly clear evidence for a particular hypothesis. This method encourages researchers to move beyond large-scale statistical analyses and into more diverse methods and toward more controlled, causal accounts by breaking a hypothesis down into smaller causal links, then addressing each link with the most appropriate method and data. This approach was illustrated

with examples based on the research into the relationship between humidity and tone. Although large-scale statistical approaches can provide the motivation to pursue a hypothesis, and ultimately a demonstration that the whole causal chain produces significant adaptation, it is unlikely that they will be able to provide convincing evidence on their own. Some examples above suggested ways in which other approaches could help, including laboratory phonetics, iterated artificial language learning experiments, corpus studies, and historical case studies.

Engaging with this range of methods and disciplines is difficult for just one researcher. It is more likely that this approach will be successful in the context of collaboration between specialists from different areas. Open access to data and statistical modeling code will also be a key to making these projects viable. It also means that large-scale interdisciplinary grants will become more important, as well as recognising the different types of contribution that authors make to a paper (theoretical, experimental, data collection, statistical, organizational etc.).

The maximum robustness approach advocates doing many analyses with as many sources of data as possible. However, it is important to note that it definitely does not advocate practices such as p-hacking, cherry-picking or presenting *post-hoc* descriptions as a priori hypotheses. Studies with good structural robustness will run many analyses, then *report them all*. The aim is to provide many alternative viewpoints, not to discover the most convenient statistic. Recent advances in meta-analysis methods are providing ways of navigating the range of results (Lewis et al., 2015). The approach is also different from the slow science movement (Lutz, 2012). While both emphasize careful and detailed consideration of theories and methods, the maximum robustness approach is more open and pragmatic. In linguistics, different phenomena are deeply interrelated and new data is a scarce resource. It is better that the analyses are published and discussed so that they can help other research, even if the conditions are not perfect. Indeed, according to the maximum robustness approach, foregrounding the potential flaws and limits of an analysis is useful. On the other hand,

while swift publishing is encouraged, this approach does require a more cautious approach to acceptance of theories. Overstated results and hasty adoption may be hard to overturn. The study above linking language to economic behavior (Chen, 2013) was quickly taken onboard by economists and the data has been reused (Santacreu-Vasut et al., 2014; Hicks et al., 2015; Pérez and Tavits, 2017), even though the original findings are now in doubt (Roberts et al., 2015, though see Mavisakalyan and Weber, 2017). The solution may mean that researchers need to spend more time refining the communication of their research, especially to non-specialist audiences. It may also require more moderate language to describe the significance of studies and a full account of its flaws. However, like language, scientific practice adapts to its wider ecology, and the current climate promotes hyperbolic discoveries over statistical grumblings. Wider changes may be necessary to support the maximum robustness method.

AUTHOR CONTRIBUTIONS

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

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00166/full#supplementary-material

REFERENCES

Allison, D., Hyman, L. M., and Mortensen, D. (2006). *Xtone: Cross-Linguistic Tonal Database*. Berkeley, CA: University of California. Available online at: http://xtone.linguistics.berkeley.edu

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Barth, D., and Kapatsinski, V. (2017). A multimodel inference approach to categorical variant choice: construction, priming and frequency effects on the choice between full and contracted forms of am, are and is. *Corpus Linguist. Linguist. Theory* 13, 1–58. doi: 10.1515/cllt-2014-0022

Bass, H., Sutherland, L., Piercy, J., and Evans, L. (1984). "Absorption of sound by the atmosphere," in *Physical Acoustics: Principles and Methods, Vol. 17 (A85-28596 12-71)*, eds R. N. Thurston and W. P. Mason (Orlando, FL: Academic Press, Inc.), 145–232.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., and Bolker, B. (2011). *lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version 0.999375-42*. Available Online at: http://CRAN.R-project.org/package=lme4

Becker-Kristal, R. (2010). *Acoustic Typology of Vowel Inventories and Dispersion Theory: Insights from a Large Cross-Linguistic Corpus*. Los Angeles, CA: University of California.

Bero, L. A., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., and Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The cochrane effective practice and organization of care review group. *BMJ* 317, 465–468.

Bickel, B. (2010). "Capturing particulars and universals in clause linkage," in *Clause-Hierarchy and Clause-Linking: The Syntax and Pragmatics Interface*, ed I. Bril (Amsterdam: John Benjamins), 51–101.

Bickel, B. (2011). "Multivariate typology and field linguistics: a case study on detransitivization in kiranti (sino-tibetan)," in *Proceedings of the Conference on Language Documentation and Linguistic Theory* (London: SOAS), 3–13.

Bickel, B. (2015). "Distributional typology: statistical inquiries into the dynamics of linguistic diversity," in *Oxford Handbook of Linguistic Analysis*, eds B. Heine and H. Narrog (Oxford: Oxford University Press), 901–923.

Bickel, B., Hildebrandt, K., and Schiering, R. (2009). "The distribution of phonological word domains: a probabilistic typology," in *Phonological Domains: Universals and Deviations*, eds J. Grijzenhout and B. Kabak (Berlin: Walter de Gruyter), 47–75.

Bickel, B., and Nichols, J. (2006). "Oceania, the pacific rim, and the theory of linguistic areas," in *Annual Meeting of the Berkeley Linguistics Society*, Vol. 32 (Berkeley, CA), 3–15.

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: evidence from case marking. *PLoS ONE* 10:e0132819. doi: 10.1371/journal.pone.0132819

Bickel, B., Zakharko, T., Bierkandt, L., and Witzlack-Makarevich, A. (2014). Semantic role clustering: an empirical assessment of semantic role types in non-default case assignment. *Stud. Lang.* 38, 485–511. doi: 10.1075/sl.38.3.03bic

Blasi, D. E., and Roberts, S. G. (2017). "Beyond binary dependencies in language structure," in *Dependencies in Language: On the Causal Ontology of Linguistic Systems*, ed N. J. Enfield (Leipzig: Language Sciences Press), 117.

Borin, L., Forsberg, M., Hammarstedt, M., Rosen, D., Schäfer, R., and Schumacher, A. (2016). "Sparv: Sprøakbanken's corpus annotation pipeline infrastructure," in *The Sixth Swedish Language Technology Conference (SLTC), Umeå University* (Umeå), 17–18.

Chen, M. K. (2013). The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets. *Am. Econ. Rev.* 103, 690–731. doi: 10.1257/aer.103.2.690

Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* 70:213.

Collins, J. (2016). Commentary: the role of language contact in creating correlations between humidity and tone. *J. Lang. Evol.* 1, 46–52. doi: 10.1093/jole/lzv012

Coupé, C. (2017). "Ecological constraints on linguistic diversity," in *Proceedings of the 50th Societas Linguistica Europea Conference* (Zürich).

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach.* Harlow: Longman.

Currie, T. E., Meade, A., Guillon, M., and Mace, R. (2013). "Cultural phylogeography of the bantu languages of sub-saharan africa," in *Proceedings of the Royal Society B*, Vol. 280 (The Royal Society).

de Boer, B. (2016). Commentary: is the effect of desiccation large enough? *J. Lang. Evol.* 1, 55–57. doi: 10.1093/jole/lzv008

de Rivarol, A. (1784). *De l'universalité de la Langue Française; Discours qui a Remporté le Prix à l'Académie de Berlin.* Berlin: Bailly et Dessenne.

Dediu, D., Janssen, R., and Moisik, S. R. (2017). Language is not isolated from its wider environment: vocal tract influences on the evolution of speech and language. *Lang. Commun.* 54, 9–20. doi: 10.1016/j.langcom.2016.10.002

Dediu, D., and Ladd, D. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc. Natl. Acad. Sci. U.S.A.* 104:10944. doi: 10.1073/pnas.0610848104

Dediu, D., and Moisik, S. R. (2016). "Anatomical biasing of click learning and production: an MRI and 3d palate imaging study," in *11th International Conference on the Evolution of Language (EvoLang XI)* (New Orleans, LA).

Dingemanse, M., and Enfield, N. J. (2015). Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguist.* 1, 98–118. doi: 10.2478/opli-2014-0007

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal principles in the repair of communication problems. *PLoS ONE* 10:e0136100. doi: 10.1371/journal.pone.0136100

Donohue, M. (2016). Commentary: culture mediates the effects of humidity on language. *J. Lang. Evol.* 1, 57–60. doi: 10.1093/jole/lzv009

Donohue, M., Hetherington, R., Mc Elvenny, J., and Dawson, V. (2013). *World Phonotactics Database.* Canberra, ACT: Australian National University. Available online at: http://phonotactics.anu.edu.au

Dorie, V. (2011). *blme: Bayesian Linear Mixed-Effects Models.* Available online at: http://CRAN.R-project.org/package=blme

Easton, C. L., and Stebbins, T. N., (2015). *On Becoming an Object of Study*, Amsterdam: John Benjamins Publishing Company.

Ember, C. R. (2016a). Commentary: considering language as expressive culture. *J. Lang. Evol.* 1, 60–61. doi: 10.1093/jole/lzv005

Ember, C. R. (2016b). *Explaining Human Culture.* New Haven, CT: Human Relations Area Files. Available online at: http://hraf.yale.edu/ehc

Ennever, T., Meakins, F., and Round, E. R. (2017). A replicable acoustic measure of lenition and the nature of variability in gurindji stops. *Lab. Phonol.* 8:32. doi: 10.5334/labphon.18

Everett, C. (2017). Languages in drier climates use fewer vowels. *Front. Psychol.* 8:1285. doi: 10.3389/fpsyg.2017.01285

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1322–1327. doi: 10.1073/pnas.1417413112

Everett, C., Blasi, D. E., and Roberts, S. G. (2016a). Language evolution and climate: the case of desiccation and tone. *J. Lang. Evol.* 1, 33–46. doi: 10.1093/jole/lzv004

Everett, C., Blasi, D. E., and Roberts, S. G. (2016b). Response: climate and language: has the discourse shifted? *J. Lang. Evol* 1, 83–87. doi: 10.1093/jole/lzv013

Ey, E., and Fischer, J. (2009). The 'acoustic adaptation hypothesis': review of the evidence from birds, anurans and mammals. *Bioacoustics* 19, 21–48. doi: 10.1080/09524622.2009.9753613

Fletcher, P., Lee, T., Leung, C., and Stokes, S. (1996). *Milestones in the Learning of Spoken Cantonese by Pre-School Children.* Hong Kong: Language Fund.

Fletcher, P., Leung, S. C. S., Stokes, S., and Weizman, Z. (2000). *Cantonese Pre-school Language Development: A Guide.* Report of a project funded by the Language Fund, 1996–1999, Hong Kong. Available online at: http://repository.vtc.edu.hk/thei-fac-gen-ed-sp/223

Fought, J. G., Munroe, R. L., Fought, C. R., and Good, E. M. (2004). Sonority and climate in a world sample of languages: findings and prospects. *Cross Cult. Res.* 38, 27–51. doi: 10.1177/1069397103259439

Gavin, M. C., Botero, C. A., Bowern, C., Colwell, R. K., Dunn, M., Dunn, R. R., et al. (2013). Toward a mechanistic understanding of linguistic diversity. *BioScience* 63, 524–535. doi: 10.1525/bio.2013.63.7.6

Gavin, M. C., Rangel, T. F., Bowern, C., Colwell, R. K., Kirby, K. R., Botero, C. A., et al. (2017). Process-based modelling shows how climate and demography shape language diversity. *Global Ecol. Biogeogr.* 26, 584–591. doi: 10.1111/geb.12563

Güldemann, T. (2011). "Proto-bantu and proto-niger-congo: Macro-areal typology and linguistic reconstruction," in *Geographical Typology and Linguistic Areas, with Special Reference to Africa*, eds O. Hieda, C. König, and H. Nakagawa (Amsterdam: John Benjamins Publishing), 109–141.

Gussenhoven, C. (2016). Commentary: tonal complexity in non-tonal languages. *J. Lang. Evol.* 1, 62–64. doi: 10.1093/jole/lzv016

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the mcmcglmm R package. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i02

Hammarstrom, H. (2010). A full-scale test of the language farming dispersal hypothesis. *Diachronica* 27, 197–213. doi: 10.1075/bct.46.02ham

Hammarström, H. (2016). Commentary: there is no demonstrable effect of desiccation [commentary on "language evolution and climate: the case of desiccation and tone"]. *J. Lang. Evol.* 1, 65–69. doi: 10.1093/jole/lzv015

Hammarström, H., Forkel, R., and Haspelmath, M. (2017). *Glottolog 3.0.* Max Planck Institute for the Science of Human History.

Haspelmath, M., and Tadmor, U. (2009). *Loanwords in the World's Languages: a Comparative Handbook.* Leipzig: Walter de Gruyter.

Hemler, R. J., Wieneke, G. H., and Dejonckere, P. H. (1997). The effect of relative humidity of inhaled air onacoustic parameters of voice in normal subjects. *J. Voice* 11, 295–300.

Hicks, D. L., Santacreu-Vasut, E., and Shoham, A. (2015). Does mother tongue make for women's work? linguistics, household labor, and gender identity. *J. Econ. Behav. Organ.* 110, 19–44. doi: 10.1016/j.jebo.2014.11.010

Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen, K., and Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the uralic languages. *J. Evol. Biol.* 26, 1244–1253. doi: 10.1111/jeb.12107

Hoover, K. D., and Perez, S. J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bull. Econ. Stat.* 66, 765–798. doi: 10.1111/j.1468-0084.2004.101_1.x

Irvine, L., Roberts, S. G., and Kirby, S. (2013). "A robustness approach to theory building: a case study of language evolution," in *The 35th Annual Meeting of the Cognitive Science Society (CogSci 2013)* (Berlin), 2614–2619.

Janssen, R., Dediu, D., and Moisik, S. R. (2016). "Simple agents are able to replicate speech sounds using 3D vocal tract model," in *11th International Conference on the Evolution of Language (EvoLang XI)* (New Orleans, LA)

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The ncep/ncar 40-year reanalysis project. *Bull. Am. Meteor. Soc.* 77, 437–471.

Kandler, A., and Steele, J. (2008). Ecological models of language competition. *Biol. Theory* 3, 164–173. doi: 10.1162/biot.2008.3.2.164

Key, M. R., and Comrie, B., (eds.). (2015). *IDS*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Khan, K. S., Ter Riet, G., Glanville, J., Sowden, A. J., Kleijnen, J., et al. (2001). *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Arrying Out or Commissioning Reviews*. Number 4 (2n), NHS Centre for Reviews and Dissemination.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Ladd, D. R. (2016). Commentary: tone languages and laryngeal precision. *J. Lang. Evol.* 1, 70–72. doi: 10.1093/jole/lzv014

Ladd, D. R., Roberts, S. G., and Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Ann. Rev. Linguist.* 1, 221–241. doi: 10.1146/annurev-linguist-030514-124819

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Leamer, E. E. (1985). Sensitivity analyses would help. *Am. Econ. Rev.* 75, 308–313.

Lee, T., Wong, C., Leung, S., Man, P., Cheung, A., Szeto, K., et al. (1996). *The Development of Grammatical Competence in Cantonese-Speaking Children (Rgc Project No. Cuhk 2/91)*. Hong Kong: Report of RGC Earmarked Grant.

Levins, R. (1966). The strategy of model building in population biology. *Am. Sci.* 54, 421–431.

Lewis, M., Braginsky, M., Bergmann, C., Tsuji, S., Cristia, A., and Frank, M. (2015). "Metalab: a tool for power analysis and experimental planning in developmental research," in *Proceedings of 40th BUCLD* (Boston, MA).

Leydon, C., Sivasankar, M., Falciglia, D. L., Atkins, C., and Fisher, K. V. (2009). Vocal fold surface hydration: a review. *J. Voice* 23, 658–665. doi: 10.1016/j.jvoice.2008.03.010

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6:e1000100. doi: 10.1136/bmj. b2700

Lutz, J.-F. (2012). Slow science. *Nat. Chem.* 4, 588–589. doi: 10.1038/nchem.1415

Macklin-Cordes, J. L., and Round, E. R. (2015). "High-definition phonotactics reflect linguistic pasts," in *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics* (Tubingen: University of Tubingen).

Maddieson, I. (2013). "Vowel quality inventories," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: http://wals.info/chapter/2

Maddieson, I., and Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation hypothesis. *J. Acoust. Soc. Am.* 138, 1838–1838. doi: 10.1121/1.4933848

Mavisakalyan, A., and Weber, C. (2017). Linguistic structures and economic outcomes. *J. Econ. Surv.* doi: 10.1111/joes.12247. [Epub ahead of print].

Michaelowa, A. (2001). "The impact of short-term climate change on british and french agriculture and population in the first half of the 18th century," in *History and Climate*, eds P. Jones, A. Ogilvie, T. Davies, and K. Briffa (New York, NY: Springer), 201–217.

Moisik, S. R., and Dediu, D. (2017). Anatomical biasing and clicks: evidence from biomechanical modeling. *J. Lang. Evol.* 2, 37–51. doi: 10.1093/jole/lzx004

Montes Rodriguez, M. E. (2004). Fonología y dialectologia del tikuna. *Amerindia* 29–30, 97–116. Available online at: https://www.vjf.cnrs.fr/sedyl/amerindia/articles/pdf/A_29-30_04.pdf

Moran, S. (2012). *Phonetics Information Base and Lexicon*. Ph.D. thesis, University of Washington.

Moran, S. (2016). Commentary: issues of time, tone, roots and replicability. *J. Lang. Evol.* 1, 73–76. doi: 10.1093/jole/lzv011

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language* 88, 877–893. doi: 10.1353/lan.2012.0087

Moran, S., McCloy, D., and Wright, R., (eds.). (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Morgan, L. C. (1954). *The Effect of Temperature and Humidity on Hearing Acuity*. LSU Historical Dissertations and Theses. Available online at: https://digitalcommons.lsu.edu/gradschool_disstheses/8121

Morton, E. S. (1975). Ecological sources of selection on avian sounds. *Am. Natural.* 109, 17–34.

Munroe, R. L., Fought, J. G., and Macaulay, R. K. S. (2009). Warm climates and sonority classes: not simply more vowels and fewer consonants. *Cross Cult. Res.* 43, 123–133. doi: 10.1177/1069397109331485

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference, Vol. 47*. Cambridge: Cambridge University Press.

Pérez, E. O., and Tavits, M. (2017). Language shapes people's time perspective and support for future-oriented policies. *Am. J. Polit. Sci.* 61, 715–727. doi: 10.1111/ajps.12290

Pinkerton, J. (1789). *An Enquiry into the History of Scotland Preceding the Reign of Malcolm iii. or the Year 1056: Including the Authentic History of that Period: In Two Volumes*. Vol. 1. London: John Nichols.

Progovac, L., and Ratliff, M. (2016). Commentary: Beyond tone and climate: broadening the framework. *J. Lang. Evol.* 1, 77–79. doi: 10.1093/jole/lzv006

R Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://cran.r-project.org

Regier, T., Carstensen, A., and Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS ONE* 11:e0151138. doi: 10.1371/journal.pone.0151138

Roberts, S. G. (2018). "CHIELD: Causal hypotheses in evolutionary linguistics database," in *The Evolution of Language: Proceedings of the 12th International Conference*, eds C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, and T. Verhoef (Torun).

Roberts, S. G., Winters, J., and Chen, K. (2015). Future tense and economic decisions: controlling for cultural evolution. *PLoS ONE* 10:e0132145. doi: 10.1371/journal.pone.0132145

Sala-i Martin, X. X. (1997). I just ran two million regressions. *Am. Econ. Rev.* 87, 178–183.

Santacreu-Vasut, E., Shenkar, O., and Shoham, A. (2014). Linguistic gender marking and its international business ramifications. *J. Int. Bus. Stud.* 45, 1170–1178. doi: 10.1057/jibs.2014.5

Silberzahn, R., and Uhlmann, E. L. (2015). Crowdsourced research: many hands make tight work. *Nature* 526, 189–191. doi: 10.1038/526189a

Slonimska, A., and Roberts, S. G. (2016). "A case for systematic sound symbolism in pragmatics: the role of the first phoneme in question prediction in context," in *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (London).

Slonimska, A., and Roberts, S. G. (2017). A case for systematic sound symbolism in pragmatics: Universals in wh-words. *J. Pragmat.* 116, 1–20. doi: 10.1016/j.pragma.2017.04.004

Snell-Rood, E. C. (2012). The effect of climate on acoustic signals: does atmospheric sound absorption matter for bird song and bat echolocation? *J. Acoust. Soc. Am.* 131, 1650–1658. doi: 10.1121/1.3672695

Spruit, M. R. (2006). Measuring syntactic variation in Dutch dialects. *Liter. Linguist. Comput.* 21, 493–506. doi: 10.1093/llc/fql043

Sundarrajan, A., Fujiki, R. B., Loerch, S. E., Venkatraman, A., and Sivasankar, M. P. (2017). Vocal loading and environmental humidity effects in older adults. *J. Voice* 31, 707–713. doi: 10.1016/j.jvoice.2017.02.002

Tamariz, M. (2017a). "Biased and unbiased transmission in language evolution," in *Proceedings of the Minds, Mechanisms and Interaction in the Evolution of Language Workshop* (Nijmegen).

Tamariz, M. (2017b). Experimental studies on the cultural evolution of language. *Ann. Rev. Ling.* 3, 389–407. doi: 10.1146/annurev-linguistics-011516-033807

Thoma, D., and Tytus, A. E. (2017). How cross-linguistic differences in the grammaticalization of future time reference influence intertemporal choices. *Cogn. Sci.* doi: 10.1111/cogs.12525. [Epub ahead of print].

Torreira, F., Roberts, S. G., and Hammarström, H. (2014). "Functional trade-off between lexical tone and intonation: typological evidence from polar-question marking," in *4th International Symposium on Tonal Aspects of Language* (Nijmegen).

Wan, I.-P., and Jaeger, J. (1998). Speech errors and the representation of tone in mandarin chinese. *Phonology* 15, 417–461.

Weide, R. (2005). *The Carnegie Mellon Pronouncing Dictionary [cmudict. 0.6].* Carnegie Mellon University. Avaliable online at: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Weirich, M., and Simpson, A. P. (2014). Differences in acoustic vowel space and the perception of speech tempo. *J. Phonet.* 43, 1–10. doi: 10.1016/j.wocn.2014.01.001

Weisberg, M. (2006). Robustness analysis. *Philos. Sci.* 73, 730–742. doi: 10.1086/518628

Weisberg, M., and Reisman, K. (2008). The robust volterra principle. *Philos. sci.* 75, 106–131. doi: 10.1086/588395

Weizman, Z., and Fletcher, P. (2000). "A comparative study of language development: English and cantonese pre-schoolers in hong kong," in *Committee on Research and Conference Grants, University of Hong Kong* (Hong Kong).

Wichmann, S., Müller, A., Wett, A., Julia Bischoffberger, V. V., Brown, C. H., Holman, E. W., et al. (2013). *The ASJP Database (version 16).* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Wimsatt, W. C. (2012). "Robustness, reliability, and overdetermination (1981)," in *Characterizing the Robustness of Science. Boston Studies in the Philosophy of Science, Vol. 292*, eds L. Soler, E. Trizio, T. Nickles, and W. Wimsatt (Dordrecht: Springer), 61–87.

Winter, B., and Wedel, A. (2016). Commentary: Desiccation and tone within linguistic theory and language contact research. *J. Lang. Evol.* 1, 80–82. doi: 10.1093/jole/lzv010

Woodward, J. (2006). Some varieties of robustness. *J. Econ. Methodol.* 13, 219–240. doi: 10.1080/13501780600733376

Modeling Linguistic Variables With Regression Models: Addressing Non-Gaussian Distributions, Non-independent Observations, and Non-linear Predictors With Random Effects and Generalized Additive Models for Location, Scale, and Shape

Christophe Coupé*

Laboratory Dynamique du Langage, CNRS and University of Lyon, Lyon, France

As statistical approaches are getting increasingly used in linguistics, attention must be paid to the choice of methods and algorithms used. This is especially true since they require assumptions to be satisfied to provide valid results, and because scientific articles still often fall short of reporting whether such assumptions are met. Progress is being, however, made in various directions, one of them being the introduction of techniques able to model data that cannot be properly analyzed with simpler linear regression models. We report recent advances in statistical modeling in linguistics. We first describe *linear mixed-effects regression models* (LMM), which address grouping of observations, and *generalized linear mixed-effects models* (GLMM), which offer a family of distributions for the dependent variable. *Generalized additive models* (GAM) are then introduced, which allow modeling non-linear parametric or non-parametric relationships between the dependent variable and the predictors. We then highlight the possibilities offered by *generalized additive models for location, scale, and shape* (GAMLSS). We explain how they make it possible to go beyond common distributions, such as Gaussian or Poisson, and offer the appropriate inferential framework to account for 'difficult' variables such as count data with strong overdispersion. We also demonstrate how they offer interesting perspectives on data when not only the mean of the dependent variable is modeled, but also its variance, skewness, and kurtosis. As an illustration, the case of phonemic inventory size is analyzed throughout the article. For over 1,500 languages, we consider as predictors the number of speakers, the distance from Africa, an estimation of the intensity of language contact, and linguistic relationships. We discuss the use of random effects to account for genealogical relationships, the choice of appropriate distributions to model count data, and non-linear relationships. Relying on GAMLSS, we assess a range of candidate distributions,

including the Sichel, Delaporte, Box-Cox Green and Cole, and Box-Cox *t* distributions. We find that the Box-Cox *t* distribution, with appropriate modeling of its parameters, best fits the conditional distribution of phonemic inventory size. We finally discuss the specificities of phoneme counts, weak effects, and how GAMLSS should be considered for other linguistic variables.

THE GROWING WEIGHT OF STATISTICS IN LINGUISTICS

Different reasons can be put forward for why data-driven approaches are gaining more prominence in the whole linguistic field. First, large digital datasets such as *WALS* (Dryer and Haspelmath, 2013), *ASJP* (Wichmann et al., 2016), *Lapsyd* (Maddieson et al., 2013), or *D-Place* (Kirby et al., 2016) are freely and readily available for computational analysis. Second, personal computers now offer high computational power, along with efficient and open-source statistical software, like the *R language and environment for statistical computing and graphics* (R Development Core Team, 2017). In particular, advanced modeling techniques that were either still under development or computationally out of reach with affordable computers two decades ago are becoming accessible. Third, such techniques are exported from fields such as econometrics, ecology or genetics to linguistics. While the trend of 'big data' is already well established in subfields of linguistics such as text mining, it has also more recently gained prominence in studies of language diversity or language change. It is for example becoming increasingly common to publish studies investigating more than a thousand languages (e.g., Wichmann et al., 2011; Moran et al., 2012). This is true in particular when the relevance of non-linguistic factors such as sociodemographic ones is being investigated.

With these approaches comes a number of issues regarding the choice of appropriate statistical modeling for the questions at stake. The illusion of truth is dangerous, especially when algorithms deliver arrays of *p*-values without warning of possible misspecifications or violated assumptions. Such issues are a component of the crisis of confidence in psychology (e.g., Earp and Trafimow, 2015): widespread failure to replicate previous studies may be due to different factors, but one of them is likely the inappropriate use of statistical models (e.g., Greenland et al., 2016). This is compounded by the fact that articles often do not report whether the authors have properly checked the assumptions of their tests, nor give sufficient information to replicate the experiment.

A CASE STUDY: THE SIZE OF PHONEMIC INVENTORIES

What drives linguistic diversity? What phenomena, and in particular what external factors, explain the distribution of linguistic structures across the globe? These questions are at the heart of linguistics, and can be considered at various levels of linguistic analysis, either with qualitative or more quantitative approaches. At the phonological level, one of these approaches consists in studying phonemic inventories, and how their size varies across linguistic families and areas. Phonemic inventory size has thus been tentatively related to two non-linguistic variables, namely population size (Hay and Bauer, 2007) and the distance from Africa (Atkinson, 2011a,b), with reference in the second case to modern humans' migrations out of this continent during the last 100,000 years. Both proposals have led to substantial debates (e.g., Pericliev, 2004; Bybee, 2011; Donohue and Nichols, 2011; Moran et al., 2012), both at a theoretical and at a methodological level. Beyond that, language contact and subsequent borrowing – or lack of it –, but also inheritance from parent languages, are obvious partial answers to why a phonemic inventory may be small or large.

In the next sections, we perform as series of regression analyses of phonemic inventory size, in order to illustrate the potentialities and limits of various approaches. In order to do this, we built a dataset of 1529 languages containing 681 languages from the *Lapsyd* database (Maddieson et al., 2013), complemented by 846 languages from the *Phoible* database (Moran et al., 2014). This dataset compiles information for a number of predictors:

- Linguistic families and number of speakers extracted from the *WLMS* dataset (Global Mapping International SIL International, 2012);
- Distance from Africa computed following Atkinson's methodology with a departure point located in eastern Africa (36°W, 8°N) and great circle distances constrained by specific passage points (Sinai region, Bering Strait etc.) (Atkinson, 2011b);
- A measure of local linguistic density, equal for each language to the number of languages spoken less than 50 km away, on the basis of the polygons delimiting the respective areas of these languages, again given by the WLMS dataset; computations were performed with QGIS (Quantum GIS Development Team, 2017).

There were 139 linguistic families, including a number of families restricted to a single language in the case of isolates or creoles. Several transformations were applied to the continuous variables: (i) a natural logarithm transformation was applied to the number of speakers, (ii) a cubic root transformation was applied to the local linguistic density, since it allows to expand the range of values without the issues raised by the log transformation, especially with 0 values, (iii) a scaling without centering was applied to all continuous variables – which is to say, we divided the values of each variable by the standard deviation

of this variable, in order to be able to compare their respective effect sizes in the models.

The choice of predictors reflects recent proposals in the relevant literature, and includes heavily debated variables such as Atkinson's distance from Africa. Together, these predictors provide a rich testbed for the various models considered hereafter; conversely, these models may shed new light on current issues in linguistic diversity, at least at a statistical level with better modeling of the putative influence of geographic and social factors.

Figure 1 provides an overview of *Number of Speakers, Local linguistic density, Distance from Africa,* and *Phonemic inventory size*, and of their one-to-one relationships. The density function of *Distance from Africa* is noticeable because of its three components. These components are related to the distribution of languages on the planet according to the distance from the reference point in eastern Africa. The leftmost component encompasses languages from families such as the Nilo-Saharan, Niger-Congo, and Afro-Asiatic families (left side of the component), but also the Indo-European, Dravidian, Sino-Tibetan, and Austro-Asiatic families (right side of the component). The second component relates mostly to the Trans-New-Guinean, Australian, and Austronesian families, while the rightmost component relates to languages spoken in the Americas, such as Tupi, Macro-Ge, or Arawakan languages. With respect to the locations of passage points for the computation of distances, the "bumps" arise by contrast with regions of lower linguistic diversity, such as in western central Asia and at high latitudes, e.g., in the region of the Bering Strait.

All the regression models were built within the R environment (version 3.4.3) (R Development Core Team, 2017), using various packages that are cited in the following sections. The code used to produce the results and the figures can be found in Supplementary Presentation 1. We remain in a frequentist framework, and therefore do not refer to packages offering Bayesian approaches. Our models always include *Distance from Africa*, *Number of Speakers* and *Local linguistic density* as fixed continuous effects, and *Family* as an intercept random effect for reasons given in section 3.1. The dependent/predicted variable is always *Phonemic inventory size*, also called *Number of phonemes*. In summaries of models, *p*-values lower than 0.05 are in bold, but all exact *p*-values are given unless very small – smaller than 0.001.

ADVANCES IN STATISTICAL MODELING IN LINGUISTICS

How can one identify relationships between phonemic inventory size and the set of predictors mentioned above? Regression models are one of the main methodological answers, especially since they account for several predictors simultaneously. Indeed, the one-to-one relationships between the dependent variable and the various predictors, as exemplified in **Figure 1**, must be considered in the light of a possibly complex network of dependencies between the latter. Since the absence of strong multicollinearity is a prerequisite of regression models, we checked it by computing the variance inflation factors, or VIF, of

the continuous predictors. The three values are between 1 and 1.5, which allow safely concluding to low multicollinearity – values higher than 4 or 5 would have been problematic.

From Linear Regression Models to Mixed-Effects Linear Regression Models

As said previously, regression models relating a *dependent variable*, also known as a *response variable*, to a number of *predictors*, also known as *independent variables* or *explanatory variables*, are common tools. Different approaches, however, fall into this broad category, from straight linear regression to quantile (Cook and Manning, 2013) or ridge (de Vlaming and Groenen, 2015) regression.

There is a growing use of *linear mixed-effects models* (LMM) in linguistics (Jaeger et al., 2011; Johnson, 2014; Winter and Wieling, 2016). In these models, random effects are considered in addition to fixed effects to better account for the distribution of the dependent variable. Random effects allow in particular to account for the issue of non-independence of observations characterized by grouping, known as Galton's problem, which can lead to what is known as pseudo-replication and therefore to increased type I errors, i.e., erroneous significant results (Hurlbert, 1984). As an example in biology, closely related species are assumed to have more similar traits because of their shared ancestry and hence produce more similar residuals from the least squares regression line. Comparatively, in studies investigating a linguistic phenomenon in a large number of languages, not accounting for the increased likelihood that languages sharing a common ancestor share similar features may lead to wrong conclusions in favor of spurious results. If regression is used, this usually leads to the inclusion of linguistic family as a random effect (Atkinson, 2011b). A strategy used by linguists in the field of typology has also been to avoid non-independence by relying on sampling strategies. In Maddieson (1984, p. 158–159)'s work on phonological inventories, the genetic bias was for example controlled by the following method: "*include no pair of languages which had not developed within their own independent speech communities for at least some 1000–1500 years, but to include one language from within each group of languages which shared a history closer than that.*". In experimental linguistics, repeated measurements within subjects or within items are also now usually accounted for with random effects (Baayen et al., 2008).

At the statistical level, including random effects is a more reliable strategy than for example averaging values over subjects or items (Baayen et al., 2008). Such a strategy to bypass the independence problem indeed leads to reduced datasets and a significant loss of information. Random effects fall into random intercepts and random slopes, and with the latter, the impact of predictors entered as fixed effects can be further analyzed across groupings of observations (Barr, 2013). More generally, as underlined by Drager and Hay (2012), random effects are not only a tool to get more accurate models; actually looking at the conditional modes of their levels can provide useful information. For example, if the distribution of levels of a subject random effect reveals that lower values are mostly those of males, and higher values mostly those of females, it is very likely that sex
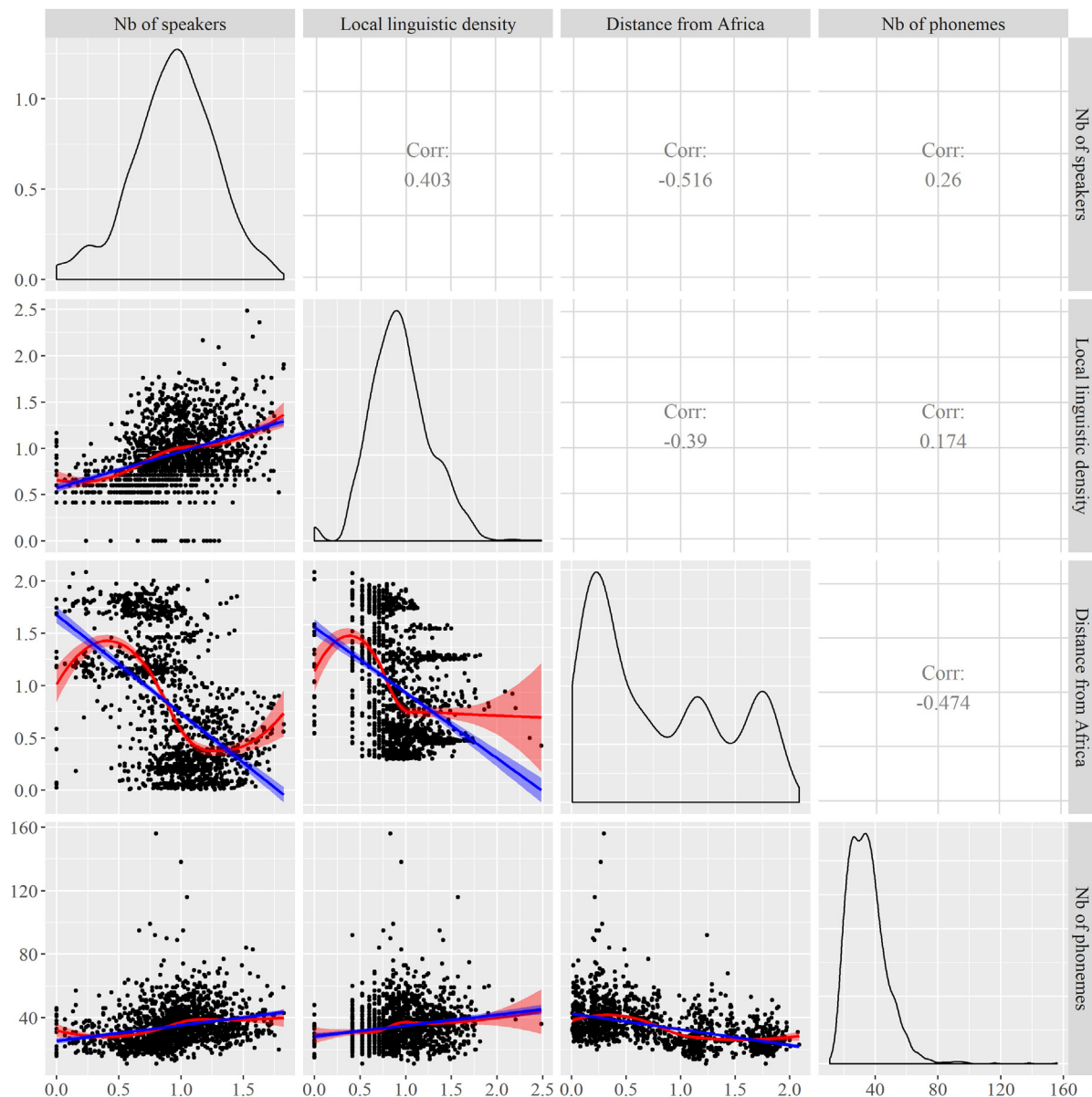
FIGURE 1 | Relationships between *Number of phonemes*, *Number of speakers*, *Distance from Africa,* and *Local language density*. The curves on the diagonal are density curves. The blue lines are linear regressions, the red curves loess regressions.

should be added as a covariate to the model. Upon doing so, the distribution of levels of the subject random effect will likely no longer display a structure according to sex, and its variance will likely be lower.

As in more complex models presented later in this article, the parameters of a LMM can be estimated with different techniques. Besides Bayesian approaches that we do not cover in this article, *maximum likelihood estimation* or MLE is a commonly employed technique. The underlying algorithms aim at finding values of the parameters which maximize the likelihood of observing the sample of data fed to the model. The higher the likelihood, the better the fit to the data. Usually, the logarithm of the likelihood

is given as a measure of the quality of the fit. The so called *log-likelihood* is always negative, and the closer it is to 0, the better the model's goodness of fit is. Conversely, the deviance, *D*, which is equal to minus twice the natural logarithm of the likelihood, is always positive; again, the closer it is to 0, the better the fit of the model.

Both log-likelihood and deviance are good indicators of the quality of the fit, but one is also often interested in the parsimony of computed models. Reaching a good fit with a high number of parameters is for example less parsimonious than reaching the same fit, or a slightly worse one, with only half of them. The *Akaike Information Criterion* or AIC is commonly used

to evaluate parsimony, and penalizes the deviance by twice the number of degrees of freedom in the model, *df*. More precisely, $AIC = 2.df + D$, and the lower the value, the more parsimonious the model. The factor 2 corresponds to a specific tradeoff, and other criteria rest on other values. The *Bayesian Information Criterion* (BIC), also known as *Schwarz Bayesian Criterion* (SBC), is equal to $\ln(n).df + D$, where *n* is the number of observations, i.e., the sample size. The previous definition of BIC, however, assumes that observations are independent, which is not true for example when data are recorded longitudinally, since there is temporal auto-correlation. In such situations, an 'effective sample size' n' must replace *n* (Jones, 2011). Compared to the AIC, the BIC more strongly penalizes models with more parameters, and model selection based on it will therefore tend to promote simpler models. The BIC is thus more conservative against overfitting. The number of degrees of freedom which is part of the computation of AIC and BIC is not easy to estimate when random effects are included in the model – one must rely on approximations such as Satterthwaite or Kenward-Roger. Both the AIC and BIC are specific instances of generalized AIC, or GAIC, which is equal to $k.df + D$, where *k* is a positive real number. There is no a priori reason to choose a specific value of *k* over another, and several measures like AIC and BIC can be used simultaneously to assess the parsimony of several models (Kuha, 2004). Information criteria are hence useful when one tries to select the most appropriate model for a given set of observations and possible predictors (Burnham et al., 2011). While there is no significance test associated with AIC or BIC, they offer more flexibility than for example likelihood ratio tests, which require to compare two models that one is nested into the other. The AIC and BIC values reported for the various models in the next sections have all been rounded up or down to the closest whole number. Two identically reported values may therefore be in fact slightly different.

Turning our attention to our test case, we can compute a LMM with the *lmer()* function provided in the lme4 package (Bates et al., 2015) – one of the better-known packages offering this possibility. *lmer()* takes as inputs the dataset and a formula specifying the predicted variable, the fixed effects and the random effects of the desired model, and outputs estimates for the various parameters of this model. The underlying algorithm uses either a maximum likelihood (ML) or a restricted/residual maximum likelihood (REML) approach. The second differs from the first in the way the variance components that belong to random effects are estimated: REML accounts for the loss in degrees of freedom corresponding to fixed effects, while ML does not. While the variances of random effects may be more accurate when REML is used, ML is the only correct approach when comparing models with different fixed effects. In our case, *Distance from Africa*, *Number of speakers,* and *Local linguistic density* are entered as fixed effects, and could not qualify as random effects given their non-categorical nature. Linguistic families (*Family*) are entered as random intercepts, since following Bolker (2015), these families are chosen from the set of all linguistic families, and we are not primarily interested in the differences, in terms of number of segments, between families – we only

wish to account for the dependencies the latter create in the data.

A random intercept for a categorical variable with *N* levels additionally requires only one parameter to be estimated – the variance, since the mean is fixed to 0 – while a fixed effect would request *N*-1 parameters. This is true if no random slope is simultaneously considered, since covariance between the random slope and the random intercept must then be estimated unless it has been constrained to take a 0 value. We did not consider random slopes in our models, both for the sake of simplicity and because we hypothesized that the impact of the fixed effects did not vary across the linguistic families. We are aware though that this choice could be contested (Barr, 2013).

Table 1 summarizes the output of the model. The lmerTest package is loaded so that the *lmer()* function returns *p*-values with Wald *t*-tests. There are two options to approximate the used degrees of freedom: the Satterthwaite approximation, and the Kenward–Roger approximation which is a slightly more conservative option. Likelihood-ratio tests (LRT), which compare the likelihood of the initial model with that of a model where a target fixed parameter has been dropped, are another option to assess significance. Keeping things simple with *t*-tests, the only *p*-value (well) below 0.05 is for the estimate of *Distance from Africa*. It appears that the further away from the reference point in Africa, the smaller the phonemic inventory size. The estimates for the two other fixed predictors are not significantly different from 0.

How much confidence should we put in these results? Their validity rests upon the satisfaction of a number of assumptions (Zuur et al., 2010), among them the normality of the residuals and their constant variance along the fitted values (homoscedasticity). In **Figure 2**, two diagnostic plots reveal that these requirements

TABLE 1 | Output of a LMM applied to the data.

Predictors	Dependent variable			
	Number of phonemes			
	Estimate	**Standard error**	***t*-value**	***p*-value**
Fixed parts				
(Intercept)	37.75	2.46	15.37	**<0.001**
Distance from Africa	−5.44	1.51	−3.61	**<0.001**
Number of speakers	−1.00	1.06	−0.94	0.348
Local linguistic density	1.39	0.97	1.42	0.155
Random parts				
σ^2		91.46		
$\tau_{00,Family}$		77.81		
N_{Family}		139		
ICC_{Family}		0.46		
Observations		1,529		
R^2/Ω_0^2		0.481/0.478		
AIC		11,435		
Deviance		11,423		

P-values lower than 0.05 are in bold.

FIGURE 2 | Diagnostics for the LMM model: raw residuals vs. fitted values **(left)** and quantile-quantile plot of these residuals **(right)**.
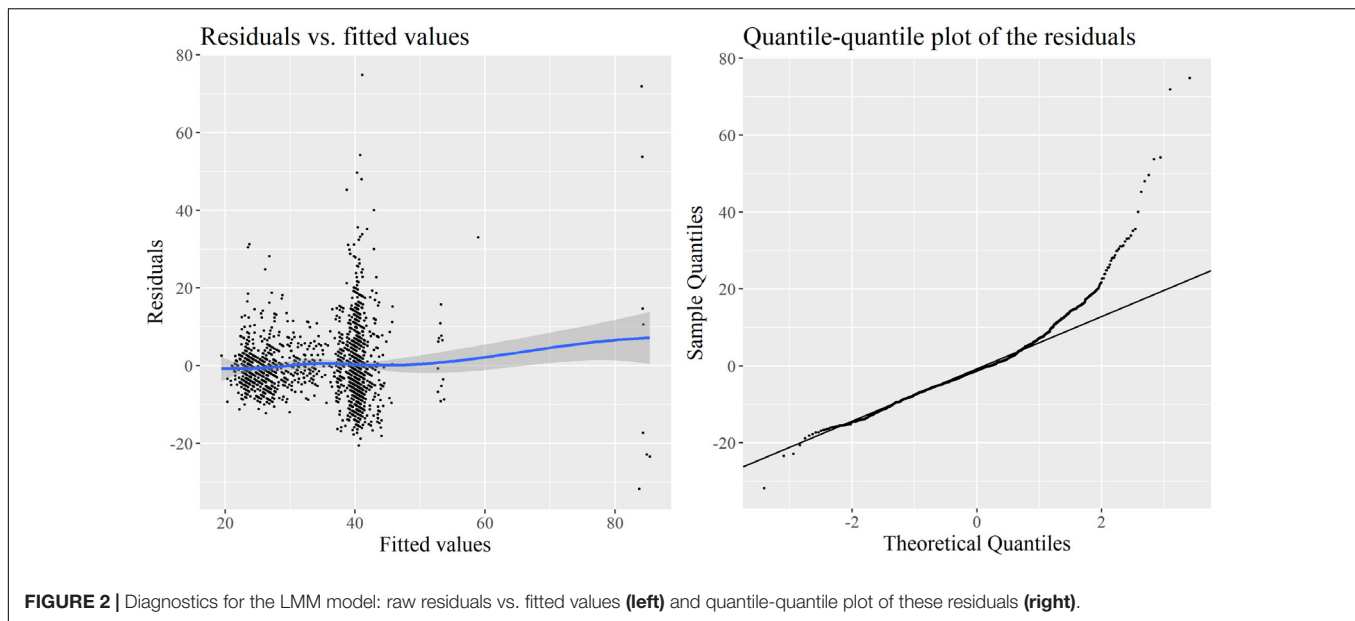
are not met: there is strong heteroscedasticity of the residuals, and a visually clear departure from normality observable in the quantile-quantile plot. The conclusions from the model should therefore be reported with caution, even if LMM are robust to a certain degree of non-normality.

In order to resolve issues of non-normality of the residuals, one commonly found strategy is to transform the dependent variable, whether it is log-transforming count data or taking the inverse of reaction times. The problem is then, however, that a predictor appearing to be significant with respect to the transformed variable is not necessarily significant with respect to the untransformed one, since the mapping between the transformed and untransformed variables is non-linear. In some cases, hypotheses and underlying processes may well concern the transformed variable and not the raw one, in which case it makes perfect sense to apply a transformation. If this is not the case, models based on a transformed dependent variable may not be very informative. All in all, applying non-linear transformations to the predicted variable as the default strategy to overcome statistical issues is therefore not recommended, although these transformations should not be completely discarded. With respect to count data, a number of articles have been published in ecology to discuss log transformation, and overall favor not transforming the data, although linear models with a log transformation often seem robust with large datasets, and may be more resistant to false positives, also known as type I errors (O'Hara and Kotze, 2010; Ives, 2015; Warton et al., 2016). Looking beyond the frequentist framework, Bayesian approaches to predictive uncertainty allow construction of credible intervals in untransformed units from a regression model with a transformed dependent variable (Gelman and Hill, 2007; Korner-Nievergelt et al., 2015). Within the frequentist framework, other modeling options are available, and are described in the next section. Given the inadequacy of the previous LMM

with respect to our test case, it makes sense to consider such options.

It is worth noting that these issues have been highlighted by some authors with respect to phonemic inventory size: an extract of the supporting online material of Cysouw et al. (2012)'s comment on Atkinson (2011b) mentions that '*It has repeatedly been observed that there is a positive correlation between the phoneme inventory size of a language and the speaker community size (S17–S19) (...) Note that for this correlation, we used the logarithm of population size and the logarithm of the phoneme inventory size. The analysis of the expected distribution of phoneme inventory size is still not settled (S20–S22), but using a logarithm seems to be preferable to using the raw numbers*' (p. 14-15). In Atkinson's study, rather than raw or log-transformed inventory sizes, an index of complexity of the phonemic inventories, including tones and with a limited range of values, was considered. The distribution of the dependent variable was therefore very different from ours, and we can argue that the raw number of phonemes provides more information than an index of complexity derived from it.

For the sake of exhaustiveness, we considered a model with the logarithm of the number of phonemes as dependent variable. Despite the transformation, the residuals are still rather unsatisfactory, although more homoscedastic and closer to normality than those of the model with untransformed numbers of phonemes. One could here argue that the log transformation is not the most appropriate, and that other approaches could be considered, such as Box-Cox transformations (Box and Cox, 1964).

In cases where relations between observations can be described with tree-like structures, phylogenetic regression methods can be used to appropriately model the expected structure of covariance between observations, and thus prevent autocorrelation (Symonds and Blomberg, 2014). These models are commonly used in biology and take advantage of the

phylogenetic trees derived from molecular data. However, as for linguistic data, especially when comparing large numbers of languages from distant families, the degree of confidence in the reconstructed tree is often low, at least in the higher branches. This perhaps explains why many studies rely on family level groups in mixed-effects models, despite this being only a very partial account of the relationships between languages. A slight improvement resides in considering several levels of classification, for example with subfamilies nested within families, but again this is only a partial account of the expected covariance between languages.

Generalized Linear Mixed-Effects Regression Models

Generalized linear models (GLM), either with or without random effects, are also on the rise. As their name suggest, they extend linear models, in that they allow the dependent variable to follow a distribution other than Gaussian (the Gaussian distribution which is also called normal distribution). They are particularly, but not only, useful in cases where the predicted variable takes its values in a restricted domain: the set of integer values, the domain of positive real numbers etc. The binomial regression is one case, and suits probabilities or a dependent categorical variable taking two values (Johnson, 2008; Morrison and Kondaurova, 2009). Considering the case of response times, Lo and Andrews (2015) explain how generalized models can come to the rescue of scholars facing two inappropriate choices: analyzing a raw dependent variable when this leads to violation of the assumptions of the linear model, or transforming this raw variable to meet these assumptions (as discussed in section "From Linear Regression Models to Mixed-Effects Linear Regression Models"). The appropriate generalized linear model offers a distribution of error terms leading to the satisfaction of assumptions without transformation. In addition to the conditional distribution of the dependent variable, a link function can also be specified; fixed factors can then linearly predict the result of the application of this function to the observed response, rather than the observed response itself. Among the more common link functions are the logarithm, square-root and inverse functions. Choosing link functions other than the identity function, however, leads once again to the evaluation of predictors with respect to a transformed dependent variable. When including random effects, GLM are usually called generalized linear mixed models, or GLMM.

The commonly available distributions in statistical packages dealing with GLM belong to the exponential family of distributions, such as the normal, Bernoulli, exponential, inverse-Gaussian, chi-squared, Poisson, or binomial (in this latter case, only when the number of trials is known) distributions.

Phonemic inventory size falls into the domain of count data, and it makes sense therefore to consider distributions over positive integers rather than over real numbers. The Poisson distribution is the better known option in such cases. In cases where the counts are small, i.e., close to 0, considering a distribution over real numbers would be dangerous, since predictions of the related model could be non-sensical negative

values. A distribution over positive real numbers seems more appropriate, but exponential distributions like inverse-Gaussian or Gamma are not suited to count data close or equal to 0, unless in very specific cases. When count values are far from 0, however, continuous distributions may be considered, as it is the case for *Phonemic inventory size* – the smallest value is 11, the largest value 156, and the median 33. They may then give better results than discrete distributions. Given these considerations, we thus fitted to our data a Poisson regression, an inverse-Gaussian regression, and a Gamma regression, each time with an identity link function. This choice was motivated by the positive skewness of the distribution of inventory sizes. We used the *glmer()* function of the lme4 package (Bates et al., 2015), in which a few distributions of the exponential family can be specified, including the three previous ones. *glmer()* takes the same inputs as *lmer()* plus the chosen distribution.

The inverse-Gaussian distribution turned out to give the lowest deviance, which was much lower than that of the Poisson regression (10,693 vs. 11,653). The corresponding results (with restricted maximum likelihood – REML) are reported in **Table 2**. They depart from those of the previous LMM in that all the estimates for the fixed effects are closer to 0. The effect of *Distance from Africa* is still significant, but with a higher *p*-value, while *Number of speakers* and *Local linguistic density* are far from being significant. In addition to estimates for fixed predictors being closer to 0, all standard errors are smaller. This observation is a good point for the model.

Again, a number of assumptions must be met for the output of the model to be acceptable. **Figure 3** contains two diagnostic plots for the inverse-Gaussian regression. Heteroscedasticity is more moderate than in the first LMM, but it appears that once again, the distribution of residuals departs from normality, although the problem is much less important than previously, as indicated by the range of sample values. The Gamma regression and the

---

TABLE 2 | Output of an inverse-Gaussian GLMM applied to the data.

Predictors	Dependent variable			
	Number of phonemes			
	Estimate	**Standard error**	**z-value**	**p-value**
Fixed parts				
(Intercept)	35.65	2.33	15.26	**<0.001**
Distance from Africa	−3.52	1.36	−2.59	**0.010**
Number of speakers	−0.04	0.83	−0.05	0.957
Local linguistic density	0.15	0.62	0.18	0.855
Random parts				
$\tau_{00,Family}$		38.47		
N_{Family}		139		
ICC_{Family}		1.00		
Observations		1,529		
AIC		10,705		
Deviance		10,693		

P-values lower than 0.05 are in bold.

FIGURE 3 | Diagnostics for the inverse-Gaussian GLMM: deviance residuals vs. fitted values **(left)** and quantile–quantile plot of these residuals **(right)**.
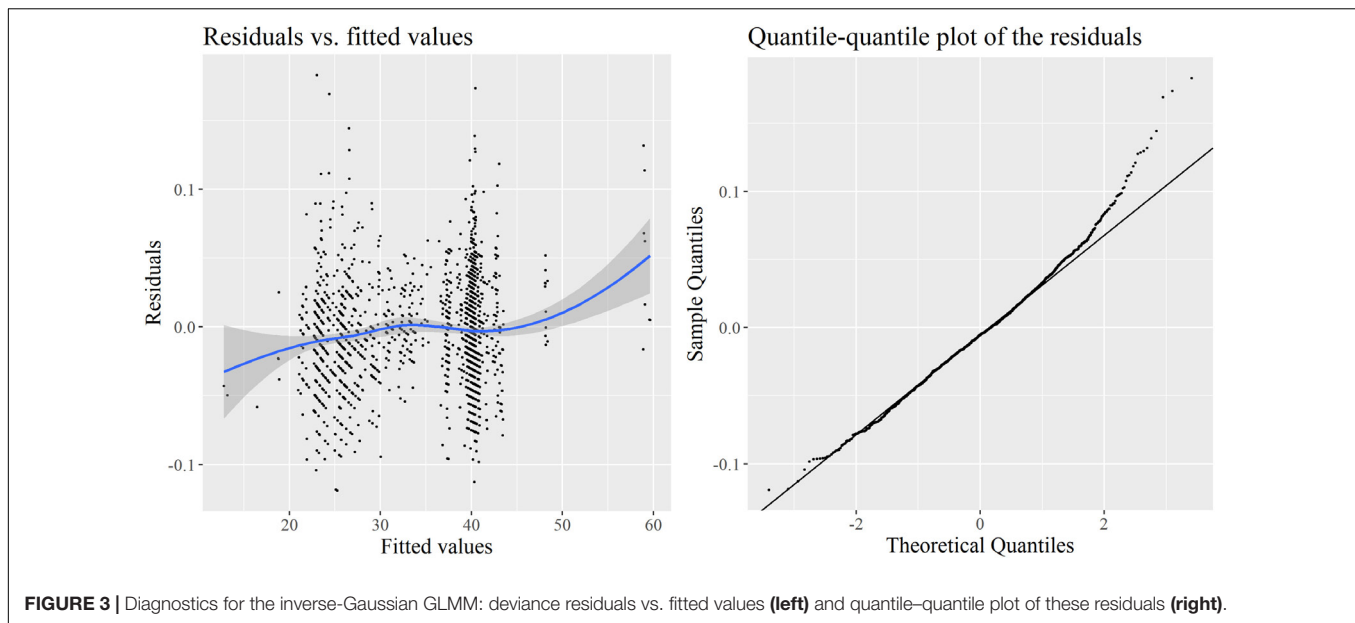
Poisson regression are not better in this respect. In the second case in particular, this is actually not surprising when one knows that the variance of the Poisson distribution is equal to its mean. The marginal distribution of *Phonemic inventory size* has a mean of 34.8, and a variance of 164.8: this is a clear case of strong overdispersion, which makes the Poisson distribution a very unlikely candidate for the regression.

Generalized Additive Models (GAM)

Generalized additive models (GAM) are a family of models which were designed in the 1980s and are widely used today in a range of scientific fields (Hastie and Tibshirani, 1986). They are slowly making their way to linguistics, and a few authors recommend their use, for example in speech analysis (Sóskuthy, 2017).

Generalized additive models are at the intersection between additive models and generalized linear models. They are relevant when the relationship between a continuous predictor and the dependent variable is not adequately described by a linear regression (Wood, 2011; Winter and Wieling, 2016). Adopting a linear regression for a non-linear relationship is dangerous, since it creates autocorrelation patterns in the residuals, and therefore possibly unreliable estimates and confidence intervals for the model parameters (Sóskuthy, 2017). In some cases, non-linear relationships between a predictor and the dependent variable can be expressed by a simple polynomial of this predictor, and LMM or GLMM are then enough, but this is not always the case. GAM address this difficulty by allowing the presence of smoothing functions, or smoothers, in the linear predictor component of the regression model, along with "unsmoothed" covariates. The general equation of a GAM can thus be written:

$$g(E(Y)) = I + s_1(x_1) + \ldots + s_n(x_n) + \varepsilon$$

where $x_1 \ldots x_n$ are the predictors, $s_1(x_1), \ldots, sn(xn)$ the smooth terms relating to these predictors, I the intercept, ε the remaining

error term, Y the dependent variable, $E(Y)$ the expected value and g the link function.

The smooth terms can be either parametric (and this includes the linear and polynomial cases), semi-parametric or non-parametric, univariate or multivariate (in the latter case, to deal with interaction effects); they are overall very unconstrained and therefore very flexible. While this requires noticeably more observations, it can account for predictors and their influence more accurately. However, especially in the case of intricate non-linearities, interpreting the underlying causes can become much harder.

Among the more common parametric smoothers, one finds polynomials, fractional polynomials, piecewise polynomials, or B-splines. Non-parametric smoothers include local regression smoothers, such as the loess regression, which rely on a sliding window to extract local estimators, much in the way speech signals are analyzed to produce spectrograms. They also include penalized smoothers: for a single variable, cubic splines, P-splines, penalized B-splines, penalized categorical variables, Gaussian Markov random fields etc.; for several variables, thin plate regression splines, tensor product splines, varying coefficients etc. (Stasinopoulos et al., 2017, p. 257). While the differences between all these smoothers are beyond the scope of this article, it matters to say that the so-called penalization aims at finding the best value for the smoothing parameter, which controls the amount of smoothing, i.e., the degree of fitting of the smooth term to the raw predictor(s), unless this degree is specified externally by the user. The effective degrees of freedom (*edf*) can be referred to describe the amount of smoothing. The goal is here to avoid both underfitting and overfitting – the bias/variance tradeoff, so that the model can generalize well to data other than the sample used to build it.

Random effects can be included in GAM, in particular under the form of a specific penalized smoother (Stasinopoulos et al., 2017, p. 346). Random slopes can also be considered. One then

speaks of generalized additive mixed models (GAMM), or "mixed GAM." Significantly in a GAM(M), the smooth function of a predictor is estimated while taking into account all other predictors, whether smoothed or not.

In R, common packages for GAM(M) are gam, mgcv, or gamm4 (Wood, 2011), with differences in the underlying MLE algorithms. In mgcv, the function *gamm()* calls to the *lme()* function of the package nlme to estimate random effects, while *gamm4()* calls to *lmer()* or *glmer()*, all these secondary functions being related to LMM or GLMM. As said earlier, random effects can also be specified directly with a penalized smoothing function. It can be noted that the mgcv package enables the use of other distributions than those already mentioned, such as the Tweedie distribution, the zero-inflated Poisson distribution etc. (Wood et al., 2016).

Since the algorithms for MLE differ in GAM(M) and GLM(M), it makes sense to first check the output of an inverse-Gaussian GAM without smoothing functions. We used the *gam()* function of the mgcv package, with a random effect smoother for *Family*. **Table 3** gives the various elements of the model; the random effect clearly appears as a (very significant) smooth term. One can detect variations in the estimates, standard errors and *p*-values; in particular, the estimate for *Distance from Africa* is significantly larger than in the GLMM model. This illustrates the sensitivity of the results to the algorithm, and therefore reminds us to be cautious when concluding on the basis of only "slightly significant" *p*-values. As for GLMM, a Poisson GAM and a Gamma GAM both had higher deviance than the inverse-Gaussian GAM.

Looking back at the various relationships presented in **Figure 1**, several relationships between the predictors and *Phonemic inventory size* suggest that smooth terms may be relevant. The question, however, is whether the non-linear relationship observed on the surface between an isolated predictor and the dependent variable is intrinsic, or whether it

is actually linear under the surface, but appears as non-linear due to the interlaced influence of other predictors. Considering several predictors and smooth terms in a single model allows one to disentangle the various influences at play. As a next step, we thus considered an inverse-Gaussian GAM with smoothers. Finding the most appropriate smoother(s) requires comparing different options and models with measures such as AIC or BIC, and it is generally advisable to estimate the smoothing parameter automatically, i.e., try a penalized version of the smoother. For the sake of simplicity here, we only compared two smoothers that we applied homogeneously to our three continuous fixed effects: cubic splines and P-splines. Regarding the former, the penalty was modified so as to shrink toward zero when the smoothing parameter goes to infinity. Concretely, this meant that an absence of relationship was correctly identified, i.e., with 0 effective degrees of freedom, rather than modeled with one degree of freedom as in standard cubic splines. We actually compared three approaches: penalized cubic splines, penalized P-splines, and cubic splines with a fixed smoothing parameter corresponding to two effective degrees of freedom, i.e., the minimum possible value, corresponding to polynomials of degree 2 ($k = 3$ in the specification of the model). Cubic splines and P-splines are common penalized smoothers, hence our choice; for more information on the differences between them, see (Stasinopoulos et al., 2017, p. 279).

Table 4 reports the outputs of the three models, and **Figure 4** the various smoothing terms for *Distance from Africa*, *Number of speakers*, and *Local linguistic density*. Regarding the numbers in **Table 4**, one should be careful with the standard errors and *p*-values reported for smooth terms. Indeed, these values are unreliable when the smoothing parameters have been penalized by the algorithm, because the uncertainty in the optimization of these parameters is not taken into account when assessing the null hypothesis. In consequence, *p*-values can be too low – again with potential type-I errors leading to falsely rejecting the null hypothesis. Likelihood ratio tests are more conservative than Wald chi-square tests, but results should still be examined with caution. A requirement in the presence of smoothing terms is to perform significance tests with un-penalized smooths, specifying the degree of smoothing as equal to the value obtained previously with penalization (Stasinopoulos et al., 2017, p. 125).

The various graphs in **Figure 4** illustrate the subtleties of using GAM and choosing the right smoothers. As expected, unpenalized cubic splines smooth terms with a fixed number of two degrees of freedom result in relationships which display little "wiggliness". In particular, they suggest a decreasing linear relationship between *Distance from Africa* and *Phonemic inventory size*, other predictors being accounted for. However, despite using less degrees of freedom (113.3 vs. 114.7 and 121.4 for penalized P-splines and cubic splines, respectively), the model has a higher AIC (10,682) than models with penalized P-splines and cubic splines (10,662 and 10,649, respectively). Contrary to what one could have expected, the degrees of freedom are actually only slightly lower than those of the two other models – with a difference of only 1.4 with the P-splines model. A closer look reveals that constraining the smoothness of continuous predictors is counterbalanced by more degrees of freedom used

TABLE 3 | Output of an inverse-Gaussian GAMM without smooth terms.

Predictors	Dependent variable			
	Number of phonemes			
	Estimate	Standard error	z-value	p-value
Parametric coefficients				
(Intercept)	36.25	2.21	16.43	**<0.001**
Distance from Africa	−5.20	1.29	−4.02	**<0.001**
Number of speakers	0.14	0.89	0.15	0.876
Local linguistic density	−0.08	0.83	−0.10	0.922
Smooth term	*edf*	*Ref.df*	*F*	*p*-value
s(Family)	105	138	6.23	**<0.001**
Adjusted R^2		0.416		
AIC		10,679		
Deviance explained		56.5%		

P-values lower than 0.05 are in bold.

TABLE 4 | Output of three inverse-Gaussian GAMM: with cubic splines for continuous predictors (top), with P-splines (middle), with cubic splines and a smoothing parameter fixed to 3 (bottom); in all models, a random effect smoother is applied to the predictor *Family*.

Predictors	Dependent variable			
	Number of phonemes			
Parametric coefficients	**Estimate**	**Standard error**	**z-value**	**p-value**
Cubic splines, starting from *k* = 10				
(Intercept)	33.58	1.15	29.12	**<0.001**
Smooth terms	*edf*	*Ref.df*	*F*	*p-value*
s(Distance from Africa)	8.91	9.00	3017.56	**<0.001**
s(Number of speakers)	0.00	9.00	0.00	1.000
s(Local linguistic density)	8.72	9.00	9.98	0.325
s(Family)	101.69	138	3.87	**<0.001**
Adjusted R^2		0.427		
AIC		10,649		
Deviance explained		57.9%		
P-splines, starting from *k* = 10				
(Intercept)	33.68	1.14	29.47	**<0.001**
Smooth terms	*edf*	*Ref.df*	*F*	*p-value*
s(Distance from Africa)	5.82	6.52	6.67	**<0.001**
s(Number of speakers)	1.70	2.10	0.49	0.653
s(Local linguistic density)	4.47	5.20	1.52	0.207
s(Family)	100.70	138	3.66	**<0.001**
Adjusted R^2		0.418		
AIC		10,662		
Deviance explained		57.2%		
Cubic splines, *k* = 3				
(Intercept)	32.37	1.13	28.55	**<0.001**
Smooth terms	*edf*	*Ref.df*	*F*	*p-value*
s(Distance from Africa)	2	2	8.47	**<0.001**
s(Number of speakers)	2	2	0.42	0.655
s(Local linguistic density)	2	2	0.76	0.466
s(Family)	105.3	138	4.87	**<0.001**
Adjusted R^2		0.416		
AIC		10,683		
Deviance explained		56.5%		

P-values lower than 0.05 are in bold.

by the random effect *Family* (105.3 vs. 100.7 and 101.7 for penalized P-splines and cubic splines). Additionally, comparing the three models shows that 2 degrees of freedom is too much

for *Number of* speakers: The penalized cubic splines model indicates an absence of relationship for this predictor (0 degrees of freedom), while the P-splines model returns 1.7 degrees of freedom. Altogether, these observations suggest that constraining the smooth terms to low degrees of freedom is not a very reasonable choice, and that the related model should rather be left aside. There is more generally no strong argument for choosing a priori 2 rather than 3 or 4 degrees, and penalizing the smooth term is a more neutral approach than starting by constraining the model with imprecise assumptions at the quantitative level.

Comparing now the two models with penalization, one sees that cubic splines lead to high degrees of non-linearity for *Distance from Africa* and *Local linguistic density*, which is reflected by the larger values of the effective degrees of freedom of these two smooth terms (8.90 and 8.72, respectively, to be compared to 5.82 and 4.47 for P-splines), while discarding an influence of *Number of speakers* (owing to the modified penalty introduced above). It looks as if canceling the influence of this predictor resulted in increased non-linearity in the two other continuous predictors. Different smooth functions thus result in different optimizations, something which is likely possible because of the complex correlations between *Distance from Africa*, *Number of speakers* and *Local linguistic density* (see **Figure 1**). Overall, the cubic splines model has the lowest AIC and should therefore be preferred in theory, although it does not provide any simple explanation for the shape of the non-linear relationship between for example *Distance from Africa* and *Phonemic inventory size*. While one may argue that the latter globally decreases with the former, things appear to be more complex than a linear relationship, and this while other predictors have been accounted for. P-splines lead to simpler smooth terms, but interpretation is still difficult. These results are interesting with respect to previous studies in the literature, which have always considered linear predictors rather than smooth terms. Some of the observed effects, as well as some of the contradictory results in different studies, may stem from an inappropriate modeling of non-linear relationships.

Does adding smooth terms to the regression model solve the issue of the non-normality of the residuals? In all previous GAMM models, residuals remain problematic, in a way very similar to those observed in **Figure 3** for the inverse-Gaussian GLMM. Previous observations with cubic splines and P-splines should therefore be treated with caution, and this calls for yet another modeling tool.

GENERALIZED ADDITIVE MODELS FOR LOCATION, SCALE, AND SHAPE (GAMLSS)

Overview

Generalized additive models for location, scale and shapes are an extension of GAM(M) which allows one to consider a wide range of options for the conditional distribution of the dependent variable, while GLM(M) and GAM(M) are
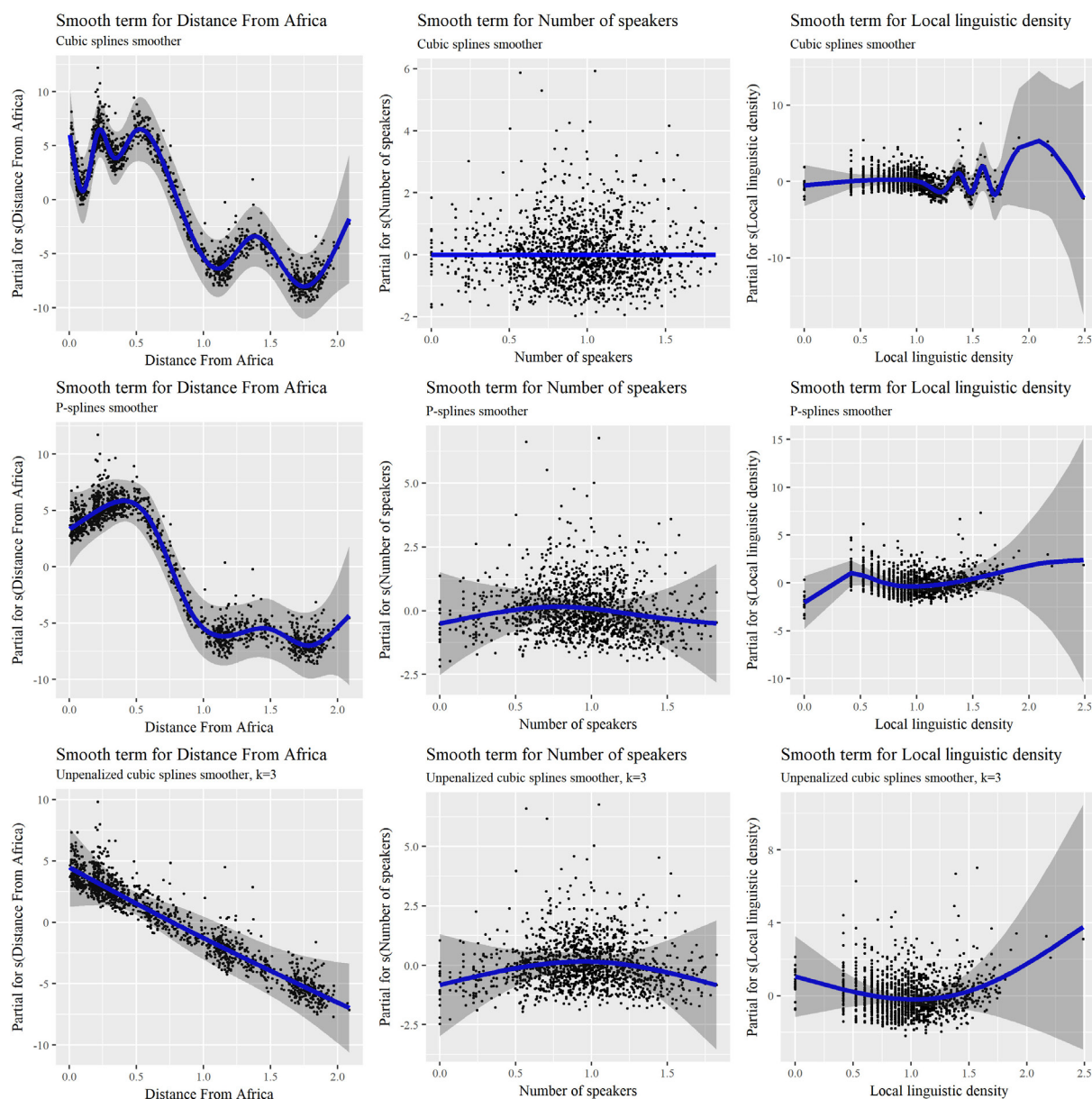
FIGURE 4 | Smooth terms for *Distance from Africa*, *Number of Speakers,* and *Local linguistic density*, for three smoothing approaches in an inverse-Gaussian GAMM: cubic splines (top), P-splines (middle), and cubic splines with a fixed smoothing parameter equal to 3.

restricted to the exponential family of distributions (Rigby and Stasinopoulos, 2005). Besides their range of values – all real numbers, positive real numbers, real numbers between 0 and 1 etc. –, distributions can be contrasted on the basis of their number of parameters: the Poisson distribution is defined with a single parameter, the Gaussian, Gamma, inverse-Gaussian distributions by two parameters etc. Some distributions, such as the generalized Gamma distribution – of which the Gamma and inverse-Gaussian distributions are two specific instances – or the exponential Gaussian distribution, rely on three parameters, while yet other distributions are defined by four parameters, such as the Johnson SU distribution. The terms location, scale, and

shape refer to these various parameters, and are connected, but not necessarily equal, to the four moments of a distribution, namely the mean, the variance, the skewness, and the kurtosis. In the Poisson distribution, the single parameter is a location parameter, equal to the mean, and the scale and shape of the distribution are fixed – this corresponds to the fact that in a Poisson distribution, the variance is equal to the mean, the skewness to the square root of the mean, and the excess kurtosis (the kurtosis minus 3) to the inverse of the mean. In the Gaussian distribution, the mean and variance can be defined independently from each other and are the location and scale parameters, while the skewness and kurtosis, i.e., the shape, are both fixed, equal

to the values 0 and 3, respectively. GAMLSS offer a large variety of distributions with 1, 2, 3, or 4 parameters, classically noted μ, σ, ν, and τ. While only μ is modeled in (G)LM(M) and GAM(M), in GAMLSS all four parameters can be modeled, either with linear parametric, non-linear parametric or non-parametric (smooth) functions of the predictors (Rigby et al., 2007). Normal random effects, but also non-parametric random effects can be considered. Mixtures of distributions can also be used. At the heart of the GAMLSS, algorithms have been designed to fill two tasks: maximize a penalized log-likelihood function addressing the estimates of fixed and random parameters, and evaluate the various smoothing parameters appropriately (Rigby et al., 2007; Stasinopoulos et al., 2017). These two operations cannot be disconnected, and various options are available to perform them in an imbricated way.

An example of the use of GAMLSS is given by Zha et al. (2016) in their analysis of motor vehicle crash data. The predicted variable consists in count data of crashes in highway segments in the United States over the course of several years. As previously stated, the Poisson regression is what usually comes first to mind when count data needs to be assessed. However, as seen for phonemic inventory size, the overdispersion is very high for the number of crashes. The negative binomial distribution better accounts for overdispersion, but by using GAMLSS, Zha et al. (2016) show that a Poisson-Inverse Gaussian provides a better fit and similar predictive performance. They thus suggest that it should be used in subsequent studies to obtain better estimates of the role of predictors. Another example is response times in psycholinguistic experiments. While Lo and Andrews (2015) report that inverse Gaussian and Gamma distributions are equivalent good fits for response times due to theoretical reasons, analysis of experimental data reveals that the distribution of residuals is not always satisfactory, especially because of the long tail of the distribution corresponding to long response times. Relying on distributions better accounting for the skewness of the target distribution, such as the generalized Gamma distribution, leads to more satisfying results in terms of normality of the residuals. Finally, Rigby et al. (2008) discuss various approaches to modeling overdispersed count data, among others 3-parameter Sichel and Delaporte distributions, as well as a 4-parameter distribution, the Poisson-shifted generalized inverse Gaussian distribution.

As for the overall philosophy of GAMLSS, it is interesting to quote Stasinopoulos et al. (2017, p. 26–27): "*GAMLSS provides greater flexibility in regression modeling, but with this flexibility comes more responsibility for the statistician. This is not a bad thing. The philosophy of GAMLSS is to allow the practitioner to have a wide choice of regression models.*"

In R, GAMLSS are available through several packages. The main package is named gamlss, but associated packages such as gamlss.add, gamlss.cens, gamlss.mx, gamlss.spatial etc. allow extending the main functionalities: generation of censored or truncated versions of the main distributions, additional smooth functions such as neural networks or decision trees, use of mixture distributions etc.

Models built with the aforementioned *lmer()*, *glmer()* or *gam()* functions can all be reproduced within the GAMLSS framework.

Given the differences in the algorithms, outputs may, however, slightly differ from one model to the next.

Investigating the Marginal Distribution of Phonemic Inventory Size

A first step in contemplating the use of GAMLSS to study phonemic inventory size is to pay a closer look at the distribution of the latter. The distribution of the dependent variable independently from any predictor is called the marginal distribution.

The *histDist()* and *fitDist()* functions of the gamlss package come in handy to investigate what theoretical distribution comes closest to the empirical one. The first one takes as its main inputs a vector of values and the name of a distribution, and returns how well the values fit the distribution, as expressed by the global deviance, the AIC and BIC of the fit. The second allows one to find the best fit among a list of distributions, and also returns the AIC of the different fitting attempts.

We used these two functions to compare different distributions. On the one hand, we considered distributions adapted to count data available in the gamlss.dist package (loaded by default with the gamlss package). There are over 25 available distributions, among them:

- The 1-parameter Poisson distribution (PO);
- The 2-parameter negative binomial distribution; the types I and II parametrizations (NBI and NBII) available in gamlss led to the same result, and we took the type I;
- The 2-parameter Poisson-Inverse Gaussian distribution (PIG);
- The 3-parameter Delaporte distribution (DEL);
- The 3-parameter Sichel distribution; we considered the second parametrization (SICHEL) offered in gamlss in order for the mean of the distribution to be equal to μ.

We also checked all the distributions adapted to positive real numbers. However, some distributions are based on parameters that are difficult to relate to the four moments mean, variance, skewness, and kurtosis. A location parameter, μ, equal to the mean of the distribution offers easier interpretations, and can be related to LMM, GLMM, and GAMM which all model the mean, and only the mean, of the distribution. This is the case for all previously reported discrete distributions (although with a specific parametrization for the Sichel distribution), but not for all continuous distributions – some of them, however, model the median, which is easy to interpret. Given this constraint of interpretability, we especially paid attention to:

- The 2-parameter inverse-Gaussian distribution (IG), following previous results with GLMM and GAM;
- The 3-parameter Generalized inverse-Gaussian (GIG), a generalization of IG with the mean as location parameter;
- The 3-parameter Box-Cox Cole and Green distribution (BCCG), with the median as location parameter;
- The 4-parameter Box-Cox *t* distribution (BCT), with the median as location parameter;
- The 4-parameter Box-Cox power exponential (BCPE), with the median as location parameter.
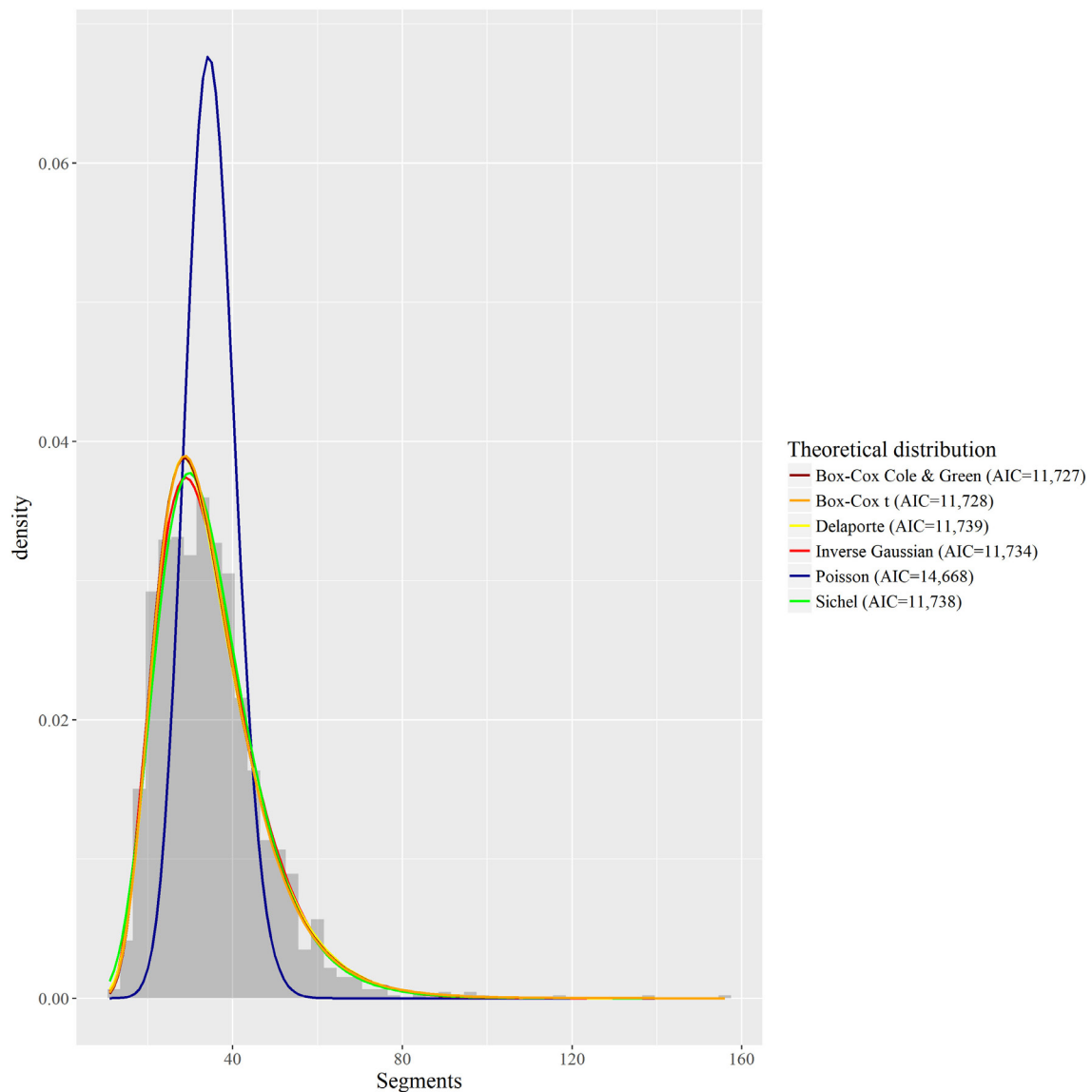
FIGURE 5 | Fitting of several discrete and continuous theoretical distributions to the empirical distribution of *Phonemic inventory size*.

The intuition behind testing these various distributions was that those with more parameters would better be able to account for the thick right tail of the distribution, i.e., the positive skewness of this distribution. **Figure 5** summarizes the fits of the two most adequate discrete distributions, of the two most adequate continuous distributions, and of the Poisson and inverse-Gaussian distributions that were tested in previous models.

Among discrete distributions, the Sichel distribution has the lowest AIC (11,738), but is followed very closely by the Delaporte distribution (AIC = 11,739). The Poisson distribution has a much poorer fit (AIC = 14,668), which is in line with our previous results with GLMM and GAMM. Among continuous positive distributions, the BCCG distribution has the best fit in terms of AIC (11,727), followed by the Generalized Gamma

(AIC = 11,727) which location cannot be easily related to the mean or median, and the BCT distribution (AIC = 11,728). The inverse-Gaussian distribution appears further away in the ranking (AIC = 11,734), but its distance to the best distributions is in no way comparable with how the Poisson distribution differs from the Sichel or Delaporte distributions. As visible on **Figure 5**, except for the Poisson distribution, all displayed theoretical distributions seem rather close to the empirical distribution. One can also observe here that strictly referring to AIC values, the BCCG and BCT distributions provide better fits that the SICHEL and DEL distributions.

Do these results suggest that the BCCG should be the distribution to use in a GAMLSS with our various predictors? One must be cautious here, since the marginal distribution is not the same as the conditional distribution of the dependent

variable, i.e., its distribution when factoring in the various predictors. The question is whether the overdispersion can be explained by one or several of these predictors, or whether it is to some extent independent of them. In the second case, overdispersion will still be manifest in the conditional distribution, and will require treatment with a distribution with the right number of parameters. In the first case, given its degrees of freedom, this distribution will likely still provide good fitting. To this extent, the results obtained with the marginal distribution can serve as a guide in the choice of the target conditional distribution.

Fitting a GAMLSS to Predict Phonemic Inventory Size

In practice, many decisions have to be made regarding the modeling options offered by GAMLSS, from choosing the distribution to choosing the link function, the additive terms and the smoothing parameters. Stasinopoulos et al. (2017, p. 380–384) provide valuable guidelines to operate adequate choices, although no strict sequence of operations can be followed blindly.

In our case, in the previous section, we first investigated the marginal distribution of the dependent variable to narrow down possible choices of distributions. Given the results, one can reasonably focus on a few distributions, namely the Sichel, Delaporte, Box-Cox Cole and Green, and Box-Cox t distributions. We also included the inverse-Gaussian distribution for the sake of comparison with previous models. Second, regarding the link function, we thought that keeping an identity link was useful to relate estimates of the models to actual number of phonemes, without the difficulties related to transforming the dependent variable – or the relationship between it and the predictors – as mentioned earlier in this article. Various link functions can actually be compared with AIC. In distributions requiring positive values, link functions such as the logarithm also prevent convergence issues that are otherwise difficult to address. Third, which additive terms to consider was like in all previous models related to current debates in the literature, which in no way means that other predictors would not be relevant. Various methods of model selection are available, some of them mixing backward, forward, and stepwise procedures across the various parameters of the distribution (Stasinopoulos et al., 2017, p. 385–402). However, besides the fact that some scholars disagree with the concept of model selection overall, the presence of a random effect for *Family* is somehow problematic. Indeed, the way this random effect is estimated in the model – a local normal approximation to likelihood, also known as penalized quasi likelihood – is different from what occurs in common LMM or GLMM – a global estimation to likelihood. The consequence is that dropping a continuous predictor can lead to a change in the penalization of the random effect, such that a strong effect, which should be retained by the selection procedure, may be abandoned. Because of this, we chose not to rely on selection procedures, but rather compare a number of models of increasing complexity. Thus, for each distribution, we considered a model with our predictors only for location (μ), a model with predictors additionally introduced for scale

(σ), then, when possible, models with predictors additionally considered for shape parameters (ν then τ). As for smoothing finally, we considered P-splines smooth functions – cubic splines proved difficult to work with –, with a modified penalty so as to shrink toward zero when the smoothing parameter went to infinity – the *pbz()* smooth function in gamlss (Stasinopoulos et al., 2017, p. 274–275). The advantage of these smooth terms was that the estimation could lead to linear terms, or even to constant terms when no influence of a predictor was detected, other predictors being accounted for. Some parameter selection was thus present.

Table 5 reports the deviance, the degrees of freedom used for the various parameters, the total number of used degrees of freedom, as well as the AIC and BIC of the various models tested. (DEL, μ, σ, and ν) refers for example to a model with the Delaporte distribution, and μ, σ, and ν modeled with our predictors. There were issues of convergence with Sichel models that we could not address, which explains why they are not discussed in what follows. In terms of deviance, the (BCT, μ, σ, and ν) and (BCT, μ, σ, ν, and τ) models had the lowest deviance. These two models were actually identical, which is explained by the fact that all predictors introduced to model τ ended up being estimated with 0 degrees of freedom – in other words, τ was best modeled with an intercept only. In terms of AIC, i.e., taking into account the number of degrees of freedom used by the models, the (DEL, μ, σ, and ν), (BCT, μ and σ) and (BCT, μ, σ, and ν) models were the best, with only a slight difference between them. Finally, the BIC pointed to the three Delaporte models as the most parsimonious.

Which of the previous models to choose, especially given the contradictions between the AIC and BIC? We first decided to prefer (BCT, μ, σ, and ν) over (BCT, μ and σ), since deviance was lower in the first model and since skewness could be better investigated with it. Checking an important assumption – the normality of the residuals – helped us to make a final choice between (BCT, μ, σ, and ν) and Delaporte models. **Figure 6** displays two diagnostic plots of the residuals – one to check homoscedasticity and the other to assess normality – for the (DEL, μ, σ, and ν) and (BCT, μ and σ) models, with (IG, μ) additionally as a reference. While, as previously seen, residuals strongly deviate from normality in (IG, μ), they are much better in (DEL, μ, σ, and ν) and (BCT, μ and σ). However, there is still some deviation in (DEL, μ, σ, and ν). **Figure 7**, which displays detrended quantile-quantile plots – also known as worm plots – provides a much clearer view of the problems of (IG, μ) and (DEL, μ, σ, and ν). In a worm plot, 95% of the dots must be within the 95% confidence interval defined by the two elliptic curves in the figure. This is not the case for the two models. By comparison, the residuals of the (BCT, μ and σ) model are very satisfying, which motivated our decision to adopt this model as the most relevant to further investigate our predictors.

Looking at the various effective degrees of freedom of the smooth terms, it appeared that many terms were actually equivalent to linear predictors, and the model could be simplified and described as follows:

TABLE 5 | Comparisons of various GAMLSS models with different distributions and different levels of modeling of parameters.

Model	Global Deviance	*df* for μ	df for σ	df for ν	df for τ	*df*	AIC	BIC
IG, μ	10,450	104.47	1.00	0	0	105.47	10,661	11,224
IG, μ and σ	10,246	117.55	52.38	0	0	169.93	10,586	11,492
DEL, μ	10,456	79.73	1.00	1	0	81.73	10,619	**11,055**
DEL, μ and σ	10,356	83.24	20.42	1	0	104.66	10,565	**11,123**
DEL, μ, σ, and ν	10,344	83.14	18.86	4.00	0	106.00	**10,556**	**11,121**
BCCG, μ	10,424	105.68	1.00	1	0	107.68	10,640	11,214
BCCG, μ and σ	10,222	121.08	48.71	1	0	170.79	10,563	11,474
BCCG, μ, σ, and ν	10,219	121.38	49.42	3.00	0	173.80	10,567	11,494
BCT, μ	10,403	109.62	1.00	1	1	112.62	10,628	11,228
BCT, μ and σ	10,199	123.14	53.59	1	1	178.73	**10,557**	11,510
BCT, μ, σ, and ν	10,184	124.54	55.24	6.22	1	187.00	**10,558**	11,555
BCT, μ, σ, ν, and τ	10,184	124.54	55.24	6.22	1	187.00	**10,558**	11,555

In each model, a penalized P-spline smooth function is used for the three continuous predictors, and a penalized random effect smoother for the categorical variable. The three lowest AIC and BIC are in bold. IG, inverse-Gaussian; DEL, Delaporte; BCCG, Box-Cox Green and Cole; and BCT, Box-Cox t.
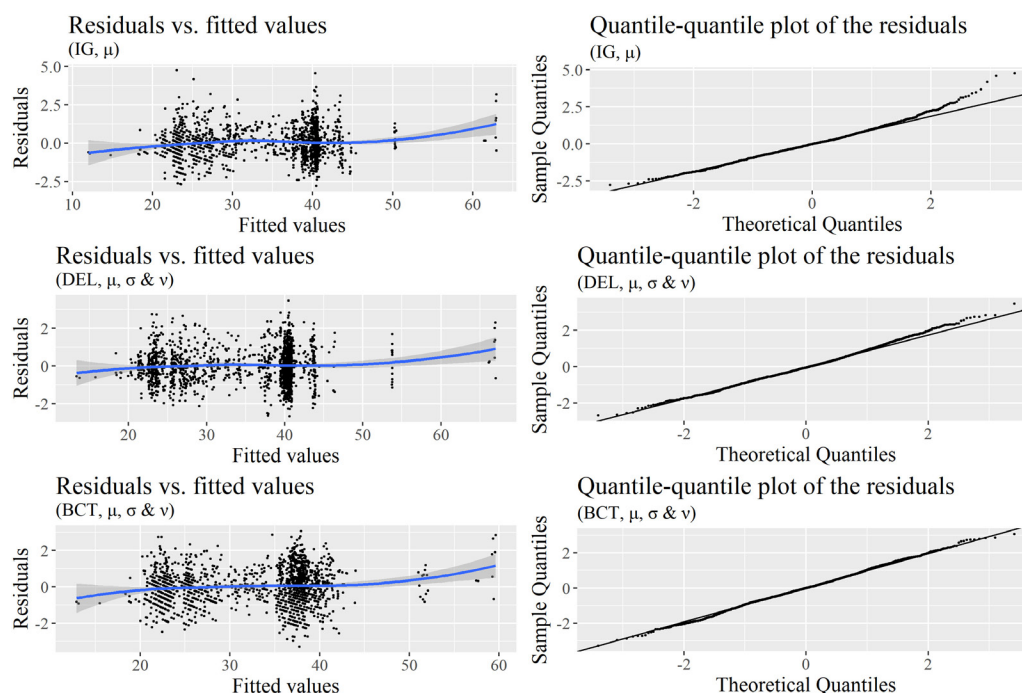
FIGURE 6 | Diagnostics for the (IG, μ), (DEL, μ, σ, and ν), and (BCT, μ, σ, and ν) models reported in **Table 5**: Normalized quantile residuals vs. fitted values **(left)** and quantile-quantile plot of these residuals **(right)**.

- For μ, a smooth term is relevant for *Distance from Africa* and *Family*, and *Number of speakers* and *Local linguistic density* can be included without smoothing;
- For σ, a smooth terms is relevant for *Family*, and the three continuous predictors do not require smoothing;
- For ν, *Family* can be excluded, a smooth term is relevant for *Number of speakers*, but not for *Distance from Africa* and *Local linguistic density*.

Table 6 reports the outputs of this model. Several predictors appear as statistically significant, however, Stasinopoulos et al. (2017, p. 18) warn that *p*-values should be inspected with caution

when smooth terms are present. Indeed, the values given for a smooth term correspond to its linear part, and not to its total contribution. Additionally, reminiscent of what was said for GAM, the values for non-smoothed terms do not account for the uncertainty attached to the estimation of the smoothing terms. A partial solution to this problem is to consider likelihood-ratio tests to assess the significance of the predictors once the degrees of freedom of the smooth terms have been fixed to the values previously estimated with penalization (Stasinopoulos et al., 2017, p. 125). With such fixed smooth terms, dropping a predictor does not result in these smooth terms "reacting" to the drop by increasing their degrees of freedom. The *drop1()* function

FIGURE 7 | Detrended quantile–quantile plots – also known as worm plots – of the normalized quantile residuals for the (IG, μ), (DEL, μ, σ, and ν) and (BCT, μ, σ, and ν) models reported in **Table 5**.
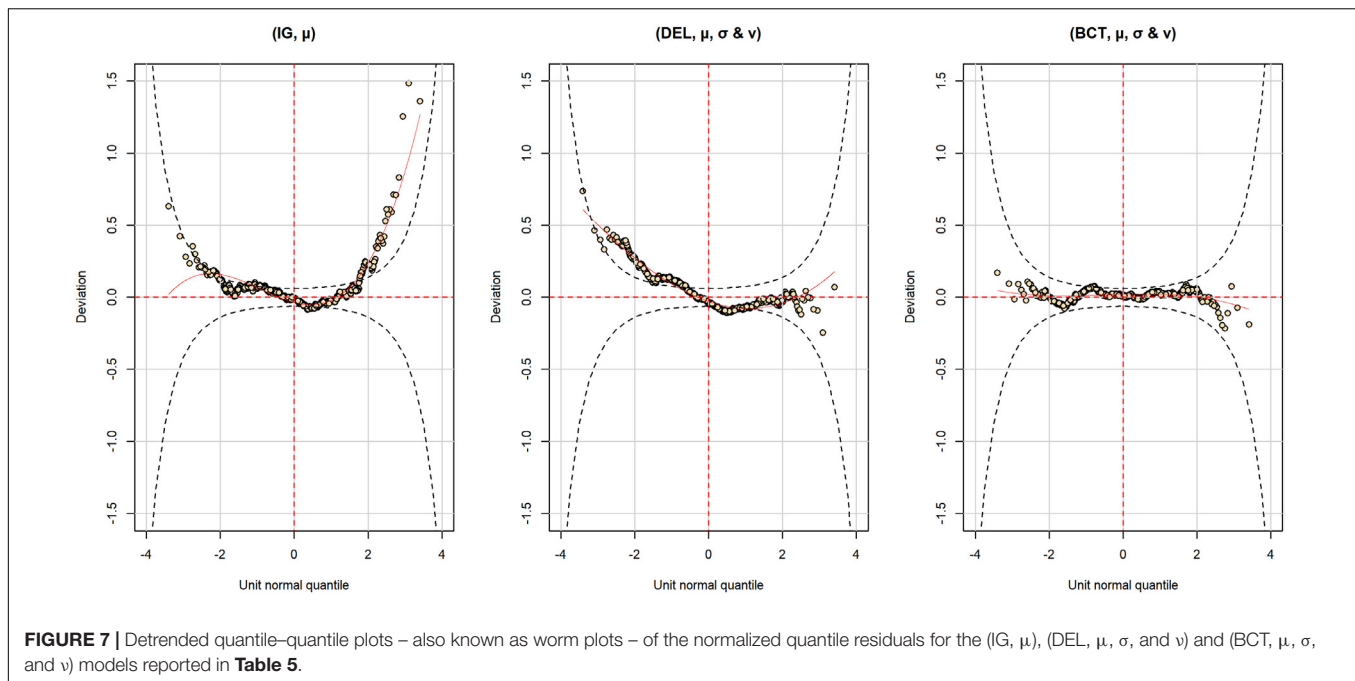
can be used to drop predictors one by one, whether in μ, σ, or ν, and obtain the *p*-value of the chi^2 test involving the full model and the nested model without the dropped predictor (the difference in degrees of freedom is used for the test). **Table 7** reports the output of this function for our chosen model (described in **Table 6**).

Regarding the median of the distribution, the smooth term for *Distance from Africa* is highly significant, while *Local linguistic density* is barely significant and *Number of speakers* is not. With τ constant, σ is approximately proportional to the coefficient of variation (the variance divided by the mean), and is significantly influenced by all predictors but *Number of speakers*. Finally, no predictor reaches the 0.05 significance threshold for ν. One can observe that for P-splines smooth terms, the difference in degrees of freedom between the full model and the model without the smooth term is equal to the fixed number of degrees of this smooth term minus 1. This is because the fixed number of degrees includes one degree for the intercept; when the smooth term is dropped, an intercept remains, hence the "minus 1." One can also ponder here over the benefits of GAMLSS models which, in addition to predictions for the mean or median of the distribution, can also provide information regarding other moments of the distribution. In our case, a conclusion is that the coefficient of variation of the distribution significantly decreases as *Distance from Africa* increases, which means that inventories are more homogeneous in terms of size the further away from Africa, other factors being accounted for.

In order to better understand what is suggested by the model, it is necessary to look at the partial terms reproduced in **Figure 8**. The median of *Phonemic inventory size* is non-linearly related to *Distance from Africa*, and the two local maxima of the non-linear relation are not easy to interpret.

As for GAMM, a linear decrease is not confirmed by the observed pattern. A sharp decrease can, however, be observed at some distance away from Africa. Relations for *Number of speakers* and *Local linguistic density* are linear. While the first one was assessed as not significant, the second one barely is, with an increase of the median *Phonemic inventory size* as the local linguistic density increases. This result was absent in previous LMM, GLMM, and GAMM models. This could be due to less satisfying statistical approaches, but should also serve as a warning of the limited trust one should put in this result.

DISCUSSION

Three aspects can be put forward in discussing the previous results and observations.

The first aspect concerns the specific nature of our target dependent variable, i.e., phonemic inventory size. The very large inventories of some languages, and the overdispersion of the connected variable, can be in good part explained by how features are combined into phonemes. The notion of feature economy states that "*speech sounds tend to be organized by a principle of feature economy, according to which languages maximize the combinatory possibilities of a few phonological features to generate large numbers of speech sounds*" (Clements, 2003, p. 371). According to this principle, very large inventories are so because some features are used intensively and produce series of phonemes "in mirror," e.g., the vocalic feature of nasalization is put to use so that all vowels without secondary features have their nasalized counterparts. Multiplicative processes are therefore at the origin of at least some the variance and overdispersion of phonemic inventory size.

TABLE 6 | Output of a GAMLSS with (i) Box-Cox t distribution, (ii) μ, σ, and ν modeled with either linear predictors or penalized P-splines smooth functions of these predictors, and a penalized random effect smoother for the categorical variable *Family* when necessary, (iii) τ modeled as intercept only.

Predictors	Dependent variable			
	Number of phonemes			
Parametric coefficients	**Estimate**	**Standard error**	**t-value**	**p-value**
μ (link function: identity)				
(Intercept)	36.40	0.79	46.16	**<0.001**
s(Distance from Africa)	−6.05	0.30	−20.19	**<0.001**
Number of speakers	−0.19	0.52	−0.36	0.716
Local linguistic density	1.27	0.52	2.47	**0.014**
σ (link function: log)				
(Intercept)	−1.47	0.10	−15.16	**<0.001**
Distance from Africa	−0.26	0.04	−7.14	**<0.001**
Number of speakers	0.05	0.07	0.78	0.432
Local linguistic density	−0.14	0.07	−2.09	**0.037**
ν (link function: identity)				
(Intercept)	−1.00	0.51	−1.97	**0.048**
Distance from Africa	0.15	0.19	0.78	0.438
s(Number of speakers)	0.25	0.34	0.74	0.456
Local linguistic density	−0.02	0.30	−0.08	0.932
τ (link function: log)				
(Intercept)	11.97	3.39	3.53	**<0.001**
Smooth terms			**edf**	**σ_B**
μ (link function: identity)				
s(Distance from Africa)			8.83	1.08
s(Family)			113.71	6.96
σ (link function: log)				
s(Family)			51.23	0.31
ν (link function: identity)				
s(Number of speakers)			4.22	0.11
Global deviance		10,184		
AIC		10,558		
BIC		11,555		

Regarding the parametric coefficients, the coefficient of a smoothing term and its standard error refer to its linear component. P-values lower than 0.05 are in bold.

TABLE 7 | Likelihood ratio tests (LRT) for the predictors of the (BCT, μ, σ, and ν) GAMLSS model described in **Table 6**.

	df	**AIC**	**LRT**	**p-value**
μ				
Starting model		10,558		
s(Distance from Africa)	7.82	10,571	29.28	**<0.001**
Number of speakers	1	10,556	0.22	0.640
Local linguistic density	1	10,560	4.11	**0.043**
s(Family)	113.71	10,874	544.30	**<0.001**
σ				
Starting model		10,558		
Distance from Africa	1	10,561	5.81	**0.016**
Number of speakers	1	10,556	0.12	0.726
Local linguistic density	1	10,561	5.39	**0.020**
s(Family)	51.23	10,600	145.24	**<0.001**
ν				
Starting model		10,558		
Distance from Africa	1	10,557	1.39	0.239
s(Number of speakers)	3.22	10,558	7.29	**0.074**
Local linguistic density	1	10,556	0.09	0.765

P-values lower than 0.05 are in bold.

the mechanisms hypothesized with the inclusion of a predictor. For example, referring to the impact of the number of speakers, does one conceive this impact at the level of phonemes, or at the level of features? In the latter case, the transformation would perhaps be justified. In the former, some situations could appear as less convincing. Although this hypothesis is far-fetched and is only put forward to the sake of argumentation, one could argue that having a larger number of speakers does not increase the number of features at the basis of the phonemic inventory, but rather influences the way speakers combine these features, in such a way that the system tends to display greater feature economy. Along the same line of thought, with respect to linguistic contact and the putative effect of the local linguistic density, the meaningful question would be whether speakers mostly borrow phonemes or features from other languages. In any case, one of the messages of this article is that models do exist that allow one to model "difficult" variables without resorting to transformation.

To move further in this direction, future work will consist in extracting the features of each phonemic inventory used in the test case of this article. It will then become possible to study the distribution of feature inventory size, much in the way phonemic inventory size was scrutinized in the previous sections. There are no multiplicative processes at the level of features, and it will therefore be relevant to evaluate the overdispersion of the marginal and conditional distributions. If overdispersion is still present and high, a possible conclusion will be that the overdispersion of phonemic inventory size derives from multiplicative processes when combining features, but also from the properties of the systems of features themselves.

A second point is the issue of weak effects in regression modeling. As it appears from our various analyses, *Distance*

From this observation, one could argue that applying a transformation to the dependent variable makes sense, even if it is not an easy question to answer which transformation is respectful of the specific multiplicative processes at play. However, this transformation may run counter to the nature of

FIGURE 8 | Additive terms for *Distance from Africa*, *Number of Speakers,* and *Local linguistic density* for the μ and σ components of the (BCT, μ, σ, and ν) model reported in **Tables 5**, **6**. A P-splines smooth function is applied to *Distance from Africa* for μ.

from Africa appears as a very significant effect in all models. One can assume that very strong and significant effects will be observed even with imperfect models. However, what about weaker effects, with significance close to the 5% threshold? Another predictors of our models, *Local linguistic density*, has *p*-values (well) above 0.05 in less satisfying models, and a *p*-value barely below 0.05 in the supposedly most appropriate model. Drawing conclusion about weak effects is very dependent on the model, especially if one clings to the 5% significance threshold, and also on the use of one test of significance over another: Wald *t*-tests, likelihood ratio tests, parametric bootstrapping etc. (Luke, 2017). On the one hand, some scientists advocate for moving away from the "null ritual" and the 5% threshold (Gigerenzer et al., 2004; Baker, 2016; Greenland et al., 2016; Wasserstein and Lazar, 2016), in which case differences between *p*-values slightly below or above 0.05 do not matter much. On the other hand, a conclusion is that weak signals are at the mercy of the chosen model, and thus this model should be chosen and assessed with care. For example, in the case of phonemic inventories, in addition to the assumptions we tested for residuals, potential spatial autocorrelation should be accounted for in order to minimize related type I errors. We have not addressed this concern in the previous models, but some options are available, whether it's moving to regression models including spatial correlation structures, or including specific predictors such as the 'weighted areal normalized phonological diversity' proposed by Jaeger et al. (2011). All in all, with respect to our test case, whether language contact significantly affects phonemic inventory size through borrowing remains to us an open question. What geo-linguistic measures best capture language borrowing is a connected question that requires further investigation.

Finally, we argue that linguistics and psycholinguistics could benefit from the use of GAMLSS when regression models are envisaged to explore a phenomenon. The adequacy of the Delaporte distribution to model phonemic inventory size in no way means that this distribution in particular is *the* solution to a large number of problems. Rather, we have tried to highlight the reasoning that led us to consider this distribution, and why other options – LMM, GLMM, GAMM, GAMLSS with other distributions – were not as much appropriate. In other contexts, similar investigations would lead to another distribution or narrow choice of distributions. One domain of application already mentioned in Section "Overview" is the study of response times in psycholinguistics. In addition to finding appropriate theoretical distributions for the very specific distribution of reaction times (Moscoso Del Prado Martín, 2009; Baayen and Milin, 2010), a potentially fruitful advantage of GAMLSS is their

ability to not only model mean, but also variance and skewness. Relating the mean of response times as dependent variable to a number of factors such as number of phonological neighbors, frequency, number of letters etc. is very common, but doing the same for the variance or the skewness could help further unravel the way cognitive treatment unfolds and linguistic information is processed.

Besides psycholinguistics, work in preparation suggests that another variable which can benefit from GAMLSS is speech rate. Indeed, speech rate – the number of syllables uttered by second – presents interesting variations between speakers and languages (Pellegrino et al., 2011; Coupé et al., 2014), but distributions in speakers and languages also suggest meaningful patterns of skewing, where the amount and orientation of skewing is connected to the mean value of the speech rate.

More generally, we have little doubt that many other variables, either continuous, discrete or count data, can benefit from both the smooth functions and distributions of GAMLSS.

CONCLUSION

Various statistical tools are available to linguists willing to explain how a given linguistic variable varies across its domain. We highlighted how GAMLSS models, which are still very rarely used in the language sciences, could be put to use to depict 'complex' variables such as phonemic inventory size. This seems especially relevant when non-linguistic causes of linguistic diversity such as climatic or sociodemographic factors are considered, since their study can often be conducted with regression models. The distributions offered by GAMLSS can be more appropriate from a methodological point of view, and both the possibility to include additive terms and the possibility to model the scale and shape of the distribution in addition to its location can be put to use to better understand the behavior of a system.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the author, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

The author designed the work, assembled the studied dataset from other sources of data, conducted the different analyses, and wrote the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00513/full#supplementary-material

REFERENCES

Atkinson, Q. D. (2011a). Linking spatial patterns of language variation to ancient demography and population migrations. *Linguist. Typol.* 15, 321–332. doi: 10.1515/LITY.2011.022

Atkinson, Q. D. (2011b). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–349. doi: 10.1126/science.1199295

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

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28. doi: 10.1287/mksc.12.4.395

Baker, M. (2016). Statisticians issue warning over misuse of P values. *Nature* 531:151. doi: 10.1038/nature.2016.19503

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* 4:328. doi: 10.3389/fpsyg.2013.00328

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bolker, B. M. (2015). "Linear and generalized linear mixed models," in *Ecological Statistics: Contemporary Theory and Application*, eds

G. A. Fox, S. Negrete-Yankelevich, and V. J. Sosa (Oxford: Oxford University Press), 309–333. doi: 10.1093/acprof:oso/9780199672547.003.0014

Box, G., and Cox, D. (1964). An analysis of transformations. *J. R. Stat. Soc. Ser. B* 26, 211–252.

Burnham, K. P., Anderson, D. R., and Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65, 23–35. doi: 10.1007/s00265-010-1029-6

Bybee, J. (2011). How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate. *Linguist. Typol.* 15, 147–153. doi: 10.1515/LITY.2011.009

Clements, G. N. (2003). "Feature economy as a phonological universal," in *Proceedings of the 15th International Congress of Phonetic Sciences*, (Barcelona: Futurgraphic), 371–374.

Cook, B. L., and Manning, W. G. (2013). Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. *Biostat. Psychiatry* 25, 55–59. doi: 10.3969/j.issn.1002-0829.2013.01.011

Coupé, C., Oh, Y. M., Pellegrino, F., and Marsico, E. (2014). "Cross-linguistic investigations of oral and silent reading," in *Proceedings of the 15th Interspeech Conference*, Écully, 514–518.

Cysouw, M., Dediu, D., and Moran, S. (2012). Comment on "Phonemic diversity supports a serial founder effect model of language expansion from Africa". *Science* 335:657. doi: 10.1126/science.1208841

de Vlaming, R., and Groenen, P. J. F. (2015). The current and future use of ridge regression for prediction in quantitative genetics. *Biomed. Res. Int.* 2015:143712. doi: 10.1155/2015/143712

Donohue, M., and Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguist. Typol.* 15, 161–170. doi: 10.1515/LITY.2011.011

Drager, K., and Hay, J. (2012). Exploiting random intercepts: two case studies in sociophonetics. *Lang. Var. Change* 24, 59–78. doi: 10.1017/S0954394512000014

Dryer, M. S., and Haspelmath, M. (2013). *The World Atlas of Language Structures*. Available at: http://wals.info

Earp, B. D., and Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Front. Psychol.* 6:621. doi: 10.3389/fpsyg.2015.00621

Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevelhierarchical Models*, Vol. 1. Cambridge: Cambridge University Press.

Gigerenzer, G., Krauss, S., and Vitouch, O. (2004). "The null ritual. What you always wanted to know about significance testing but were afraid to ask," in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan (Thousans Oaks, CA: Sage), 391–408. doi: 10.4135/9781412986311.n21

Global Mapping International SIL International (2012). *World Language Mapping System. Language Area and Point Data for Geographic Information Systems (GIS)*. Available at: http://www.worldgeodatasets.com/language/

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* 31, 337–350. doi: 10.1007/s10654-016-0149-3

Hastie, T., and Tibshirani, R. (1986). Generalized additive models. *Stat. Sci.* 1, 297–318. doi: 10.1016/j.csda.2010.05.004

Hay, J., and Bauer, L. (2007). Phoneme inventory size and population size. *Language* 83, 388–400. doi: 10.1353/lan.2007.0071

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54, 187–212. doi: 10.2307/1942661

Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods Ecol. Evol.* 6, 828–835. doi: 10.1111/2041-210X.12386

Jaeger, T. F., Graff, P., Croft, W., and Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguist. Typol.* 15, 281–320. doi: 10.1515/LITY.2011.021

Johnson, D. E. (2014). *Progress in Regression: Why Natural Language Data Calls for Mixed-Effects Models*. Available at: http://www.danielezrajohnson.com/

Johnson, K. (2008). *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Jones, R. H. (2011). Bayesian information criterion for longitudinal and clustered data. *Stat. Med.* 30, 3050–3056. doi: 10.1002/sim.4323

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., et al. (2016). D-PLACE: a global database of cultural, linguistic and environmental diversity. *PLoS One* 11:e0158391. doi: 10.1371/journal.pone.0158391

Korner-Nievergelt, F., Roth, T., von Felten, S., Guélat, J., Almasi, B., and Korner-Nievergelt, P. (2015). *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan*. Cambridge, MA: Academic Press.

Kuha, J. (2004). AIC and BIC: comparisons of assumptions and performance. *Sociol. Methods Res.* 33, 188–229. doi: 10.1177/0049124103262065

Lo, S., and Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6:1171. doi: 10.3389/fpsyg.2015.01171

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511753459

Maddieson, I., Flavier, S., Marsico, E., Coupé, C., and Pellegrino, F. (2013). "LAPSyD: Lyon - albuquerque phonological systems database," in *Proceedings of the 14th Interspeech Conference*, Albuquerque, NM, 3022–3026.

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language* 88, 877–893. doi: 10.1353/lan.2012.0087

Moran, S., McCloy, D., and Wright, R. (2014). *PHOIBLE Online*. Available at: http://phoible.org/

Morrison, G. S., and Kondaurova, M. V. (2009). Analysis of categorical response data: use logistic regression rather than endpoint-difference scores or discriminant analysis (L). *J. Acoust. Soc. Am.* 126, 2159–2162. doi: 10.1121/1.3216917

Moscoso Del Prado Martín, F. (2009). *A Theory of Reaction Time Distributions*. Available at: http://cogprints.org/6326/

O'Hara, R. B., and Kotze, D. J. (2010). Do not log-transform count data. *Methods Ecol. Evol.* 1, 118–122. doi: 10.1111/j.2041-210X.2010.00021.x

Pellegrino, F., Coupé, C., and Marsico, E. (2011). A cross-language perspective on speech information rate. *Language* 87, 539–558. doi: 10.1353/lan.2011.0057

Pericliev, V. (2004). There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. *Linguist. Typol.* 8, 376–383. doi: 10.1515/lity.2004.8.3.376

Quantum GIS Development Team (2017). *Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project*. Available at: http://qgis.org/en/site/

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rigby, R. A., and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Appl. Stat.* 54, 507–554. doi: 10.1016/j.placenta.2014.02.001

Rigby, R. A., Stasinopoulos, D. M., and Akantziliotou, C. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Comput. Stat. Data Anal.* 53, 381–393. doi: 10.1016/j.csda.2008.07.043

Rigby, R. A., Stasinopoulos, D. M., and Lane, P. W. (2007). Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C Appl. Stat.* 54, 507–554. doi: 10.1111/j.1467-9876.2005.00510.x

Sóskuthy, M. (2017). *Generalised Additive Mixed Model for Dynamic Analysis in Linguistics: A Practical Introduction*. Available at: http://eprints.whiterose.ac.uk/113858/

Stasinopoulos, M., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F. (2017). *Flexible Regression and Smoothing Using GAMLSS in R*. Boca Raton, FL: CRC Press.

Symonds, M. R. E., and Blomberg, S. P. (2014). "A primer on phylogenetic generalised least squares," in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, ed. L. Z. Garamszegi (Berlin: Springer), 105–130.

Warton, D. I., Lyons, M., Stoklosa, J., and Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods Ecol. Evol.* 7, 882–890. doi: 10.1111/2041-210X.12552

Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's statement on p-values: context, Process, and Purpose. *Am. Stat.* 70, 129–133. doi: 10.1080/00031305.2016.1154108

Wichmann, S., Holman, E. W., and Brown, C. H. (2016). *The ASJP Database*. Available at: http://asjp.clld.org/

Wichmann, S., Rama, T., and Holman, E. W. (2011). Phonological diversity, word length, and population sizes across languages: the ASJP evidence. *Linguist. Typol.* 15, 177–197. doi: 10.1515/LITY.2011.013

Winter, B., and Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *J. Lang. Evol.* 1, 7–18. doi: 10.1093/jole/lzv003

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc.* 73, 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *J. Am. Stat. Assoc.* 111, 1548–1563. doi: 10.1080/01621459.2016.1180986

Zha, L., Lord, D., and Zou, Y. (2016). The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. *J. Transp. Saf. Secur.* 8, 18–35. doi: 10.1080/19439962.2014.977502

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* 1, 3–14. doi: 10.1111/j.2041-210X.2009.00001.x

Language Adapts to Environment: Sonority and Temperature

Ian Maddieson *

Department of Linguistics, University of New Mexico, Albuquerque, NM, United States

The phonetic patterns of human spoken languages have been claimed to be in part shaped by environmental conditions in the locales where they are spoken. This follows predictions of the Acoustic Adaptation Hypothesis, previously mainly applied to the study of bird song, which proposes that differential transmission conditions in different environments explain some of the frequency and temporal variation between and within species' songs. Prior discussion of the relevance of the Acoustic Adaptation Hypothesis to human language has related such characteristics as the total size of the consonant inventory and the complexity of the permitted maximum syllable structure, rather than patterns in continuous speech, to environmental variables. Thus the relative frequency with which more complex structures occur is not taken into account. This study looks at brief samples of spoken material from 100 languages, dividing the speech into sonorous and obstruent time fractions. The percentage of sonorous material is the sonority score. This score correlates quite strongly with mean annual temperature in the area where the languages are spoken, with higher temperatures going together with higher sonority scores. The role of tree cover and annual precipitation, found to be important in earlier work, is not found to be significant in this data. This result may be explained if absorption and scattering are more important than reflection. Atmospheric absorption is greater at higher temperatures and peaks at higher frequencies with increasing temperature. Small-scale local perturbations (eddies) in the atmosphere created by high air temperatures also degrade the high-frequency spectral characteristics that are critical to distinguishing between obstruent consonants, leading to reduction in contrasts between them, and fewer clusters containing obstruent strings.

Keywords: acoustic adaptation hypothesis, language and environment, sonority, running speech, temperature

BACKGROUND

Any communication system using an acoustic channel is inevitably subject to filtering and masking effects which modify the faithfulness of the transmission of a signal. Once any acoustic signal is emitted from its source its characteristics will be modified by a wide variety of factors before it reaches any recipient. When considering sounds transmitted through open air, the temperature and density of the air, the nature of the ground surface below and the presence of obstacles and their surface characteristics are among the various factors that impact both the spectral and temporal characteristics of a signal (Harris, 1966, 1967; Aylor, 1972; Marten and Marler, 1977; Marten et al., 1977; Piercy et al., 1977; Wiley and Richards, 1978; Richards and Wiley, 1980; Martens and Michelsen, 1981; Bass et al., 1984; Martens, 1992; Attenborough et al., 1995, 2011; Embleton, 1996; Sutherland and Daigle, 1998; Wilson et al., 1999; Salomons, 2001; Naguib, 2003; Albert, 2004).

Moreover, the presence of any competing sounds in the environment can affect a hearer's perception of the properties of a signal. Sound is generated by wind, rainfall, flowing water, birds, insects and other creatures, among other sources. Environmental sounds of this kind can selectively mask some characteristics of an acoustic signal in natural settings (Winkler, 2001; Slabbekoorn, 2004a,b; Brumm and Slabbekoorn, 2005).

While a good deal of the research on outdoor sound propagation has been directed to addressing practical issues relevant to humans, such as the mitigation of vehicle or aircraft noise (Salomons, 2001) or the calculation of the source of weapons fire (Beck et al., 2011), a considerable amount of work has also been devoted to the potential effects of both filtering and masking on the design of biological acoustical communication systems. Several basic principles have been put forward (Bradbury and Vehrencamp, 1988; Hauser, 1996; Römer, 2001; Ryan and Kime, 2002). The Acoustic Niche Hypothesis (Krause, 1987, 1993; Farina et al., 2011) proposes that different species tend to avoid competition for the same frequency band and time window, which reduces the impact of masking. Related to this proposition, several studies have shown that song birds in urban areas seem to be raising the pitch of their songs in response to the pervasive presence of lower-frequency human-generated machine noise (Slabbekoorn and Peet, 2003; Wood and Yezerinac, 2006) and Slabbekoorn and Smith (2002) suggest that little greenbul (*Andropadus virens*) populations adapt their songs to lessen interference from ambient noise. The Acoustic Adaptation Hypothesis (AAH) proposes that the acoustic communications of biological organisms are in part shaped by the transmission characteristics of the environment in which they are employed. There seems a broad consensus that the evidence for this is particularly clear with respect to bird song, the AAH having been particularly studied in this context (e.g., Chappuis, 1971; Morton, 1975; Seddon, 2005; Boncoraglio and Saino, 2007). This research has indicated that such factors as the typical density of vegetation in a species' habitat correlate with both spectral and temporal properties of bird songs. In the spectral domain, Boncoraglio and Saino's (2007) meta-analysis of multiple studies found that *"Maximum, minimum, [and] peak frequency and frequency range [are] found to be significantly lower in closed compared with open habitats"*. The temporal structure of bird songs also correlates with habitat: for example, Badyaev and Leaf (1997) found that among a group of warblers *"species occupying closed habitats avoided the use of rapidly modulated signals and had song structures that minimized reverberation."* It is not so apparent that mammals and anurans typically display any such effect (Waser and Brown, 1986; Daniel and Blumstein, 1988; Ey and Fischer, 2009; Peters et al., 2009; Peters and Peters, 2010). This difference seems likely to be due to the fact that bird song is often much more structured, sequentially complex and varied in pitch than the calls of many mammals and anurans, and so has more features that could be disrupted in poor transmission conditions.

The overall thrust of the AAH is that in environments that are generally hostile to the faithful transmission of acoustic signals the nature of those signals will tend to become simpler in form. Importantly, since many of the factors that modify signals selectively impede transmission of higher frequencies

more than of lower ones, components of a signal that involve higher frequencies are the most likely to be simplified (e.g., Dabelsteen et al., 1993; Nemeth et al., 2001). It has been suggested that the AAH may also apply to human languages (Maddieson, 2012; Coupé, 2015; Maddieson and Coupé, 2015; Coupé and Maddieson, 2016). Suggestions that non-linguistic factors have relevance to language structure have a long history, but until recently the importance of the environmental transmission characteristics had not received much attention (but see Munroe et al., 1996, 2009; Munroe and Silander, 1999; Fought et al., 2004 on a connection between climate and language structure). Maddieson and Coupé (2015) found that both the number of consonants in a phonological inventory and the complexity of syllable onsets and codas are significantly correlated with mean annual temperature and precipitation as well as maximum tree cover in the areas where the languages are spoken. These factors are, naturally enough, correlated, as vegetation requires sun and water to thrive. For this reason a principal components analysis was performed to reduce the number of variables. Consonant inventory size and syllable complexity were also combined into a consonant-heaviness index. There is a highly significant relationship ($R^2 = 0.196$, $p < 0.0001$) between Principal Component 1 and the consonant-heaviness scores in a sample of 663 languages from the LAPSyD database (Maddieson et al., 2013) used by Maddieson and Coupé. Higher levels of consonant-heaviness broadly coincide with lower temperature, precipitation and tree cover (as well as with higher altitude and greater rugosity). This result is consistent with what is known about the effects of the environmental factors mentioned earlier. Consonants, especially obstruents, are more critically dependent on high frequency spectral components for their identification, and more complex syllable margins also lead to more rapid alternations of amplitude and spectral pattern. Hence it plausible that these properties would tend to be simplified where faithfulness of transmission is reduced.

However, this result was based on looking at the overall size of a consonant inventory and the maximal permitted length of syllable onsets and codas. Languages might have large inventories of obstruents and permit complex syllables but make only extremely rare use of these possibilities in the stream of speech. This paper presents a follow-up which examines if the proportion of obstruency vs. sonority in the speech stream in languages also correlates with environmental factors. Short spoken texts are compared using a sample of 100 + languages. The hypothesis under investigation is that in environments which impede faithful transmission, especially of higher frequencies, languages will favor a higher proportion of sonority. This will over time tend to differentiate the lexical forms of the words in languages spoken in environments which favor fidelity of transmission from those spoken in areas that impede faithful transmission of spectral and temporal complexity.

MATERIALS

The texts used in this study are drawn from the recordings available from the Global Recordings Network (GRN), an evangelical Christian organization that provides recordings of didactic religious materials intended to be used to spread a

particular sort of Christian faith via recordings made in the native languages of the target audiences. These recordings provide a very useful sample of a wide variety of languages in a relatively standard format. Many of the texts are re-tellings of stories from the Bible, both from Old and New Testament books. They usually involve a single speaker speaking at a moderately rapid rate, but some include more than one voice. More of the speakers are male than female. At some points sound effects and music may be also included, and some have accidental background noise or are of low quality, but a great many of the recordings are clear and have a very good signal to noise ratio. Most of the recordings in this collection are available for download in mp3 format, which sacrifices some fidelity to the quality of the original but is quite satisfactory for the present purposes, provided the original recording was done under good conditions.

There are some drawbacks to using these recordings, especially in that no details concerning the speakers are known. Some inferences concerning age and gender can be made based on the voices heard, but it is not known, for example, what other languages a given speaker may speak in addition to the target language, how much they use that language, or at what age they learned it. It is also evident that some of the recordings have been edited, particularly by truncating the signal at the onset and end of utterances. The nature of the subject matter also leads to a relatively high number of non-indigenous proper names of persons and places being used, e.g., Noah, Jesus, Adam. However, if there are "foreign accent" effects or other factors that make the recording a less than ideal exemplar of the language, these are considered as introducing statistical noise that would make it harder to confirm the hypothesis.

Each recording sample was divided into essentially sonorant and obstruent portions, as well as non-speech interludes. Sonorant and obstruent classifications were based on an auditory identification of the nature of the segments, coupled with close inspection of shape and amplitude changes in the waveform and of the spectral pattern. Files were examined using Praat (Boersma and Weenink, 2017). Vowels, voiced nasals, voiced central and lateral approximants and voiced rhotics were classed as sonorant. All stops, fricatives, and affricates as well as voiceless segments of other types are classed as obstruent. Bursts and any aspiration or affrication following a stop release as well as any preaspiration are included in the obstruent duration. The stop portion of a prenasalized stop or nasal + stop sequence was counted as obstruent, no matter how short, and the nasal portion as sonorant. As in any exercise to divide a continuous speech stream into discrete segments there are difficulties. The most acute issues concern delimitating onset and offset of segments at the margins of utterances. In most cases the articulatory onset of an utterance-initial stop is not apparent in the acoustic record, but since the hypothesis concerns the lexical shape of items an imputed articulation onset is assigned about 70 ms before a visible acoustic signature such as a burst; less if pre-voicing is apparent before the consonant release. At pre-pausal boundaries there is often an extended duration in which speech fades off into non-speech, often with devoicing, especially when the final segment is vocalic, although glottal constriction may also occur in such positions. Decisions as to the end of utterances were mainly

based on where the auditory impression of a specific segment identity was lost. On occasion, it was difficult to decide if there was final devoicing or glottalization of a vowel or the syllable was closed by a final /h/ or /ʔ/ segment. Again, if such decisions are made in error, this is likely to weaken the probability of the hypothesis being confirmed.

A short extract from the recording used for the Aleut language is shown in **Figure 1** to illustrate the procedure. The waveform and spectrogram (0–7 KHz) of a short (1.7 s) fragment are shown with two annotation tiers. The second of these shows the division into the obstruent (o), sonorant (s) and non-speech (n) intervals used to calculate the sonority score. The first tier shows a segmental transcription created for this exemplary figure based on the auditory identification of the segments heard. Segmental transcriptions were not regularly made; this annotation tier was normally only used to mark such things as a change of speaker or the presence of background noise or music. In this example, two issues in particular might be noted. The nasal in the sequence /ana/ in the middle of the sample appears to be pre-stopped, although this is not at all auditorily apparent. Since this is not a regular phenomenon in Aleut, unlike in, say, Eastern Arrernte, the prestopping is not considered as creating an obstruent interval. Secondly, the final /a/ is heavily glottalized and its end is indeterminate, although the auditory presence of an /a/ segment is indisputable. The end-point chosen for this vowel is a compromise between minimal and maximal options.

For each of the language samples the durations of speech fragments in obstruent and sonorant categories were summed, and the percentage of the total speech duration that was sonorant calculated. The speech samples are quite brief, consisting on average of about 1 min of actual speech (mean 66.12 s, s.d. 14.1). The mean sonority score across the samples is 65.52% (s.d. 9.02), although the range is wide, from 89.64 to 41.15%). Scores were calculated for 103 languages, but note that three of the languages whose data is included in **Figure 2** below, Towa, Guarani and Southern Qiang, are not included in subsequent analyses as they could not be matched with reliable climatic and ecological data.

The sonority scores obtained for the language sample used correlate quite well with the consonant heaviness index for the same languages in Maddieson and Coupé (2015), as shown in **Figure 2**. This correlation is highly significant ($R^2 = 0.232$, $p < 0.0001$), which indicates that the static measures of size of consonant inventory and syllable complexity predict a good part of the variance in sonority in continuous spoken language samples.

The sample of languages analyzed in the present study was selected to include a diverse range of representatives from different geographical areas and language families, and to sample the full range of values on Principal Component 1 from the Maddieson and Coupé (2015) study. Languages spoken over smaller geographical areas were preferred to ones spoken over larger areas since climatic and environmental measures are more uniform over smaller areas. Because a somewhat limited number of the recordings targeted were of usable quality, a more carefully structured sample could not be constructed. The list of languages used is included in Appendix 1 in Supplementary Material.
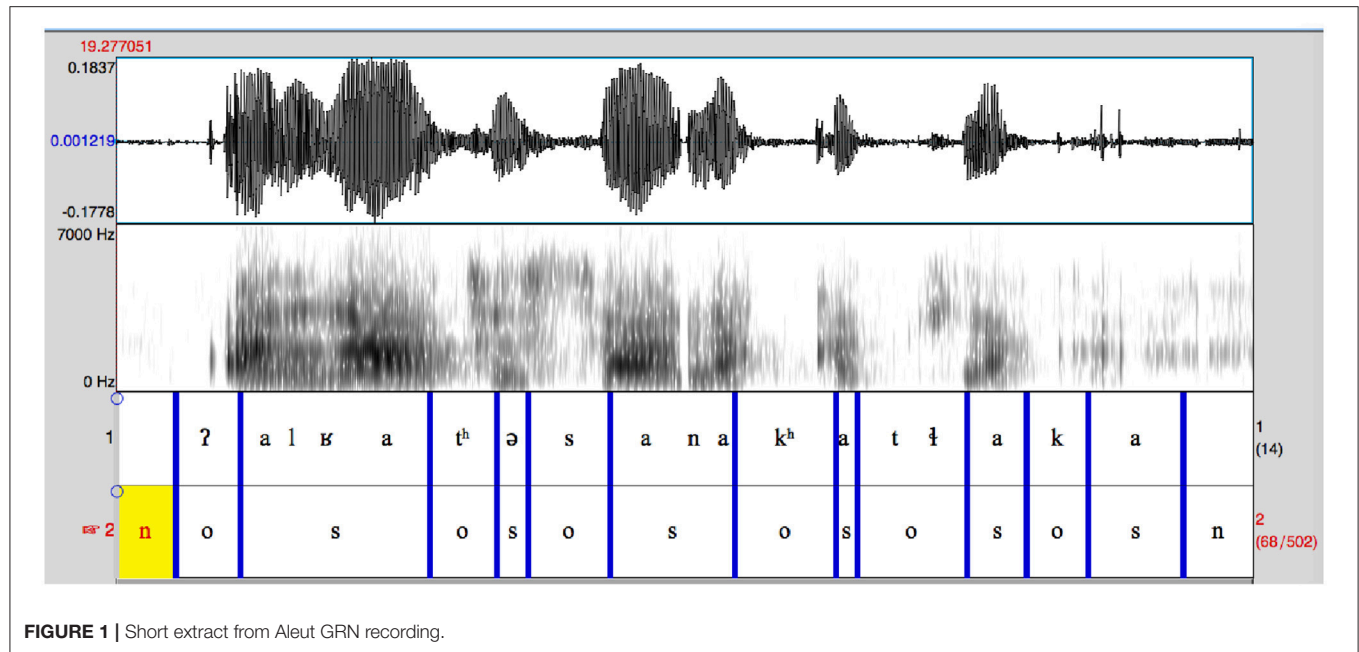
FIGURE 1 | Short extract from Aleut GRN recording.

For each language an estimate of the area where it is spoken was taken from the World Language Mapping System, a collaboration between Global Mapping International (2016) and SIL International which generates the language maps used in *The Ethnologue* (Simons and Fennig, 2017). This procedure requires forcing an alignment between languages as identified in *The Ethnologue* and those recognized by the Global Recordings Network. Inevitably, there are some discrepancies in this match, as well as with languages as represented by the descriptions included in LAPSyD. For each language area the mean values were computed for Percent Tree Cover and Elevation from values reported in 15-s bins by the Geospatial Information Authority of Japan (http://www.gsi.go.jp/kankyochiri/gm_global_e.html). Mean Temperature data in 5 s bins is from the Climate Research Unit of the University of East Anglia (available at http://www.ipcc-data.org/observ/clim/get_30yr_means.html, see New et al., 1999 for methodology) and covers the period 1961–1990. Other ecological and climatic data was obtained from the International Steering Committee for Global Mapping (http://www.iscgm.org) (disbanded in March 2017) and the UN Food and Agriculture Organization's Sustainable Development Department.

RESULTS

The salient result of this research is that the proportion of a speech sample that is sonorant in a sample of 100 languages is significantly correlated with mean annual temperature, but to a small or negligible extent with the other factors that were found to be related to consonant-heaviness in Maddieson and Coupé (2015). The significance values of simple correlations with single factors are shown in **Table 1**.

When these factors are entered together into a stepwise multiple correlation analysis only temperature is retained as
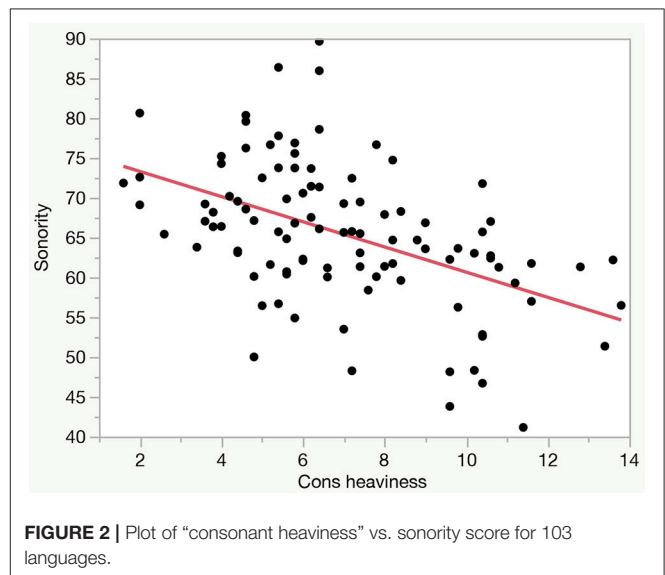
FIGURE 2 | Plot of "consonant heaviness" vs. sonority score for 103 languages.

making a significant contribution ($R^2 = 0.242$, $p < 0.0001$, after elimination of the other variables). In other words, although rugosity and elevation considered individually appear as significantly correlated with sonority in **Table 1**, this relationship disappears when factors are considered jointly—no doubt because of the well-known relationship between temperature and elevation and the fact that elevation and rugosity (roughness of terrain) are highly correlated with each other.

The linear relationship between sonority score and mean annual temperature (shown on a normalized scale reflecting deviations from global mean) for the 100 language sample is plotted in **Figure 3**.

TABLE 1 | Correlations between sonority score and climatic and environmental factors.

Factor	Significance
Temperature	<0.0001
Rugosity	0.0186
Elevation	0.0488
Precipitation	0.1618
Tree cover	0.8253

FIGURE 3 | Plot of sonority score vs. temperature for 100 languages.
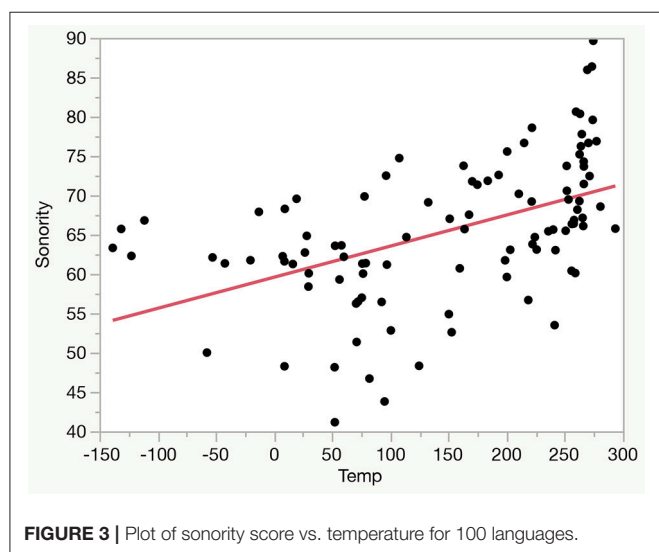
As seen in **Figure 3** there are notable deviations from the general trend, and the present data is probably best regarded as still exploratory in nature. A set of speech samples of longer duration from a larger sample of languages would represent a better test of the robustness of this relationship, and more nuanced temperature data might also be informative. However, there is a strong suggestion that languages habitually spoken in parts of the world that are hotter are more likely to have a more sonorous structure than languages spoken in cooler climates.

A standard objection to claims of any external influence on language structure is that the differences said to be associated with the external influence are simply inherited differences from ancestor states. That is, they can be explained by membership in different language families. In the present case, this is difficult to refute. The 100 language sample used includes languages from 49 different highest-level family classifications. When these 49 family affiliations are included as individual predictors of sonority, it is not surprising that the overwhelming majority of the variance can be associated with individual family affiliation since there are so many parameters present in the statistical model. However the effect of temperature remains significant ($p = 0.0203$) when language family is included as a random effect in a mixed-effects model. But related languages tend to be spoken in contiguous areas, and are therefore more likely to be spoken under somewhat similar environmental conditions. This can be

seen in **Figure 4** which plots sonority and temperature for the 7 families from which 5 or more languages are included in the sample. The left panel shows that languages from the same family tend to have somewhat similar sonority scores, with, for example, Altaic and Indo-European below the average and Australian, Niger-Congo and Trans-New Guinea above. The right panel plots the mean annual temperature for the same languages. A similar pattern emerges, with Altaic and Indo-European below the average and Australian, Niger-Congo and Trans-New Guinea above. Austronesian and Sino-Tibetan straddle the means. In this subset of data, sonority and temperature values are quite highly correlated, $R^2 = 0.4097$. While inherited aspects of the segment inventories and syllabic structures undoubtedly account for some of the similarity in sonority scores within families, it cannot be argued that mean temperature is a heritable linguistic trait. Thus perhaps the question should be to what extent might within-family similarities themselves be accounted for (at least in part) by environmental conditions.

DISCUSSION

Why would higher average temperature lead to the use of more sonorous sounds? There are various factors at play. First is the fact that atmospheric absorption increases at higher temperatures and it peaks at higher frequencies as the temperature increases (Harris, 1966). This will perturb the fidelity of transmission of frequencies higher in the speech range more than those in a lower range. In addition there is the impact of the turbulence in the air that is associated with higher temperature. Under some conditions heat-induced air turbulence can be seen by the naked eye as a disturbance to the visualization of objects at a distance (though bending of light rays also contributes to this visual effect). Studying the effects of atmospheric turbulence is problematical, since by its very nature turbulence is random, and moreover these effects can never be isolated in practice from other effects, such as ground reflectivity and atmospheric absorption. However, Daigle et al. (1986, p. 622) do suggest that under the experimental outdoor conditions they studied "the dominant mechanism responsible for the measured sound-pressure levels at high frequencies is scattering by atmospheric turbulence" and that these higher frequencies could be attenuated by as much as 20 dB from the source strength (cf Daigle et al., 1983). Ingård (1953) also reported strong attenuation of higher frequencies due to wind turbulence based on earlier studies. Turbulence also disrupts the temporal pattern of acoustic signals, particularly disrupting the integrity of rapidly changing signals. Selective effects of absorption and turbulence on higher frequencies naturally cause more problems for the faithful identification of speech components whose recognition depends on these higher frequencies, perhaps most especially for the burst spectra of consonants and the noise of sibilant fricatives. Sonorants on the other hand are more typically identifiable from lower-frequency elements, and have more slowly-changing temporal structure, and hence are less distorted by these factors.

In addition to these effects refraction due to temperature gradients may also play a role. Under normal daytime conditions,
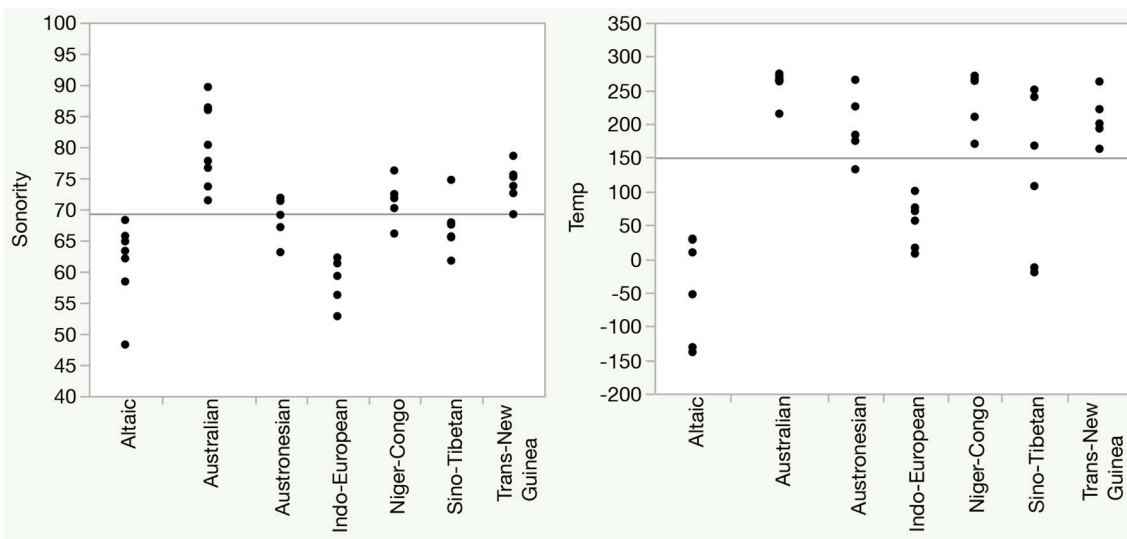
FIGURE 4 | Sonority scores **(Left)** and normalized mean annual temperature (relative to global mean), **(Right)** for languages in families with 5 or more languages in the sample.

there is a negative temperature gradient in the atmosphere—air nearer the ground is warmer than that higher up (e.g., Fowells, 1948). This causes an upward refraction of sound waves since the speed of sound is higher in warmer air (e.g., Lamancusa, 2010). Further, in general the temperature gradient ("lapse rate") is greater when ground temperature is higher, for example closer to the tropics (Mokhov and Aperov, 2006). The consequence of this is that overall sound energy is decreased more with distance. The normal daytime temperature gradient therefore generally diminishes the strength of a close-to-ground signal and degrades its perceptibility, but the more so the higher the temperature is, rendering accurate signal recognition more difficult.

As for the process by which such environmental effects shape the structure of languages, this is probably best regarded as a case where the role of the listener is paramount (Ohala, 1981, 2012). If the transmission conditions make it difficult to distinguish between different consonants, and different clusters of consonants, then the templates for given lexical items will likely converge on fewer distinct forms, because with sufficient exposure to tokens degraded during transmission a listener no longer considers them distinct. Over time, this will tend to restructure the phonological shape of words toward having smaller consonant inventories and simpler syllable structures. Naturally, this process is more likely to shape linguistic structure where speakers spend significant time outdoors. The period of human history during which a settled agricultural lifestyle was the predominant economic model—well after the "Neolithic Revolution" (Childe, 1936; Diamond and Bellwood, 2003) but before the Industrial Revolution had run its course—seems the most favorable time-frame within which the process would have impacted the shape of languages. In many cases a simple agricultural economy involves long hours of outdoor labor, tending crops and animals. In 1996 Munroe et al. (Munroe

et al., 1996; cf Ember and Ember, 1999) had suggested that more outdoor time was linked to simpler syllable structure, but did not link this in an explanatory way to environmental conditions. This paper presents a reasoned argument to support their speculation.

This paper also argues that acoustic adaptation occurs between different groups of the same species, in this case speakers of different human languages, whereas the majority of work on the AAH has examined between-species differences. However within-species effects are not unique. A number of studies of bird species that live in varied habitats have reported that their song patterns vary according to their environment in a similar way to that found across species. Hunter and Krebs (1979) examined songs of great tit (*parus major*) populations in widely dispersed sites from Morocco and Iran to Spain, Norway and the U. K. and found that birds inhabiting denser forest environments had songs with a lower maximum frequency, narrower frequency range and fewer notes per phrase than birds inhabiting more open woodland or hedgerows. Nicholls and Goldizen (2006) studied satin bowerbird (*Ptilonorhynchus violaceus*) populations along the east coast of Queensland, Australia, and found significant effects of variation in local habitat on song structure: "*Lower frequencies and less frequency modulation were utilized in denser habitats such as rainforest, and higher frequencies and more frequency modulation were used in the more open eucalypt dominated habitats.*" Within-species effects have also been reported, inter alia, by Wasserman (1979), Anderson and Connor (1985), and Tubaro and Segura (1994). These studies, like most studies addressing the AAH, have emphasized the physical characteristics of the environment, such as the vegetation, rather than looking at climatic factors. It would be interesting to see if adding analysis of factors such as temperature and precipitation would add to the insights derived by looking primarily at the characteristics of local vegetation

types in accounting for these differences. Note that global relative mean temperature patterns are likely to be more stable over recent time than tree cover, which is strongly affected by human activity as well as climatic change.

The finding that the design of acoustic communication systems within species appears to be shaped by environmental factors indicates that these influences operate over at least a shorter time-span than the interval between "speciation events" (Mayr, 1942), but this is, of course, a highly variable and imprecise datum. On the other hand, the phonological structure of human languages is highly malleable and individual languages can change their systems in the span of a single generation (e.g., Jacewicz et al., 2011). So environmental transmission factors affecting language structures, like other triggers of language change, probably do not require a long time span to operate. However, once entrenched, the consequences of such effects may persist for a long time.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2018.00028/full#supplementary-material

REFERENCES

Albert, D. G. (2004). *Past Research on Sound Propagation Through Forests*. Hanover NH: US Army Corps of Engineers, Cold Regions Research and Engineering Laboratory.

Anderson, M. E., and Connor, R. N. (1985). Northern Cardinal song in three forest habitats in Eastern Texas. *Wilson Bull.* 97, 436–449.

Attenborough, K., Bashir, I., and Taherzadeh, S. (2011). Outdoor ground impedance models. *J. Acoust. Soc. Am.* 129, 2806–2819. doi: 10.1121/1.3569740

Attenborough, K., Taherzadeh, S., Bass, H. E., Di, X., Raspet, R., Becker, G. R., et al. (1995). Benchmark cases for outdoor sound propagation models. *J. Acoust. Soc. Am.* 97, 173–191. doi: 10.1121/1.412302

Aylor, D. (1972). Noise reduction by vegetation and ground. *J. Acoust. Soc. Am.* 51, 197–205. doi: 10.1121/1.1912830

Badyaev, A. V., and Leaf, E. S. (1997). Habitat associations of song characteristics in Phylloscopus and Hippolais warblers. *Auk* 114, 40–46. doi: 10.2307/4089063

Bass, H. E., Sutherland, L. C., and Piercy, J. (1984). "Absorption of sound by the atmosphere," in *Physical Acoustics: Principles and Methods*, Vol. 17, eds W. P. Mason and R. N. Thurston (New York, NY: Academic Press), 145–232.

Beck, S. D., Nakasone, H., and Marr, K. W. (2011). Variations in recorded acoustic gunshot waveforms generated by small firearms. *J. Acoust. Soc. Am.* 129, 1748–1759 doi: 10.1121/1.3557045

Boersma, P., and Weenink, D. (2017). *Praat: Doing Phonetics by Computer*. Available online at: http://www.fon.hum.uva.nl/praat/

Boncoraglio, G., and Saino, N. (2007). Habitat structure and the evolution of bird song: a meta-analysis of the evidence for the acoustic adaptation hypothesis. *Funct. Ecol.* 21, 134–142. doi: 10.1111/j.1365-2435.2006.01207.x

Bradbury, J. W., and Vehrencamp, S. L. (1988). *Principles of Animal Communication*. Sunderland: Sinauer.

Brumm, H., and Slabbekoorn, H. (2005). Acoustic communication in noise. *Adv. Study Behav.* 35, 151–209. doi: 10.1016/S0065-3454(05)35004-2

Chappuis, C. (1971). Un exemple de l'influence du milieu sur les emissions vocales des oiseaux: l'evolution des chants en foret equitoriale. *Terre et Vie* 25, 183–202.

Childe, V. G. (1936). *Man Makes Himself*. London: Watts.

Coupé, C. (2015). "Testing the acoustic adaptation hypothesis with GIS and spatial regression models," in *Causality in the Language Sciences Meeting, Max Planck Institute for Evolutionary Anthropology* (Leipzig).

Coupé, C., and Maddieson, I. (2016). "Quelle adaptation acoustique pour les langues du monde?" in *Actes du 13ème Congrès Français d'Acoustique* (Le Mans).

Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (1993). Habitat-induced degradation of sound signals: quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song. *J. Acoust. Soc. Am.* 93, 2206–2220. doi: 10.1121/1.406682

Daigle, G. A., Embleton, T. F. W., and Piercy, J. E. (1986). Propagation of sound in the presence of gradients and turbulence near the ground. *Acoust. Soc. Am.* 79, 613–627. doi: 10.1121/1.393451

Daigle, G. A., Piercy, J. E., and Embleton, T. F. W. (1983). Line-of-sight propagation through atmospheric turbulence near the ground. *J. Acoust. Soc. Am.* 74, 1505–1513. doi: 10.1121/1.390152

Daniel, J. C., and Blumstein, D. T. (1988). A test of the acoustic adaptation hypothesis in four species of marmots. *Anim. Behav.* 56, 1517–1528. doi: 10.1006/anbe.1998.0929

Diamond, J., and Bellwood, P. (2003). Farmers and their languages: the first expansions. *Science* 300, 597–603. doi: 10.1126/science.1078208

Ember, M., and Ember, C. R. (1999). Cross-language predictors of consonant-vowel syllables. *Am. Anthropol.* 101, 730–742. doi: 10.1525/aa.1999.101.4.730

Embleton, T. F. W. (1996). Tutorial on sound propagation outdoors. *J. Acoust. Soc. Am.* 100, 31–48. doi: 10.1121/1.415879

Ey, E., and Fischer, J. (2009). The "Acoustic Adaptation Hypothesis" — a review of the evidence from birds, anurans and mammals. *Bioacoustics* 19, 1–2. doi: 10.1080/09524622.2009.9753613

Farina, A., Lattanzi, E., Malavasi, R., Pieretti, N., and Piccioli, L. (2011). Avian soundscapes and cognitive landscapes: theory, application and ecological perspectives. *Landsc. Ecol.* 26, 1257–1267. doi: 10.1007/s10980-011-9617-z

Fought, J. G., Munroe, R. L., Fought, C. R., and Good, E. M. (2004). Sonority and climate in a world sample of languages. *Cross Cult. Res.* 38, 27–51. doi: 10.1177/1069397103259439

Fowells, H. A. (1948). *The Temperature Gradient in a Forest*. California Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture. Available online at: https://www.fs.fed.us/psw/publications/cfres/cfres_1948_fowells001.pdf

Global Mapping International (2016). *World Language Mapping System*, Version 17. Available online at: http://www.worldgeodatasets.com/files/9313/0400/5767/WGDS_WLMS.pdf

Harris, C. M. (1966). Absorption of sound in air versus humidity and temperature. *J. Acoust. Soc. Am.* 40, 148–159. doi: 10.1121/1.1910031

Harris, C. M. (1967). *Absorption of Sound in Air Versus Humidity and Temperature*. Washington, DC: National Aeronautics and Space Administration.

Hauser, M. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hunter, M. L., and Krebs, J. R. (1979). Geographical variation in the song of the Great Tit (Parus major) in relation to ecological factors. *J. Anim. Ecol.* 48, 759–785. doi: 10.2307/4194

Ingård, U. (1953). A review of the influence of meteorological conditions on sound propagation. *J. Acoust. Soc. Am.* 25, 405–411. doi: 10.1121/1.1907055

Jacewicz, E., Fox, R. A., and Salmons, J. (2011). Vowel change across three age groups of speakers in three regional varieties of American English. *J. Phonol.* 39, 683–693. doi: 10.1016/j.wocn.2011.07.003

Krause, B. (1987). Bioacoustics, habitat ambience in ecological balance. *Whole Earth Rev.* 57, 14–18.

Krause, B. (1993). The niche hypothesis: a hidden symphony of animal sounds, the origins of musical expression and the health of habitats. *Explor. J.* 71, 156–160.

Lamancusa, J. S. (2010) *Outdoor Sound Propagation.* State College, PA: State University. Available online at: http://www.mne.psu.edu/lamancusa/me458/10_osp.pdf (Accessed March 30, 2018).

Maddieson, I. (2012). On the origin and distribution of complexity in phonological structure. Grenoble: Journées d'Études de la Parole. Available online at: http://aclweb.org/anthology/F/F12/F12-4004.pdf

Maddieson, I., and Coupé, C. (2015). "Human language diversity and the acoustic adaptation hypothesis," in *Proceedings of Meetings on Acoustics,* (Jacksonville, FL).

Maddieson, I., Flavier, S., Marsico, E., Coupé, C., and Pellegrino, F. (2013). "LAPSyD: Lyon-albuquerque phonological systems database," in *Proceedings of Interspeech 2013* (Lyon).

Marten, K., and Marler, P. (1977). Sound transmission and its significance for animal vocalization. I. Temperate habitats. *Behav. Ecol. Sociobiol.* 2, 271–290. doi: 10.1007/BF00299740

Marten, K., Quine, D., and Marler, P. (1977). Sound transmission and its significance for animal vocalization.II. Tropical forest habitats. *Behav. Ecol. Sociobiol.* 2, 291–302. doi: 10.1007/BF00299741

Martens, M. J. M. (1992). "Sound propagation in the natural environment, animal acoustic communication and possible impact for pollination," in *Sexual Plant Reproduction,* eds M. Cresti, and A. Tiezzi (Berlin: Springer), 225–232.

Martens, M. J. M., and Michelsen, A. (1981). Absorption of acoustic energy by plant leaves. *J. Acoust. Soc. Am.* 69, 303–306. doi: 10.1121/1.385313

Mayr, E. (1942). *Systematics and the Origin of Species from the Viewpoint of a Zoologist.* New York, NY: Columbia University Press.

Mokhov, I. I., and Aperov, M. G. (2006). Tropospheric lapse rate and its relation to surface temperature from reanalysis data. *Izvest. Atmosph. Oceanic Phys.* 42, 430–438. doi: 10.1134/S0001433806040037

Morton, E. S. (1975). Ecological sources of selection on avian sounds. *Am. Nat.* 109, 17–34. doi: 10.1086/282971

Munroe, R. L., Fought, J. G., and Macaulay, R. K. S. (2009). Warm climates and sonority classes: not simply more vowels and fewer consonants. *Cross Cult. Res.* 43, 123–133. doi: 10.1177/1069397109331485

Munroe, R. L., Munroe, R. H., and Winters, S. (1996). Cross-cultural correlates of the consonant-vowel syllable. *Cross Cult. Res.* 30, 60–83. doi: 10.1177/106939719603000103

Munroe, R. L., and Silander, M. (1999). Climate and the consonant-vowel (CV) syllable: replication within language families. *Cross Cult. Res.* 33, 43–62. doi: 10.1177/106939719903300104

Naguib, M. (2003). Reverberation of rapid and slow trills: implications for signal adaptations to long-range communication. *J. Acoust. Soc. Am.* 113, 1749–1756. doi: 10.1121/1.1539050

Nemeth, E., Winkler, H., and Dabelsteen, T. (2001). Differential degradation of antbird songs in a neotropical rainforest: adaptation to perch height? *J. Acoust. Soc. Am.* 110, 3263–3274. doi: 10.1121/1.1420385

New, M., Hulme, M., and Jones, P. (1999). Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Clim.* 12, 829–856. doi: 10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2

Nicholls, J. A., and Goldizen, A. W. (2006). Habitat type and density influence vocal signal design in satin bowerbirds. *J. Anim. Ecol.* 75, 549–558. doi: 10.1111/j.1365-2656.2006.01075.x

Ohala, J. J. (1981). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior* (Chicago: Chicago Linguistic Society), 178–203.

Ohala, J. J. (2012). "The listener as a source of sound change: an update," in *The Initiation of Sound Change: Perception, Production, and Social Factors,* eds M. J. Solé, and D. Recasens (Amsterdam: John Benjamins), 21–36.

Peters, G., Baum, L., Peters, M. K., and Tonkin-Leyhausen, B., (2009). Spectral characteristics of intense mew calls in cat species of the genus Felis (Mammalia: Carnivora: Felidae). *J. Ethol.* 27, 221–237. doi: 10.1007/s10164-008-0107-y

Peters, G., and Peters, M. K. (2010). Long-distance call evolution in the Felidae: effects of body weight, habitat, and phylogeny. *Biol. J. Linnean Soc.* 101, 487–500. doi: 10.1111/j.1095-8312.2010.01520.x

Piercy, J. E., Embleton, T. F., and Suthorland, L. G. (1977). Review of noise propagation in the atmosphere. *J. Acoust. Soc. Am.* 61, 1403–1418. doi: 10.1121/1.381455

Richards, D. G., and Wiley, R. H. (1980). Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *Am. Nat.* 115, 381–399. doi: 10.1086/283568

Römer, H. (2001). "Ecological constraints for sound communication: from grasshoppers to elephants," in *Ecology of Sensing*, eds F. G. Barth, and A. Schmid (Berlin: Springer), 59–78.

Ryan, M. J., and Kime, N. M. (2002). "Selection on long-distance acoustic signals," in *Acoustic Communication*, eds A. M. Simons, A. N. Popper, and R. R. Fay (New York, NY: Springer), 225–274.

Salomons, E. M. (2001). *Computational Atmospheric Acoustics.* Dordrecht: Kluwer.

Seddon, N. (2005). Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution* 59, 200–215. doi: 10.1111/j.0014-3820.2005.tb00906.x

Simons, G. F., and Fennig, C. D. (eds). (2017). *Ethnologue, Languages of the World, 20th Edn.* Dallas, TX: SIL International. Available online at http://www.ethnologue.com (requires subscription).

Slabbekoorn, H., and Peet, M. (2003). Birds sing at higher pitch in urban noise. *Nature* 424:267. doi: 10.1038/424267a

Slabbekoorn, H. (2004a). Habitat-dependent ambient noise: consistent spectral profiles in two African forest types. *J. Acoust. Soc. Am.* 116, 3727–3733. doi: 10.1121/1.1811121

Slabbekoorn, H. (2004b). "Singing in the wild: the ecology of birdsong," in *Nature's Music: The Science of Birdsong,* eds P. Marker, and H. W. Slabbekoorn (Amsterdam: Academic Press), 178–205.

Slabbekoorn, H., and Smith, T. B. (2002). Habitat-dependent song divergence in the Little Greenbul: an analysis of environmental selection pressures on acoustic signals. *Evolution* 56, 1849–1858. doi: 10.1111/j.0014-3820.2002.tb00199.x

Sutherland, L. C. and Daigle, G. A. (1998). "Atmospheric sound propagation," in *Handbook of Acoustics,* ed M. J. Crocker (New York, NY: John Wiley, and Sons), 305–329.

Tubaro, P. L., and Segura, E. T. (1994). Dialect differences in the song of zonotrichia capensis in the southern pampas: a test of the Acoustic Adaptation Hypothesis. *Condor* 96, 1084–1088. doi: 10.2307/1369117

Waser, P. M., and Brown, C. H. (1986). Habitat acoustics and primate communication. *Am. J. Primatol.* 10:135–154. doi: 10.1002/ajp.1350100205

Wasserman, F. E. (1979). The relationship between habitat and song in the White-throated Sparrow. *Condor* 81, 424–426. doi: 10.2307/1366974

Wiley, R. H., and Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: implications for the evolution of animal vocalizations. *Behav. Ecol. Sociobiol.* 3, 69–94. doi: 10.1007/BF00300047

Wilson, D. K., Brasseur, J. G., and Gilbert, K. E. (1999). Acoustic scattering and the spectrum of atmospheric turbulence. *J. Acoust. Soc. Am.* 105, 30–34. doi: 10.1121/1.424594

Winkler, H. (2001). "The ecology of avian acoustic signals," in *Ecology of Sensing,* eds F. G. Barth and A. Schmid (Berlin: Springer), 79–104.

Wood, W. E., and Yezerinac, S. M. (2006). Song sparrow (Melospiza melodia) song varies with urban noise. *Auk* 123, 650–659. doi: 10.1642/0004-8038(2006)123[650:SSMMSV]2.0.CO;2

Languages in Drier Climates Use Fewer Vowels

Caleb Everett *

Department of Anthropology, University of Miami, Coral Gables, FL, United States

This study offers evidence for an environmental effect on languages while relying on continuous linguistic and continuous ecological variables. Evidence is presented for a positive association between the typical ambient humidity of a language's native locale and that language's degree of reliance on vowels. The vowel-usage rates of over 4000 language varieties were obtained, and several methods were employed to test whether these usage rates are associated with ambient humidity. The results of these methods are generally consistent with the notion that reduced ambient humidity eventually yields a reduced reliance of languages on vowels, when compared to consonants. The analysis controls simultaneously for linguistic phylogeny and contact between languages. The results dovetail with previous work, based on binned data, suggesting that consonantal phonemes are more common in some ecologies. In addition to being based on continuous data and a larger data sample, however, these findings are tied to experimental research suggesting that dry air affects the behavior of the larynx by yielding increased phonatory effort. The results of this study are also consistent with previous work suggesting an interaction of aridity and tonality. The data presented here suggest that languages may evolve, like the communication systems of other species, in ways that are influenced subtly by ecological factors. It is stressed that more work is required, however, to explore this association and to establish a causal relationship between ambient air characteristics and the development of languages.

Keywords: phonetics, environment, adaptation, psychological, language, evolution

INTRODUCTION

The communication systems of many species are known to be ecologically adaptive, being impacted by factors such as humidity (Wilkins et al., 2013). Such adaptivity is not traditionally thought to characterize human speech, however. This position of linguistic "autonomy" has now been called into question, however, by studies pointing to potential environmental influences on speech sounds. (Munroe et al., 2009; Everett, 2013; Everett et al., 2015) Such studies have confronted strong objections, in part (and to varying degrees) because of their utilization of binning strategies through which linguistic and/or geographic variables were categorized. The present study avoids such binning and offers, via several analytical methods, evidence of an association between two continuous variables: ambient humidity and vowel utilization. I suggest that this association may be motivated by the influence of dry air on the vocal folds, though more research is required to show that this influence motivates the distribution described here and in other work on this topic (Everett et al., 2015). The association between ambient humidity and vowel utilization is, we will see, unlikely due to potentially confounding factors like linguistic phylogeny or contact between

languages. While I avoid strong claims of causality, I conclude that the uncovered association merits further inquiry. In other words, the association does not demonstrate that languages adapt to ecological factors, it simply suggests that idea deserves continued consideration.

Languages with complex tonality are apparently less likely to develop in desiccated regions. In previous work, colleagues and I have argued that this distributional pattern is possibly due to subtle diachronic pressures resulting from the heightened difficulty of maintaining precise pitch when vocal folds are consistently exposed to desiccated air (Leydon et al., 2009; Everett et al., 2015). Laryngology studies do suggest that phonation/voicing is affected by dry air. In one study it was observed that the effects of dry air include increased jitter rates that may impact the production of precise pitch (Hemler et al., 1997). Yet it is still debated whether such minor effects on jitter rates actually impact pitch production in normal speech, and whether languages with complex tone really do rely on more precise pitch patterns in the speech stream (De Boer, 2016; Everett et al., 2016a,b). What is less debatable is that research in laryngology has shown the desiccation of vocal cords leads to greater perceived phonatory effort on the part of speakers and that laryngeal desiccation impacts the viscoelasticity of the vocal folds. [See the survey of some relevant findings by Leydon et al. (2009).] For instance, in a recent experiment with elderly speakers, it was observed that increasing ambient humidity to moderate levels reduced the perceived phonatory effort and vocal tiredness reported by those speakers in a loud-reading task (Sundarrajan et al., 2017). The effects uncovered in such studies surface despite relatively limited exposure to desiccated ambient air, in contrast to populations living in very arid environs. The salutary effects of humidity on phonation could help explain the pervasive pattern reported here. Future work could explore this possible connection with other methods, including experimental ones.

In addition to the association between less tonality and aridity, other correlations between geography and phonemic inventories have been observed. These include the greater frequency of ejective sounds in high elevation regions (Everett, 2013) and the higher rate of consonant-vowel syllables in languages in warm regions (Ember and Ember, 2000, 2007; Fought et al., 2004; Munroe and Silander, 2009). It is suggested below that all these correlations are interrelated and, if causally motivated by the environment, may have one underlying motivator. Despite such associations, many language researchers remain skeptical of any meaningful relationships between ecological factors and human phonologies. Since language is transmitted socially, it is still unclear how ecological factors may come to influence phonetic patterns. In previous work colleagues and I have suggested tentative mechanisms through which some effects may surface, but the likelihood of these mechanisms is admittedly open to debate (Everett et al., 2016a,b). Despite such debate, some scholars now seem open to the possibility that languages are impacted by ecology. The more general suggestion that language-external factors impact language is evidenced in other contemporary research as well, for instance in work showing a

negative correlation between population size and morphological complexity (Lupyan and Dale, 2010, 2016).

In short, the last decade has seen the publication of a variety of studies hinting that, contra traditional linguistic dogma, languages develop in ways that are sensitive to ecological pressures. Yet the relevant studies on linguistic sounds share a characteristic that some scholars have found problematic: the simple binning of languages by linguistic or ecological characteristics. [This is not true of all work examining external influences on language, however, see Lupyan and Dale (2010).] For instance, one recent study suggested that languages rely on consonants more in cold regions with less vegetation (Maddieson and Coupé, 2015). The observed correlation was found after languages were binned into categories such as "consonant-heavy," meaning that a language's ratio of consonant phonemes to vowel phonemes is high. Yet there is arguably no clear independent motivation for marking the divisions between the created categories in this and other studies. For instance, in Everett et al. (2015), colleagues and I categorized languages as having or not having "complex tonality" in order to facilitate the testing of a specific hypothesis. Yet, as we were aware, languages vary dramatically and non-discretely in the extent to which they rely on pitch for contrasting meaning, and on the extent to which precise pitch is used for other purposes. (Ladd, 2016) Such binning strategies are generally the result of methodological exigencies and the limitations of extant databases but, at least to some scholars, they minimize the inferences that can be drawn from the associations in all work so far undertaken on this topic. A similar observation may be made with respect to ecological factors, which have also been binned to facilitate the grouping of environments and test for geo-phonetic patterns. In some research, populations of speakers have been grouped as living in either "cold" or "warm/moderate" climates (Munroe et al., 1996). In a study of mine, languages were categorized dichotomously as being native to either high or low altitude regions, for a portion of the analysis (Everett, 2013). For another portion of the analysis, the altitudes of language locales were analyzed continuously, but some of the objections to the study have centered around the strategy employed for the discrete binning of language locales according to elevation regions. There is a concern that the observed correlations in such studies might have benefited from the placement of the category divisions (Dediu et al., 2017). One could debate whether this concern has been exaggerated, but it must be acknowledged that, to date, all studies on this particular topic have relied to varying degrees on the discontinuous grouping of language locales and/or language types into two or a few categories. The results of such studies face resistance, at least in part, because of this methodological tack. So a central aim of the present study is to test for a key ecological-linguistic correlation while relying entirely on continuous data.

Despite this shared methodological tack, the recent studies on this topic certainly intimate that human language may be ecologically adaptive. Alternate explanations for most of the associations are still missing, beyond pointing to the well-known existence of spurious correlations. Yet it is also well-known that the uncovering of correlations is a key tool in the scientific

arsenal, as they frequently point to relationships that merit further inquiry. Since languages exhibit a bias toward ease of articulation and some sound patterns may be more easily (even slightly) produced under certain ambient conditions, some language researchers now seem open to considering the possibility of ecological influences on speech sounds. It would appear that a clearer understanding of this issue is a desideratum for the language sciences (Evans, 2016; Greenhill, 2016). In an effort to contribute to that understanding, this study explores the association of human sound systems and their ecologies. It is the first to do so without relying on binning strategies. Languages are not grouped according to any phonetic/phonological categories, and ecologies are not discretely categorized either. Additionally, the study relies on the largest data set so far considered in such work. Analysis of that data set demonstrates that vowels are relatively less frequent (when contrasted to consonants) in languages in dry regions. This pattern appears consistent with laryngology evidence suggesting that phonation/voicing threshold pressure and perceived phonatory effort are heightened by inhaled dry air and the superficial dehydration of vocal cords (Sivasankar and Fisher, 2002; Leydon et al., 2009). It is also consistent with the recent finding that perceived phonatory effort and perceived tiredness are mitigated, amongst elderly speakers, when ambient humidity is increased (Sundarrajan et al., 2017). Since languages are biased toward less articulatory effort (Napoli, 2014), it is at least possible that they could be impacted by the heightened laryngeal challenges associated with the inhalation (especially oral inhalation) of dry ambient air. This possibility is difficult to evaluate conclusively given the many complex factors at work in language change, but I argue that it nevertheless merits further investigation.

The linguistic variable investigated here, the rate at which languages rely on vowels compared to consonants, was selected for four reasons. First, it has been suggested for some time now that languages in colder regions rely less on vowels. This suggestion was initially based on binned data with small samples, or without controlling for Galton's problem (Munroe et al., 1996; Maddieson et al., 2011). So the findings presented here relate to previous results, but address the issue with a novel approach and larger data set. Second, the linguistic variable investigated here relates to all spoken human languages. All rely on vowels, sonorant voiced sounds produced with oral aperture. Some of the previous work on this topic has focused on linguistic phenomena like complex tone and ejectives that, while not rare, only occur in a subset of the world's languages. Third, the variable considered here has a clearer potential connection to experimental work in laryngology. One issue with investigations of ecological adaptation in speech is that myriad *post hoc* explanations for uncovered correlations may be possible. [See the debate in Ember and Ember (2000) and Munroe et al. (2000).] So the linguistic variable selected should have some connection to prior nonlinguistic research. In the aforementioned study on tonality, we connected tonality to findings in laryngology suggesting that complex pitch production may be more difficult to achieve in arid regions. Yet, while little of the laryngology research relates to pitch (De Boer, 2016), a more common finding is that dry air increases perceived phonatory effort since vocal

cord usage becomes slightly more effortful after the inhalation of dry air (Erickson and Sivasankar, 2010). Such effects surface even after short exposures of the larynx to dry air. Of course the healthy human larynx is capable of achieving homeostasis and adapting to environmental pressures, so these effects may not be felt in all individuals equally and may only surface in minor ways during real-world speech situations. Yet even minor effects could potentially yield, over the long haul, functional pressures on speech. So a simple possibility exists: Languages in dry regions may exhibit a bias toward less vocal cord usage. Vowels require voicing and are the sounds that generally carry stress, which often requires greater amplitude of vocal cord vibration. Given such factors, it is worth considering whether there is a slight bias against the utilization of vowels in dry places. Since vowels are critical to the audibility of language, any vowel-reductive patterns would likely be minor. While many consonants are also voiced, many are not and consonants' degree of voicing (i.e., voice-onset-time) can vary substantially. The database relied on here does not encode voicing status for all consonants, creating further motivation for focusing on vowels. Nevertheless, some analysis of voicing in consonants is presented after the main analysis of vowels.

Fourth and finally, the linguistic variable used in this study was selected because it yields continuous data derived from transcriptions of actual words, as opposed to being derived from lists of phonemes. Phonemic inventories, which have been used in all previous studies on this topic, are actually not ideal bases for investigating potential ecological interactions of the sort being considered here. After all, they are only indirect representations of which sound patterns are most characteristic of a language, since any sound that is used in a semantically contrastive function, i.e., in a minimal pair, is included in a language's phonemic inventory regardless of the phone's frequency. So, for instance, if we were ascertaining the consonant-to-vowel (C:V) ratio of the English phonemic inventory, all consonants and vowels would carry the same weight even though it is well known that some sounds are much more common than others in speech. High C:V ratios demonstrate relative diversity of consonant types, rather than actual heightened reliance on consonants in speech. To get a sense of which sounds and sound patterns are actually most characteristic of English or another language, we have to have some way of determining the relative frequency of sounds. The specific linguistic variable introduced below, "vowel index," allows for such a determination. (In the "Discussion" section, I examine the relationship between vowel index and C:V ratio.)

The selection of the main ecological variable relied on here, specific humidity, is also well motivated. Specific humidity refers to the ratio of water in the air (See "Materials and Methods"). Epidemiological, laryngological, and anthropological studies have demonstrated that dry ambient air, particularly very dry air, has pervasive effects on the human body. These effects include increased prevalence of xerostomia, laryngitis and other vocal-tract maladies, the effects of reduced humidity on the evolution of cranial morphology, and the aforementioned effects on phonation (Sivasankar and Erickson-Levendoski, 2012; Maddux et al., 2016). The latter effects are exacerbated by oral
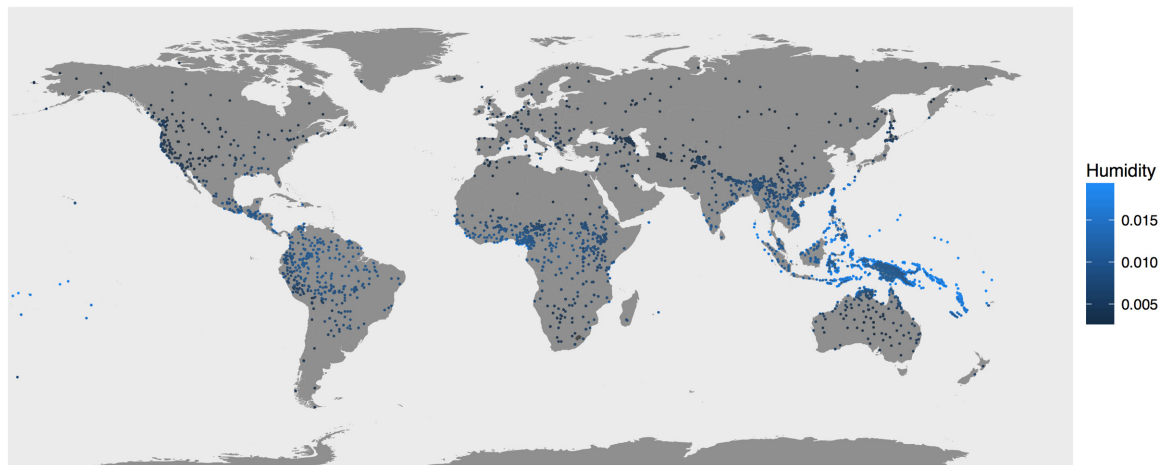
FIGURE 1 | Locations of the language varieties represented in the main analysis.

breathing that is promoted by nasal blockage that is more prevalent in dry and particularly cold-dry conditions (Mäkinen et al., 2009).

Using the continuous phonetic and ecological variables selected, the potential interaction of languages and ambient air was tested. This study relied on data from the Automated Similarity Judgment Program (Wichmann et al., 2016). The creators of the database employ a transcription system that conveys characters of the International Phonetic Alphabet with typewritten letters, with a sufficient degree of specificity as to count instances of consonants and vowels. At present the database contains phonetic transcriptions of 7221 word lists based on the analysis of many written sources including work by linguistic fieldworkers. Word lists for constructed languages and proto-languages were excluded from analysis. Some languages are represented by more than one list as multiple dialects are represented. Given the cline-like nature of the language/dialect distinction, I refer to the word lists as representing separate "language varieties" and rely extensively on methods that control for the over-representation of any language families. Each language variety is categorized according to its linguistic family as labeled in the WALS database (Dryer and Haspelmath, 2016). For each of the word lists the incidence of vowels, compared to the total transcribed vowels and consonants, was calculated. The calculation of vowels as a ratio of all sounds is referred to as the "vowel index" of each language variety. (Terms like "vowel ratio" or "consonant-to-vowel ratio" are avoided since they serve other functions in linguistics.) Word lists in the ASJP database have coordinates representing the locales to which their represented languages are thought to be native. These locales, even if not exact representations of ancestral homelands, approximate the appropriate regions. For the relatively few widespread languages, the coordinates denote the locale thought to be associated with the language's development, e.g., southeast England for English. For 4012 language varieties in the database, specific humidity rates were obtained for their presumed locales, by cross-referencing the

varieties with the humidity data in Everett et al. (2015). Of course, populations of speakers do not reside in the exact same location long-term, and climatic patterns also change over the long-term (Moran, 2016). Still, the mobility of most populations is relatively confined geographically (and cultures are typically well-adapted to particular environs) and certain climatic patterns hold regardless of weather cycles. Equatorial regions tend to be hotter and more humid, high elevations and deserts are arid, and so forth, regardless of climatological undulations. Furthermore, most major geographic factors, e.g., the Sahara and Amazonia, existed long before the languages on which this analysis is based. In short, relying on such climatological data seems the best approach available to explore this issue. Mean annual temperature data were ascertained for 6901 language varieties. Temperature is an (imperfect) proxy for specific humidity because air at colder temperatures can "hold" less water (Maddux et al., 2016). Basic tests were conducted on the larger word-list set with temperature data. Their results, some presented in the "Materials and Methods" section, are consistent with the findings discussed next, though generally less robust. This suggests that any associations with temperature may be epiphenomenal. The 4012 language varieties used in the main analysis, along with their locales' associated humidity values, are presented on the map in **Figure 1**.

The transcriptions utilized represent 40 words denoting basic semantic concepts. These include body parts, pronouns, common animals, frequent actions, as well as natural entities like 'water' and 'sun.' Some lists in the database have more than 40 words or a few less. The word lists are excellent data for testing any potential interactions with non-linguistic variables, since the words are not readily susceptible (though not immune) to contact-based effects. Some of these words are amongst the most common words in language, making them good indicators of how languages rely on particular sounds (Calude and Pagel, 2011). Recall that phonemic inventories, on which most studies of this topic have relied, do not actually offer information about the relative frequency of sounds in a given language.

RESULTS

For the 4012 word lists used in the core analysis, vowel indices ranged from 0.230 to 0.647, with a median of 0.458. The median for the larger set of 6901 lists was also 0.458. Languages aggregate around a rate of one vowel for every consonant. While it is known that CV syllables are quite common in languages (Maddieson, 2013), these figures represent, to my knowledge, the first quantification of the relative frequency of vowels and consonants across a major cross section of the world's languages. Some previous work has examined the relative frequency of vocalic and consonantal *phonemes* in much smaller sets of languages. For instance, Yegerlehner and Voegelin (1957) found, via text analysis, that the median ratio of vocalic phonemes across nine languages was 0.475. Such phoneme-based ratios evident in texts are similar to the vowel indices obtained here, though vowel indices are phonetically rather than phonologically based. Consider the two most extreme cases in Yegerlehner and Voegelin: The highest vowel ratio in their set was obtained for Maori, at 0.587. The vowel index obtained here for Maori is 0.559. The lowest vowel ratio obtained in Yegerlehner and Voegelin's small sample was for Navajo, at 0.44. The vowel index obtained for Navajo with the present methods is 0.454. So this computationally based analysis of frequent words apparently yields similar results to visual inspections of phoneme counts in texts. In the data considered here, the four language varieties with the lowest vowel indices, ranging from 0.230 to 0.258, are all Salishan languages of the Pacific Northwest. These languages are known for their complex strings of consonants (Flemming et al., 2008). The language with the highest vowel index, 0.647, is the Amazonian isolate Pirahã that is also known to exhibit some unusual phonetic characteristics (Everett, 2005). With respect to the humidity data, the median ratio for specific humidity is 0.0162, with a minimum value of 0.0025 and a maximum of eight times that, at 0.020. This marked disparity is a reminder that, while all people live at the bottom of the same ocean of air, many reside in different seas. Everett et al. (2016a) with respect to temperature, the median annual temperature is 24.2 Celsius. In **Figure 2** the 4012 languages are plotted according to humidity and vowel index.

A simple linear regression for vowel index and humidity reveals an interaction ($R^2 = 0.159$, $p = 0.000$), as evidenced by the positive slope in **Figure 2**. Since vowel indices are proportion data technically bounded at 0 and 1, simple linear regression is not the best approach. A more suitable test is beta regression, or regression of logit-transformed vowel indices. This discussion focuses on results for beta regressions, though remarkably similar results obtain for all three sorts of tests. (The proportion data in this case, while technically bounded at 0 and 1, actually occupy a fairly narrow portion of that range). In the case of the global distribution evident in **Figure 2**, a beta regression also reveals a significant interaction between humidity and vowel index (pseudo $R^2 = 0.158$, $p = 0.000$). Nevertheless, the trend in **Figure 2** could be the result of confounds like the preponderance of particular language families in dry regions. One useful approach to control for such confounds is to treat the language families (including isolates) as separate data points, so

that each family carries the same weight. The median and mean vowel index and humidity values of each family were ascertained. The means are plotted in **Figure 3**. Controlling for family in this way, the relationship between variables is actually noticeably strengthened. The beta regression for families' mean vowel indices and humidity values reveals a more striking association (pseudo $R^2 = 0.286$, $p = 0.000$). If we examine the relationship between median vowel index and median humidity values, so that the data points represent actual languages and locales, the heightened association remains (pseudo $R^2 = 0.267$, $p = 0.000$). One might object that such means and medians are misleading for large linguistic families spoken over diverse geographic regions. The most widespread families do not appreciably impact the results in **Figure 3**, however. The 229 families have an average size of 17.5 language varieties, but six families account for almost half of the total of 4012. Austronesian has 869 representative lists, Indo-European has 208, Afro-Asiatic has 185, Niger-Congo has 409, Trans-New-Guinea has 211, and Sino-Tibetan has 186. Also, 181 Australian varieties are grouped together in the database. Removing all these languages leaves us with 1763 word lists distributed across 222 families–7.9 per family, generally representing a restricted geographic region. When the mean vowel index and humidity values of only these 222 families are considered, the regression reveals the same interaction (pseudo $R^2 = 0.280$, $p = 0.000$). When the median vowel index and humidity values of these 222 families are examined, the same interaction is again observed (pseudo $R^2 = 0.263$, $p = 0.000$).

A multiple regression with the mean vowel index (logit-transformed) of each family as a dependent variable, and mean humidity, region (Eurasia, North America, South America, Africa, Australia, and the Pacific), and mean population (log-transformed) as independent variables, still reveals the interaction of humidity and vowel index ($p = 0.000$). Unsurprisingly, perhaps, no effect of population was observed. The interaction of regions and vowel indices is discussed further below.

As is evident in the regression lines in **Figure 3**, the interaction of vowel index and humidity surfaces within the four major landmasses with the greatest variances in climate. (The association does not surface within Australia, a point returned to below.) These landmasses were analyzed separately because of that variance, and because they are geographically rather than linguistically motivated. [See, e.g., Dryer (1989) for one take on the need for such geographic sampling.] For the 23 families of Africa (757 languages), mean familial vowel index and mean familial humidity are strongly associated (pseudo $R^2 = 0.343$, $p = 0.0005$). For the 31 families of Eurasia (934 languages), they are also strongly associated (pseudo $R^2 = 0.300$, $p = 0.0003$). For the 46 families of North America (299 languages), they are associated but not to the same degree (pseudo $R^2 = 0.117$, $p = 0.01$). For the 69 families of South America (407 languages), they are again strongly associated (pseudo $R^2 = 0.201$, $p = 0.00003$). If we consider instead the median vowel index and humidity values, the cross-family association remains evident within each region. It is again significant across all four regions, and once again is weakest in North America (See "Materials and Methods").
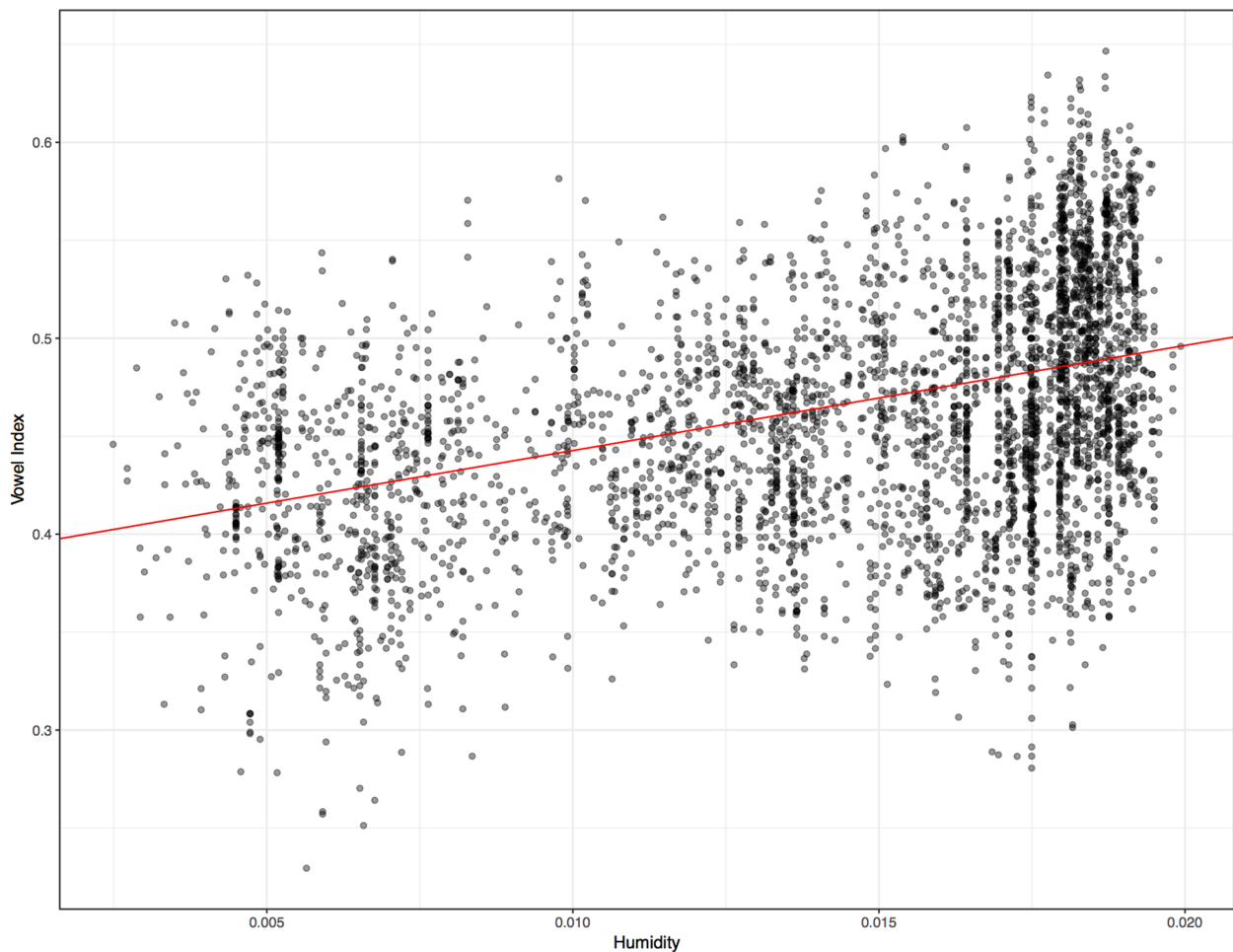
FIGURE 2 | Vowel index rates for the 4012 word lists, with associated specific humidity values. The line has a slope of 5.37 with a *y*-intercept of 0.389. For the simple regression: F-statistic = 756.2 on 1 and 4010 DF, $p = 0.000$. R^2: 0.159. A Box Cox transformation was applied to the data, to ensure homoscedasticity. The estimated lambda values was 0.9. With the fudge factor no transformations were required.

To further test the association without relying on medians or averaging, I adapted the method of random sampling used in previous work (Everett et al., 2015). This method also lends equal weight to each family. However, for this study I used random sampling at global and regional scales, so as to control for phylogeny and areal effects simultaneously. For each sample, one language per family was randomly selected and its vowel index and humidity were noted. Then the sample was analyzed with a regression contrasting vowel indices and humidity values. Both linear and beta regressions were used. (Basic linear regression was used simply to test for a positive or negative slope, for each regression.) One thousand regressions of each type were analyzed at the global level, each representing all 229 families. Critically, 1000 tests of each type were also analyzed for each of the four major landmasses with many families, thereby weighting families equally while simultaneously testing regions. The density distribution of slopes for the 5000 linear regressions, offered in **Figure 4**, shows that the pattern is evident within all major landmasses. Once again it shows itself to be weakest in North

America. Still, slopes were positive for all 5000 iterations of the simple regressions.

For the beta regressions, 1000 global tests revealed a clear interaction between humidity and vowel index. The mean pseudo R^2 value across all 1000 global iterations was 0.23. For Africa, the mean pseudo R^2 across 1000 iterations was 0.23. For Eurasia, the mean pseudo R^2 across the 1000 tests was 0.19. For South America, the average pseudo R^2 was 0.17. For North America, the average pseudo R^2 was 0.09. The association was once again found to be positive in all 5000 iterations of the test.

The association is evident not just within regions but is also evident across them. We can arrive at the mean vowel index and humidity value for each of the six major "regions" (the four major landmasses plus Australia and the Pacific) by averaging the means of each family in each region. If we then run a beta regression on the regions' average vowel indices and humidity values, controlling for family in this manner, we find a very striking association between vowel index and humidity (pseudo $R^2 = 0.917$, $p = 0.000$). (If we exclude the Pacific region, since
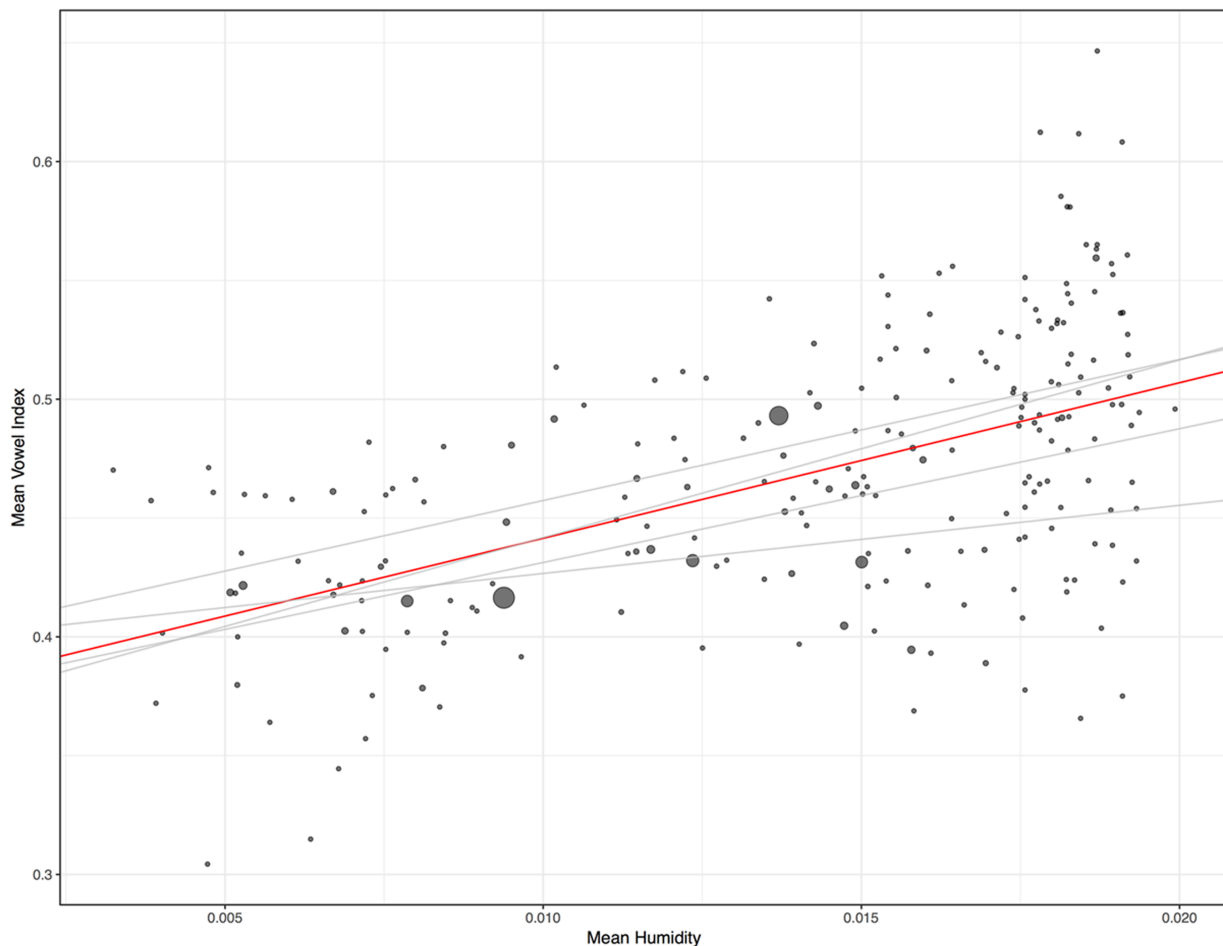
FIGURE 3 | Mean vowel index and mean humidity rates for 229 language families. Lines represent slopes of basic linear regressions. The global trend is represented in red. The red line has a slope of 6.55 with a y-intercept of 0.376. For the simple linear regression: F-statistic: 90.77 on 1 and 227 DF, $p = 0.000$. R^2: 0.286. The gray lines represent trends on each of the four main landmasses (Eurasia, Africa, South America, and North America). Dot size represents mean population (log-transformed).

it is not actually a distinguishable landmass, the association changes little: pseudo $R^2 = 0.902$, $p = 0.000$.) If we use the same approach with medians to control for relatedness, we find a similarly striking association between the regions' vowel indices and humidity values (pseudo $R^2 = 0.771$, $p < 0.00001$). (Again, if we exclude the Pacific region the association remains significant: pseudo $R^2 = 0.673$, $p = 0.001$.) At least in the case of these six separable areas, regions with lower humidity values tend to have lower vowel indices. This pattern is difficult to reconcile with the idea that the vowel-usage/humidity association is due somehow to contact between languages. While there are a limited number of data points in such a cross-continental correlation, the pattern in **Figure 5** is remarkably consistent with the notion of gradual linguistic adaptation to ecological constraints.

The Pacific "region" is not, of course, a distinct landmass amenable to intra-continental analysis. It consists of many smaller landmasses including New Guinea, Borneo, and Sumatra, as well as many islands that were only relatively recently inhabited through the Polynesian expansion. Yet the "region" is generally

characterized by high humidity values with limited variation (See Supplementary Figure 1). This relative environmental consistency motivates its inclusion in the cross-regional rankings just discussed, but is another factor that makes intra-Pacific analyses uninformative. The Australian landmass is a less straightforward case, however. In the database utilized, all language varieties in Australia are grouped together in the same family. Yet, even if we were to categorize them according to the traditional division of Pama-Nyungan/Non-Pama-Nyungan, we would be unable to uncover trends that could be said to characterize the region but not particular families. In contrast, recall that there are between 23 and 69 families represented for each of the four major continents. The Australian data are also problematic in that the Australian landmass does not exhibit the same ecological diversity as the four major continents. The driest regions in the world are extremely cold regions, which Australia lacks. Still, there are some very dry desert regions in Australia and many languages are spoken on the continent. While these regions are not characterized
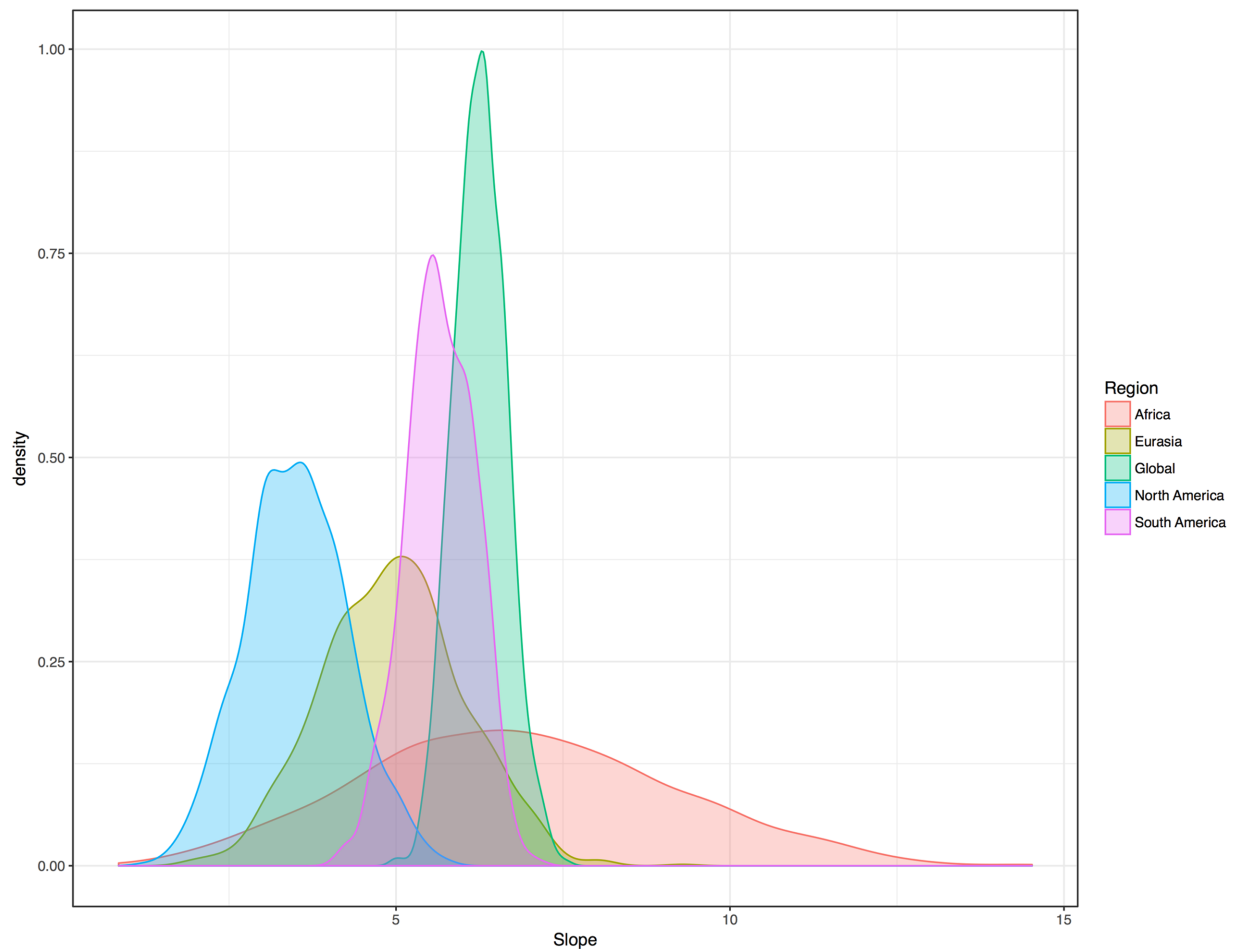
FIGURE 4 | Density distribution (Gaussian kernel) for the slopes of the randomly sampled regressions of vowel indices and humidity. Each color represents 1000 slopes. Each regression is based on a new list created from a sample of languages consisting of one randomly selected member of each family. All 5000 slopes are positive.

by the extreme aridity observed in the winters of parts of Eurasia, North America, South America, or even Africa, they are still quite dry judging from their annual averages. So a beta regression was run separately for Australia and the results of this regression are inconsistent with the other landmasses. In fact, in Australia a negative association between vowel index and humidity was observed (pseudo $R^2 = 0.124$, $p = 0.0001$). This certainly runs against the general trend and the guiding hypothesis. Still, some caution is required before giving the Australian case equal weight alongside the other continents. In addition to the already-noted issues, another relevant point should be made: The negative correlation in Australia is driven largely by the relatively low vowel indices obtained in higher humidity regions, not by objectively high vowel indices in dry regions. The lowest humidity value obtained for Australia was 0.0050, for three languages. The mean vowel index of these languages was 0.464, very close to the world median of 0.458. In other words, the languages in the dry regions of Australia do not have high vowel indices since they hover around the

median for the global sample, while some of the languages in Australia's more humid regions have lower-than-normal vowel indices. This is worth noting since the hypothesis motivating this work, as in Everett et al. (2015), is that very dry air may impact phonation in at least some real-world contexts. (In that study it was noted that tonality is not observed in Australia.) If this hypothesis is accurate, high vowel indices should be avoided in dry contexts. This expectation is not, strictly speaking, violated in Australia. Consider this: In Australia, the highest humidity value of 0.01775 was obtained for four languages. These languages had an average vowel index of 0.401, which is actually quite a bit lower than the world median. (In contrast, the vowel indices for the languages with the highest humidity values on each of the four major continents were 0.559 [Africa], 0.516 [North America], 0.508 [Eurasia], and 0.496 [South America].) When considered in the light of the values observed in the rest of the world, the Australian trend owes itself to low vowel indices in high humidity areas rather than high vowel indices in arid regions. Nevertheless, it should be acknowledged that the intra-Australian
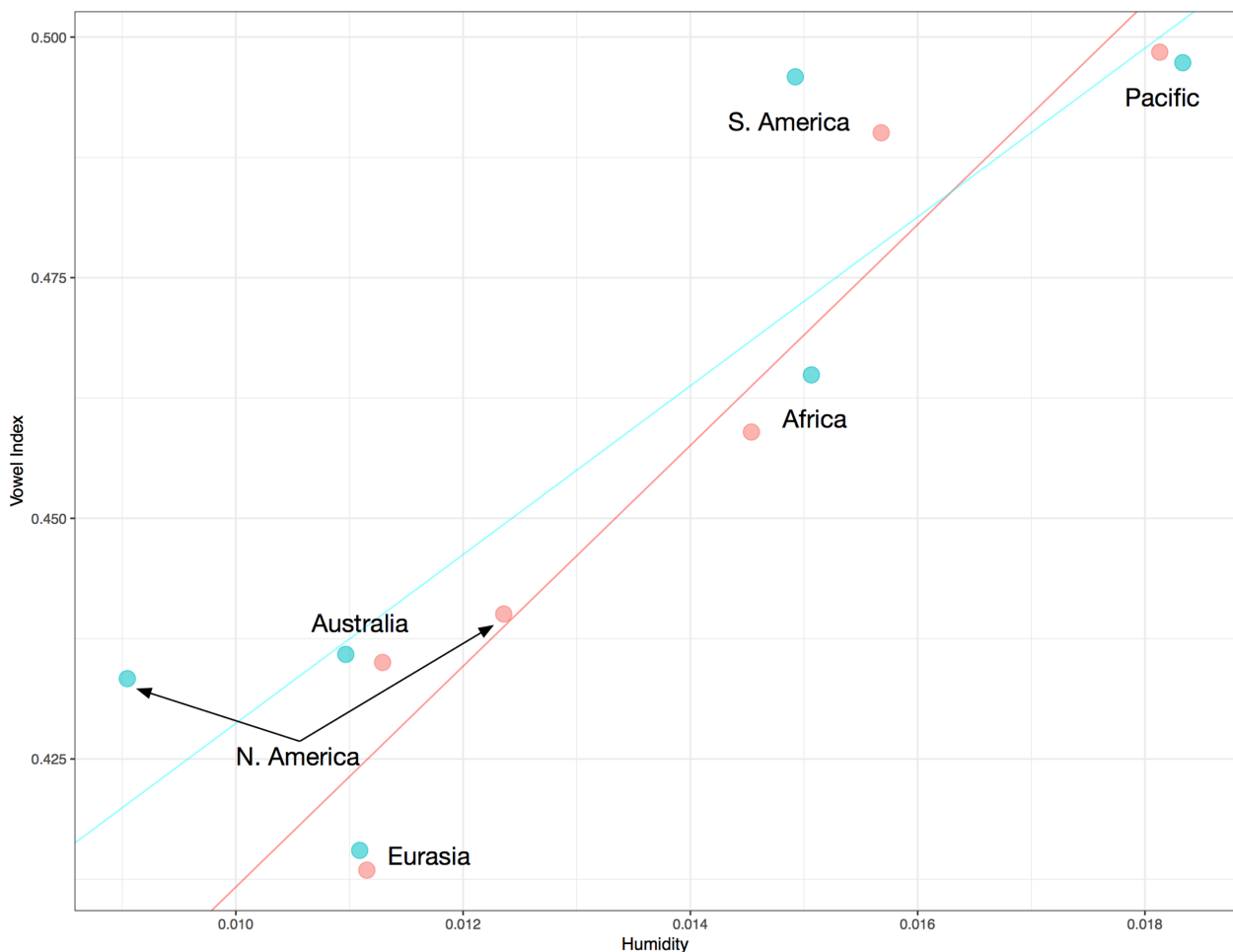
FIGURE 5 | Vowel indices by region, controlled for relatedness. Cyan dots represent the median values of a region, i.e., the median value of the families' median values. Red dots represent the mean values of the families' average values, for each region. Beta regressions of either medians or means are significant. (See text.)

trend contravenes those observed on the four major continents. It remains possible that the overall global pattern observed owes itself to coincidental trends on those continents. The Australian trend may hint at such a coincidental association, or it may simply hint at the problems of including analyses of continents with limited phylogenetic detail. The latter notion would seem to be more consistent with the cross-regional pattern in **Figure 5**, which suggests Australia is unexceptional at a less telescoped level.

It should be noted that, in some seminal typological work on the usage of correlational data, Australia has been grouped with New Guinea. In Dryer (1989), for instance, the five major global regions that are suggested for testing are Eurasia, Africa, North America, South America, and Australia/New Guinea. If that methodological tack is taken, we find that the association under examination does surface within each of the five regions. A beta regression between vowel index and humidity in Australia/New Guinea reveals a positive association (Pseudo $R^2 = 0.059$, $p < 0.00001$). This association owes itself largely to the fact that many Australian language varieties are spoken in regions that are

drier than New Guinea, and that the Australian vowel indices are generally low.

Random sampling was also used to better elucidate the nature of the global association, in a manner more similar to that used in Everett et al. (2015) vis-à-vis tonality. One member of each language family was selected at random. The resultant list of 229 languages was then ranked according to humidity. For each of 5000 generated samples of this type, one member of the highest quartile of humidity was chosen at random (language a), and one member of the lowest quartile of humidity was chosen at random (language b). The vowel index of language b was then subtracted from the vowel index of language a. For each of 5000 additional samples, the same methods were applied except that languages were randomly selected from the 2nd and 3rd quartiles of humidity rankings, for each iteration. The net result of these 10,000 contrasts is depicted in **Figure 6**. (For verbose results, see "Materials and Methods.") The 2nd and 3rd quartiles of phylogenetically controlled samples tend to have similar vowel indices, though the 3rd quartile generally has higher ones. In contrast, the lowest and highest quartile languages, in terms of
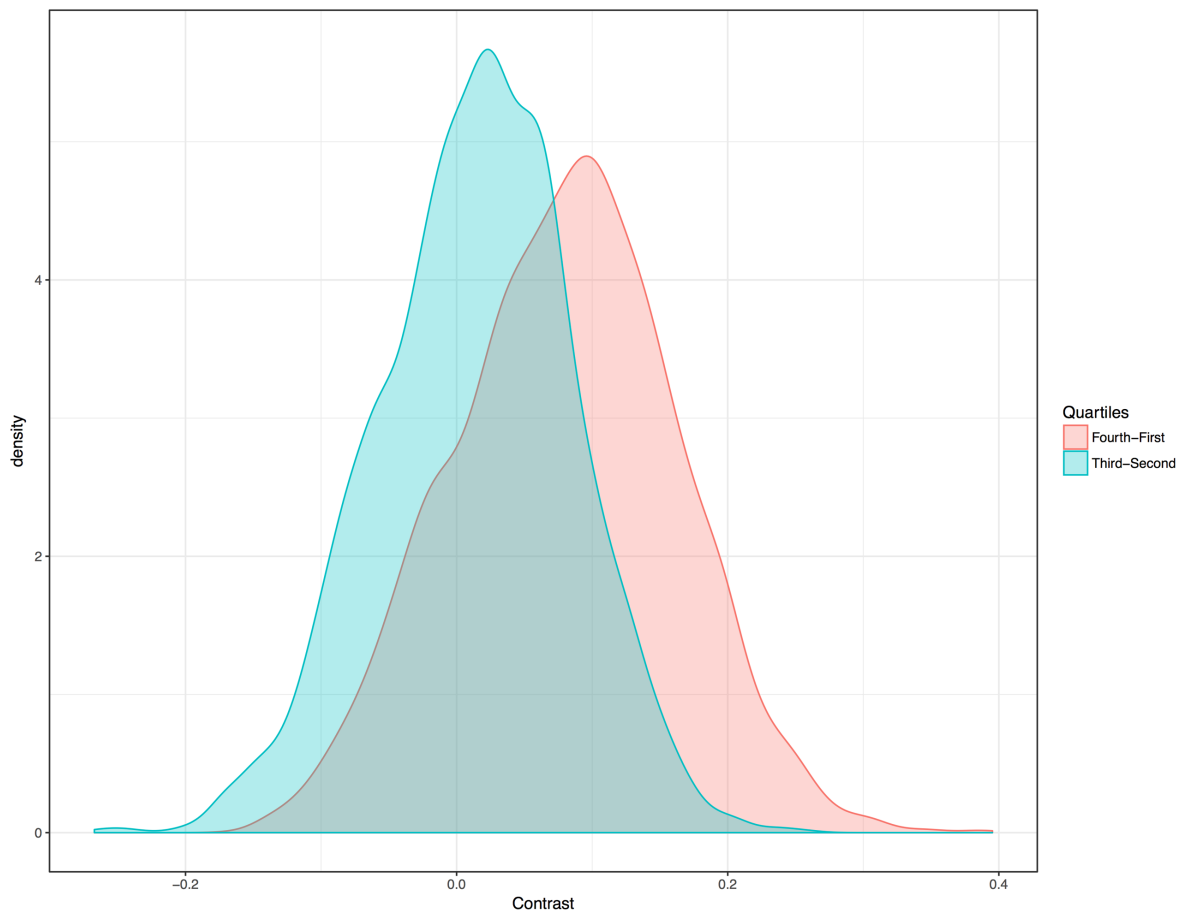
FIGURE 6 | Density distribution (Gaussian kernel) for vowel-index contrasts of languages from different humidity quartiles. Each color represents 5000 contrasts. Each pairwise contrast was created from a new list, ordered by humidity and representing one randomly selected member of each family. The vowel index of one member of a given quartile was contrasted with the vowel index of a member of the other quartile.

humidity, differ more consistently with respect to vowel index. Languages with the lowest vowel indices are clearly likely to occur in dry regions.

Languages in dry regions simply do not exhibit very high vowel indices. The upper left quadrant of **Figure 3** is blank. Or consider that 55 of the languages in the driest quartile, for the entire dataset, have vowel indices two standard deviations or more below the mean vowel index. In contrast, zero languages in the driest quartile have vowel indices two standard deviations or more above the mean. This is consistent with the notion, requiring further exploration, that languages adapt to very dry air. Of course, even basic words are not immune to contact-based effects and languages moving into dry regions may come to have low vowel indices partly because their new neighbors do. Yet such influence would not explain why those neighbors consistently had low vowel indices in the first place.

One might object that the positive correlation between humidity and vowel indices does not necessarily imply lesser overall rates of vowel usage in drier contexts. Perhaps languages in drier climates use less vowels compared to consonants, but tend to have longer words so that their overall usage of vowels

is not actually lower. This seems unlikely but, if it were the case, it would run counter to the suggestion that languages are adapting to ecological constraints. To test this possibility, the mean word length for each language variety was ascertained. Word lengths were based on the sum total of all consonants and vowels, for each of the 40 core words in the ASJP database. The average word length, by language variety, was 4.12 consonants and vowels (median 4.07), with some outlying language varieties differing in pronounced ways. (The lowest mean word length was 2.19, the highest was 9.42.) A very weak but positive correlation was observed between word length and humidity (Adjusted R^2 0.033). So it is apparently not the case that languages in drier regions have lower vowel indices but compensate with greater word lengths. If that were so, there would be a negative correlation between word length and humidity.

Finally, it is worth separately examining large linguistic families to test whether the proposed association is evident within such groups. There are six aforementioned linguistic groups with more than 100 representatives (excepting "Australian," just discussed), comprising over half the languages in the sample. Two of these have few if no representatives in the driest

regions (Austronesian and Trans-New-Guinea). Within-family beta regressions suggest the tendency surfaces weakly within each of the four remaining major families. It is significant for Niger-Congo ($p = 0.02$), and Afro-Asiatic ($p = 0.000$), but not for Sino-Tibetan ($p = 0.13$) and Indo-European ($p = 0.13$). It is actually unclear whether any potential probabilistic effects of environment should accrue within families, but perhaps they do.

Such tests of large families hint at diachronic trends consistent with the influence of environment on language. But these tests of families are admittedly crude, treating families as flat structures. Much work is required to offer clear within-family diachronic support for the suggested influence. One potential approach is to implement the family bias method described in Bickel (2013), though such an approach would require binning individual languages according to vowel indices and humidity levels. Another possibility [suggested in Everett et al. (2016a)] would be to test linguistic migrations and within-family diachronic trends against climate models. This method would require fine-grained mappings of particular language families, utilizing Bayesian phylogenetic methods and/or trees established via more traditional comparative methods. In any case, the suggestion that climate impacts languages admittedly requires fuller exploration with more robust diachronic approaches. Unfortunately, however, such approaches will also face obstacles since only a handful of large linguistic families are mapped with sufficient confidence at present. Nevertheless, the vowel-usage data offered here may assist in such explorations. Given that the suggested influence is non-deterministic, and given that many factors are at play in sound changes, within-family trends across many linguistic taxa would ideally be considered.

I have suggested that the clearest potential explanation of the vowel-index/humidity association is that based on the deleterious effects of dry air on phonation. However, acoustically oriented accounts (e.g., Maddieson and Coupé, 2015) offer similar predictions. Those accounts are consistent with the fact that vowels are less common in drier regions. If the explanatory variable is in fact increased phonatory effort, as I am suggesting, perhaps we should see differences in the distribution of voiced consonants as well. Yet predictions for consonants are weaker since they generally require less phonation, and with less amplitude. Furthermore, the predictions of a phonation-oriented account are not uniform across manners of articulation. Consider nasals: While nasals are almost always voiced, they are made with a closed oral cavity and open nasal passageway. During inhalation, this configuration promotes humidification. So an account based on the effects of aridity on phonation makes no clear predictions for nasals. Nevertheless, the possibility of the effects of dry air on reduced voicing in consonants merits further exploration. As noted above, the ASJP database collapses some consonants according to voicing. This is particularly true of fricatives. Still, some limited explorations of voicing distinctions for consonants are possible. To that end, I examined the prevalence of voiced stops, nasals, rhotics, and laterals across the 4012 word lists. (The three latter kinds of sounds are generally voiced.) For each of these four consonant types,

the total of the given consonant was divided by the number of all consonants and vowels, for each word list. Instead of a vowel index, then, this approach yielded a "voiced-stop index," a "nasal index," a "rhotics index," and a "lateral index." These indices were then tested against ambient humidity, as with the vowel index. Separate multivariate regressions, each with logit-transformed indices as dependent variables and humidity and language family as independent variables, were run. The results suggest that voiced stops are in fact slightly less common in more arid regions, even after controlling for language family in this manner ($p < 0.0001$). The same is true of rhotics ($p < 0.001$). No significant phylogenetically controlled patterns were observed for laterals and nasals. As just mentioned, though, the predictions of a desiccation-oriented account are unclear for nasals. Also, laterals may be voiceless and the voicing distinction is collapsed for laterals in the ASJP database. So the data do not lend themselves to clearly exploring degrees of phonation vis-à-vis consonants. And, as noted above, the predictions of a phonation-based account are much clearer for vowels. Nevertheless, this initial analysis of consonant data points intriguingly to the possibility of ambient pressures against the voicing of consonants as well, at least in the case of stops and rhotics. In the case of these consonants, though, it is unclear whether the patterns in question are independent of the more pervasive pattern associated with vowels, since vowels can influence the voice-onset-times of their consonant neighbors. Such issues require more substantive investigation with other databases.

DISCUSSION AND CONCLUSION

The results presented here are consistent with the possibility that human sound systems evolve in accordance with environmental pressures. The association between low humidity values and low vowel indices is very strong, across languages (**Figure 2**) or language families (**Figure 3**). At the regional level, the evidence is generally supportive as well. The association surfaces on four of five major landmasess, though Australia is a counter-example. However, the Australian results are those that should be approached with the most caution for reasons observed above. While this is not the first study to suggest a potential environmental effect on languages, it is the first to do so without binning strategies that some scholars have found objectionable. It is also the first study to rely on sounds in actual words in 1000s of languages. The study has offered several controls for language families and areal influences, simultaneously. Yet the results are merely consistent with the idea that languages adapt, probabilistically and likely over long periods, to the influence of dry air on the larynx. Much work is required to understand this possible interaction. Some degree of circumspection is warranted since these data are correlational and the multifarious factors impacting language change are already known to be complex, and to interact in complex ways. And, of course, the possibility of a coincidental correlation remains and establishing causal links between correlated linguistic and extralinguistic variables is not straightforward. (Roberts and Winters, 2013) To paraphrase an old adage, though, while correlational data

cannot establish causal relationships, they often gesticulate wildly in the direction of such relationships. Additionally, as researchers including myself have pointed out elsewhere, the traditional assumption that linguistic sounds are immune to ecological influence is problematic given the extent of ecological adaptability in human behavior (sometimes non-conscious), given that ecological adaptation characterizes the communication of other species, and given the dearth of research carefully examining this possibility. So while these results do not establish a causal influence of dry ambient air on language, they call for further consideration of this possibility and for continued exploration of this topic. They also demonstrate that global geo-phonetic correlations cannot be written off as the byproducts of convenient binning strategies.

As noted above, previous work has suggested that languages' phonemic inventories have a greater ratio of consonants in colder (and therefore drier) climates. (Maddieson et al., 2011) A higher number of consonants in a phonemic inventory points to diversity of sounds in a language, which hints (but does not demonstrate) that the language may rely on consonants more in the speech stream. To examine a potential relationship between C:V ratios and vowel indices, the WALS data on C:V ratio complexity were cross-referenced with the new vowel index data. This yielded 262 languages, across a diversity of families and regions, with both vowel indices and C:V ratios. The C:V ratios in WALS are categorized on a scale from 1 to 5, with five representing languages with high ratios of consonants in their phonemic inventories. A weak but significant interaction of C:V ratio and vowel index was observed: Languages with high vowel indices tend to have lower C:V ratios, as we might expect. (pseudo $R^2 = 0.029$, $p = 0.005$). An effect of humidity on C:V ratio was also observed, with higher C:V ratios associating with drier regions (adjusted $R^2 = 0.04$, $p = 0.001$). However, that interaction is quite modest when contrasted to the interaction of humidity and vowel index. (Given the limited number of languages with known C:V ratios and known vowel-indices, phylogenetic and areal controls could not be applied.) These findings suggest that vowel indices may more clearly reflect potential climatic effects on language, when contrasted to phonemic inventory data. It is hoped that vowel indices offer a useful new sort of data to be used in the further exploration of this topic.

In a similar vein, previous work with smaller, binned data sets has suggested that languages in warm regions rely more heavily on CV syllables for acoustic reasons (Munroe et al., 2009). While the prevalence of such syllables in warm regions may have acoustic motivations, that prevalence may also be due to the laryngeal factors discussed above. The claim that there are more CV syllables in warm places is, practically speaking, very similar to the claim that vowel indices are higher in humid places. The core of the account offered by Munroe and colleagues is this: People in colder regions tend to be closer to their interlocutors during speech events, and so need not rely as much on vowels that are so sonorant. This is an interesting claim, though I am unaware of any ethnolinguistic data supporting it. Given that there are findings in laryngology demonstrating an effect of dry air on the vocal cords, I believe that the most direct

potential explanation for the distributional findings here and in previous work on "acoustic adaptation" in speech is one that is grounded on the interaction of vocal-tract physiology and ambient air. Yet, while I submit that this is the most plausible account at present, acoustically oriented accounts cannot be ruled out (much like coincidental distributions cannot be ruled out). It should also be noted, though, that while vowel indices correlate positively with temperature, their association with humidity is more pronounced. (See discussion of temperature data in "Materials and Methods.") This suggests the former association may be epiphenomenal and would seem to further weaken the likelihood of an acoustically driven account.

The weak correlation between higher C:V ratios and lower vowel indices points to a crucial caveat required of the findings in this paper: Correlations that have previously been found via the usage of binning strategies should not be taken as independent support, alongside these, for the environmental adaptation of languages. For example, the apparent avoidance of complex tone in dry environs may be related to the association observed here since tone is conveyed via vowel pitch alterations. The ASJP database does not encode tonality, but future work could explore how interrelated these two findings are. It may be the case that the tonality findings are in a sense epiphenomenal, i.e., languages may come to rely less on tone when tone-carrying segments are less preponderant. Conversely, though, in some cases vowels may be more likely to be elided if they do not carry suprasegmental information. These sorts of tentative possibilities point to the difficulty of disentangling the interrelated patterns observed here and in previous work on tonality. Along the same lines, in Everett (2013) I suggested that ejective consonants might be more frequent at higher elevations because compression of the oral cavity is facilitated by reduced ambient air pressure. That is still an untested possibility, but it is also possible that the prevalence of ejectives in some regions is a byproduct of the relatively high frequency of consonants in arid ecologies. I am not making that claim here, instead I am simply pointing out that all these correlations described in the literature are potentially interrelated and cannot be taken as independent support for possible ecological effects on speech. Yet it is also true that the observed correlations run in the same direction, so to speak. So all these patterns are potentially explainable via one main effect of aridity on language. The results presented here, based on continuous variables, are likely the clearest signs of such an effect so far uncovered. But they are still just signs, guideposts at the beginning of an exploration. This exploration may eventually discover another variable that has not been considered in this work. Such a discovery could point to an indirect relationship between language and environment, rather than the direct one postulated here.

As I have pointed out previously (with Damián Blasi and Seán Roberts), one issue that requires further exploration is the way such patterns could surface over time. What kinds of mechanisms may actually motivate the distributional pattern described here? Are vowel elision and vowel epenthesis slightly more and less likely, respectively, to occur in very dry regions? Are innovators of sociolinguistic change more likely (however, slightly) to rely on easier-to-articulate lexical variants with reduced vowels, in

dry regions? Are elderly speakers more likely to produce easier-to-articulate variants, given that they seem particularly prone to the effects of humidity on phonation? (Sundarrajan et al., 2017). Given that vocal-tract maladies like laryngitis are more prevalent in dry environments, are some sociophonetic variants more likely to be selected for, probabilistically over centuries, in such regions? As has been asked before, are speakers of tonal languages (particularly second-language speakers) slightly less likely to precisely replicate multiple level and/or contour pitches because of increased jitter rates in very dry contexts? Such questions hint at possible avenues of research. While linguists have carefully documented many kinds of sound change, it remains to be seen how an ecological factor that impacts the vocal cords could act in concert with known processes of diachronic change in or across languages. Finally, it is also possible that acoustic/perceptual factors also play a role in a potential language-ecology interaction, though I have expressed skepticism toward that possibility here. In short, much work is required to better understand the possible language-environment interaction.

The environment has recently been shown to impact languages in a way once-dismissed by some: languages in cold regions are more likely to express a distinction between "snow" and "ice," when contrasted to languages in warm regions. Despite previous anecdotally based claims for and against the generalization, it was uncovered after careful sampling of many languages (Regier et al., 2016). As the authors of that study note, it is only when considering a probabilistic, rather than deterministic, connection between language and the environment that such patterns emerge. This conclusion would appear to hold not just with respect to some human words, but also with respect to human sounds. These new results are suggestive of another kind of probabilistic interaction between the environment and language. But the results are just pointing for now, and more research is required to find out exactly where they are pointing.

MATERIALS AND METHODS

Data and code in Supplementary Material. Coding and analysis conducted with R.

All p-values are two-tailed.

The forty basic concepts described by the word lists are presented here: https://en.wikipedia.org/wiki/Automated_Similarity_Judgment_Program.

Vowel indices were obtained for each word list via a function that summed the vowels in a given list and then divided this sum by the total number of consonants and vowels in the list. The script for this function was based on the transcription conventions used for the ASJP database, to ensure that all vowels and consonants were counted and that secondary characters were excluded (Brown et al., 2008). Tone and length are not encoded in the database. Nasal and oral vowels were treated the same for the purposes of this study. In the ASJP database, some syllable types are simplified for transcription. In particular, CV7C, CVhC, CVxC, and CVXC syllables are reduced to CVC. (The 7 is a glottal

stop, the h is a glottal fricative, the x is a velar fricative, and the X is a uvular fricative.)

As an example of how the vowel index is calculated, consider the Pirahã words in the ASJP database. There are 100 words in this case. Here is the list, with the words separated by commas: ti, gi7ai, tiatiso, gisai, gaihi, kaoi, go, ogiagao, aibai, hoi, hoi, ogi, pi7i, oihi, ipoihi, igihi, iti7isi, pibigi, giopai, tihihi, aoisi, tai, ipi, soi, isigihi, bipai, ahiai, sitoi, isapai, igai, isitai, apaitai, apapai, kosi, itaoi, kaopai, aitoi, ipopai, opoi, aosi, ko7otai, boasai, bogai, iosi, ibioi, ita7ipi, ohoai, abi, obi, aobisai, ko7o7as7∼aga, aiti, koabai, oabai, pibai, kobababopi, iho, hoagi, aitahoi, abaipi, ipopao, hoai, gai, hisi, kahaixai, ogihiai, pi, pi, a7ai, tahoasi, bigi, hoa7ai, hoa7ai, hoai, hoati, hoagaipi, agi, bigi hio7o7iai, bisi, ahoasai, bisi, kobiai, kopaiai, ahoai, hoai, agi, kabi, asi, ba7ai, hioi, and kasi.

There are 300 instances of the three Pirahã vowels. There are 464 total vowels and consonants. So the vowel index for this language is 300/464 or 0.64655.

Two symbols in the ASJP database, ∼ and $, are used to denote monosegmentality. Since this analysis was concerned with phonetic strings, these symbols were ignored. Even consonants with very short duration, like instances of prenasalization, were considered relevant. In some cases of definitive coarticulation, this choice would lead to slightly higher consonant counts. Since the goal here was to count all phonetic units, this choice was well motivated. In any case, this choice seems to have little impact on the overall results. To be sure of this, a separate analysis was run in which ∼ and $ were factored into the vowel index. In this analysis, each instance of ∼ reduced the consonant count of a list by one. Each instance of $ reduced the consonant count of a list by two. The vowel indices obtained through this method did not differ appreciably in most cases, and the key pattern evident in **Figure 3** is very similar regardless of which approach is taken with these symbols. With this slightly different approach to the vowel index, a beta regression analyzing the median vowel index (by family) according to median humidity (by family) reveals a pseudo R^2 of 0.257, $p < 0.0000001$.

Specific humidity values were originally gathered by Seán Roberts for a previous study (Everett et al., 2015). The data were obtained by averaging six decades worth of specific humidity values from the National Oceanic and Atmospheric Administration. Specific humidity is the measure of water content in the air, as a ratio of all water and air. Like absolute humidity, it is a much better indicator of water content than relative humidity (Maddux et al., 2016). Mean annual temperature data were obtained via the Bioclim package in ArcGIS by Justin Stoler, a colleague at the University of Miami.

The 4012 word lists used for the bulk of the analysis represent 2632 unique ISO codes that could be cross-referenced with the humidity data. This suggests that many ISO codes are represented with word lists for more than one dialect, a point supported by visual examination of the data set. Using ISO codes for cross-referencing ensured that dialects of the same language were tied to the same geographic region (which they are in most cases anyhow). This way, the dialects' climate data were more characteristic of their history rather than recent migrations.

The shapiro test was used to examine whether vowel indices, humidity, and temperature represented normal distributions.

None did, at least in part because most cultures are found in equatorial regions. So the median values are presented in the summaries of these variables, and analyzed alongside means in all cases.

Results of beta regression of means for all families in **Figure 3**: Log-likelihood = 371.8 on 3 Df, pseudo R^2 = 0.286. Results of beta regression of medians for all families: Log-likelihood = 365.5 on 3 Df, pseudo R^2 = 0.267. Results of beta regressions of families' median vowel index and median humidity values, by continent: Africa = pseudo R^2 = 0.311, p = 0.001. Eurasia = pseudo R^2 = 0.293, p = 0.0003. North America = pseudo R^2 = 0.093, p = 0.029. South America = pseudo R^2 = 0.208, p = 0.00002.

For the one-random-language-per-family samples, the mean slopes and R^2 values for 1000 linear regressions per region were 6.27/0.23, (world), 4.96/0.21 (Eurasia), 6.83/0.25 (Africa), 3.51/0.13 (North America), and 5.66/0.17 (South America).

For the top/bottom quartile contrasts of languages randomly selected from families, by ordered humidity: In 893/5000 simulations, the languages from the bottom quartile of humidity had higher vowel indices than those from the top. For 3/5000, they were the same. For 4104/5000 the vowel indices of the highest humidity quartile were greater. For second/third quartile contrasts: In 2041/5000 simulations, the languages from the second quartile of humidity had higher vowel indices than the third quartile. For 4/5000, they were the same. For 2955/5000, the vowel indices of the third humidity quartile were greater than those for languages from the second quartile.

For the beta regression of 6901 language varieties' vowel index and temperature values: pseudo R^2 = 0.099, p = 0.0000. This regression is less explanatory than that in **Figure 2**, suggesting that humidity is a better predictor of vowel index than temperature. A multiple regression found that temperature is a significant predictor of logit-transformed vowel indices

(p = 0.00005), even after including linguistic family as an independent variable.

For the investigation of consonants: A nasal index, a voiced-stop index, a rhotic index, and a lateral index were each calculated separately. Some zero values were obtained since some word lists lack one of the consonant types. These zero values were raised to the lowest non-zero values for each respective index, in order to allow for logit-transformation of the values.

AUTHOR CONTRIBUTIONS

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

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01285/full#supplementary-material

REFERENCES

Bickel, B. (2013). "Distributional biases in languages families," in *Language Typology and Historical Contingency*, eds B. Bickel, L. Grenoble, D. Peterson, and A. Timberlake (Amsterdam: John Benjamins), 414–444. doi: 10.1075/tsl.104

Brown, C., Holman, E., Wichmann, S., and Velupillai, V. (2008). Automated classification of the world's languages: a description of the methods and preliminary results. *Lang. Typol. Univ.* 61, 285–308. doi: 10.1524/stuf.2008.0026

Calude, A. S., and Pagel, M. (2011). How do we use language? Shared patterns in the frequency of word use across 17 world languages. *Philos. Trans. R. Soc. B* 366, 1101–1107. doi: 10.1098/rstb.2010.0315

De Boer, B. (2016). Commentary: Is the effect of desiccation large enough? *J. Lang. Evol.* 1, 55–57. doi: 10.1093/jole/lzv008

Dediu, D., Janssen, R., and Moisik, S. R. (2017). Language is not isolated from its wider environment: vocal tract influences on the evolution of speech and language. *Lang. Commun.* 54, 9–20. doi: 10.1016/j.langcom.2016.10.002

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Stud. Lang.* 13, 257–292. doi: 10.1075/sl.13.2.03dry

Dryer, M. S., and Haspelmath, M. (2016). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ember, C., and Ember, M. (2000). High CV score: regular rhythm or sonority. *Am. Anthropol.* 102, 848–851. doi: 10.1525/aa.2000.102.4.848

Ember, C., and Ember, M. (2007). Climate, econiche, and sexuality: influences of sonority in language. *Am. Anthropol.* 109, 180–185. doi: 10.1525/aa.2007.109.1.180

Erickson, E., and Sivasankar, M. (2010). Evidence for adverse phonatory change following an inhaled combination treatment. *J. Speech Lang. Hear Res.* 53, 75–83. doi: 10.1044/1092-4388(2009/09-0024)

Evans, N. (2016). Typology and coevolutionary linguistics. *Linguist. Typol.* 20, 505–520. doi: 10.1515/lingty-2016-0023

Everett, C. (2013). Evidence for direct geographic influences on linguistic sounds: the case of ejectives. *PLoS ONE* 8:e65275. doi: 10.1371/journal.pone.0065275

Everett, C., Blasi, D., and Roberts, S. (2015). Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1322–1327. doi: 10.1073/pnas.1417413112

Everett, C., Blasi, D., and Roberts, S. (2016a). Language evolution and climate: the case of desiccation and tone. *J. Lang. Evol.* 1, 33–46. doi: 10.1093/jole/lzv004

Everett, C., Blasi, D., and Roberts, S. (2016b). Response: Climate and language: Has the discourse shifted? *J. Lang. Evol.* 1, 83–87. doi: 10.1093/jole/lzv013

Everett, D. (2005). Cultural constraints on Pirahã grammar. *Curr. Anthropol.* 46, 621–646. doi: 10.1086/431525

Flemming, E., Ladefoged, P., and Thomason, S. (2008). Phonetic structures of Montana Salish. *J. Phon.* 36, 465–491. doi: 10.1016/j.wocn.2007.10.002

Fought, J. G., Munroe, R. L., Fought, C. R., and Good, E. M. (2004). Sonority and climate in a world sample of languages: findings and prospects. *Cross Cult. Res.* 38, 27–51. doi: 10.1177/1069397103259439

Greenhill, S. (2016). Overview: debating the effect of environment on language. *J. Lang. Evol.* 1, 30–32. doi: 10.1093/jole/lzv007

Hemler, R., Wieneke, G. H., and Dejonckere, P. H. (1997). The effect of relative humidity of inhaled air on acoustic parameters of voice in normal subjects. *J. Voice* 11, 295–300. doi: 10.1016/S0892-1997(97)80007-0

Ladd, R. (2016). Commentary: tone languages and laryngeal precision. *J. Lang. Evol.* 1, 70–72. doi: 10.1093/jole/lzv014

Leydon, C., Sivasankar, M., Falciglia, D. L., Atkins, C., and Fisher, K. V. (2009). Vocal fold surface hydration: a review. *J. Voice* 23, 658–665. doi: 10.1016/j.jvoice.2008.03.010

Lupyan, G., and Dale, R. A. C. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Lupyan, G., and Dale, R. A. C. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cogn. Sci.* 20, 649–660. doi: 10.1016/j.tics.2016.07.005

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

Maddieson, I., Bhattacharya, T., Eric Smith, D., and Croft, W. (2011). Geographical distribution of phonological complexity. *Linguist. Typol.* 15, 267–279. doi: 10.1515/lity.2011.020

Maddieson, I., and Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation hypothesis. *J. Acoust. Soc. Am.* 138:1838. doi: 10.1121/1.4933848

Maddux, S. D., Yokley, T. R., Svoma, B. M., and Franciscus, R. G. (2016). Absolute humidity and the human nose: a reanalysis of climate zones and their influence on nasal form and function. *Am. J. Phys. Anthropol.* 161, 309–320. doi: 10.1002/ajpa.23032

Mäkinen, T. M., Juvonen, R., Jokelainen, J., Harju, T. H., Peitso, A., Bloigu, A., et al. (2009). Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. *Respir. Med.* 103, 456–462. doi: 10.1016/j.rmed.2008.09.011

Moran, S. (2016). Commentary: issues of time, tone, roots and replicability. *J. Lang. Evol.* 1, 73–76. doi: 10.1093/jole/lzv011

Munroe, R., and Silander, M. (2009). Climate and the consonant-vowel (CV) syllable: a replication within families. *Cross Cult. Res.* 33, 43–62. doi: 10.1177/1069397109303300104

Munroe, R. L., Fought, C., and Fought, J. (2000). Rhythmicity or sonority: response to Ember and Ember's "Cross-language predictors of consonant-vowel syllables". *Am. Anthropol.* 102, 844–848. doi: 10.1525/aa.2000.102.4.844

Munroe, R. L., Fought, J., and Macaulay, R. (2009). Warm climates and sonority classes: not simply more vowels and fewer consonants. *Cross Cult. Res.* 43, 123–133. doi: 10.1177/1069397109331485

Munroe, R. L., Munroe, R. H., and Winters, S. (1996). Cross-cultural correlates of the consonant-vowel (CV) syllable. *Cross Cult. Res.* 30, 60–83. doi: 10.1177/106939719603000103

Napoli, D. J. (2014). On the linguistic effects of articulatory ease, with a focus on sign languages. *Language* 90, 424–456. doi: 10.1353/lan.2014.0026

Regier, T., Carstensen, A., and Kemp, C. (2016). Languages support efficient communication about the environment: words for snow revisited. *PLoS ONE* 11:e0151138. doi: 10.1371/journal.pone.0151138

Roberts, S., and Winters, J. (2013). Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLoS ONE* 8:e70902. doi: 10.1371/journal.pone.0070902

Sivasankar, M., and Erickson-Levendoski, E. (2012). Influence of obligatory mouth breathing, during realistic activities, on voice measures. *J. Voice* 26, e9–e13. doi: 10.1016/j.jvoice.2012.03.007

Sivasankar, M. P., and Fisher, K. (2002). Oral breathing increases Pth and vocal effort by superficial drying of vocal fold mucosa. *J. Voice* 16, 172–181. doi: 10.1016/S0892-1997(02)00087-5

Sundarrajan, A., Brinton Fujiki, R., Loerch, S., Venkatraman, A., and Sivasankar, P. (2017). Vocal loading and environmental humidity effects in older adults. *J. Voice* doi: 10.1016/j.jvoice.2017.02.002 [Epub ahead of print].

Wichmann, S., Holman, E., and Brown, C. (2016). *The ASJP Database (version 17)*. Available at: http://asjp.clld.org/

Wilkins, M. R., Seddon, N., and Safran, R. J. (2013). Evolutionary divergence in acoustic signals: causes and consequences. *Trends Ecol. Evol.* 28, 156–166. doi: 10.1016/j.tree.2012.10.002

Yegerlehner, J., and Voegelin, F. (1957). Frequencies and inventories of phonemes from nine languages. *Int. J. Am. Linguist.* 23, 85–93. doi: 10.1097/AUD.0000000000000040

Non-linguistic Conditions for Causativization as a Linguistic Attractor

Johanna Nichols [1,2,3*]

[1] Department of Slavic Languages and Literatures, University of California, Berkeley, Berkeley, CA, United States, [2] Linguistic Convergence Laboratory, Higher School of Economics, National Research University, Moscow, Russia, [3] Faculty of Arts, University of Helsinki, Helsinki, Finland

An attractor, in complex systems theory, is any state that is more easily or more often entered or acquired than departed or lost; attractor states therefore accumulate more members than non-attractors, other things being equal. In the context of language evolution, linguistic attractors include sounds, forms, and grammatical structures that are prone to be selected when sociolinguistics and language contact make it possible for speakers to choose between competing forms. The reasons why an element is an attractor are linguistic (auditory salience, ease of processing, paradigm structure, etc.), but the factors that make selection possible and propagate selected items through the speech community are non-linguistic. This paper uses the consonants in personal pronouns to show what makes for an attractor and how selection and diffusion work, then presents a survey of several language families and areas showing that the derivational morphology of pairs of verbs like *fear* and *frighten*, or Turkish *korkmak* 'fear, be afraid' and *korkutmak* 'frighten, scare', or Finnish *istua* 'sit' and *istutta* 'seat (someone)', or Spanish *sentarse* 'sit down' and *sentar* 'seat (someone)' is susceptible to selection. Specifically, the Turkish and Finnish pattern, where 'seat' is derived from 'sit' by addition of a suffix—is an attractor and a favored target of selection. This selection occurs chiefly in sociolinguistic contexts of what is defined here as linguistic symbiosis, where languages mingle in speech, which in turn is favored by certain demographic, sociocultural, and environmental factors here termed frontier conditions. Evidence is surveyed from northern Eurasia, the Caucasus, North and Central America, and the Pacific and from both modern and ancient languages to raise the hypothesis that frontier conditions and symbiosis favor causativization.

Keywords: verb, causative, language spread, mixed language, selection, attractor, linguistic symbiosis, linguistic frontier conditions

INTRODUCTION

Sociolinguistics and social context change languages. By now it is understood that absorption of an appreciable number of L2 speakers eventually leads to decomplexification of the absorbing language (the spreading one in a language shift) (Trudgill, 2011), mass bilingualism beginning in childhood can complexify languages (ibid., Dahl, 2004), dense and closed social networks retard language change while open ones foster it (Milroy and Milroy, 1985, 1992), differential degrees of social connection favor uptake and transmission of innovations (Fagyal et al., 2010), and a language whose

speakers have less reliable access to vital resources is likely to have more variation and its speakers to be more accepting of variation than one whose speakers are more secure (Hill, 2001a). But what kinds of (non)-complexity, what kinds of changes, and what kinds of variation?

Here I deal with a specific effect of sociolinguistics on grammar and in particular on grammatical categories: a type of sociolinguistic situation described below appears to favor selection of attractors. An attractor, as the term is understood in complex systems theory, is any state that is easier to enter or acquire than to leave or lose, and/or easier to retain than lose. Selection refers to both uptake and transmission, so that in the end selected features expand in frequency and range, diffusing through both the grammar and the speech community.

Figures 1, **2** (Nichols and Peterson, 2013a,b) show a known attractor that can serve as an introductory example: the phoneme /m/ figuring in personal pronoun systems with a counterposed anterior consonant such as /t/, /č/, /s/, henceforth symbolized with a generic T. Examples are Finnish *minä* 'I' and *sinä* 'you' [and similarly for most of the the sister Uralic languages of Finnish, e.g., Erzya Mordvin (central Russia) *mon, ton,* Selkup (Samoyedic branch, southern Siberia) *man*, *tan*], Georgian (Kartvelian family) *me, šen*, Latin *me* 'me' (accusative case), *te* 'you' (accusative), and many others. This pattern of first person /m/ and second person T is widespread in northern Eurasia, where it occurs in several separate language families and in most daughters of those families, but it is quite rare elsewhere (**Figure 1**). A similar pattern occurs in the western Americas, where a number of languages have first person /n/ and second person /m/, e.g., Wintu (Wintun, northern California) *ni* 'I, we', *mi* 'you'; Pipil (Uto-Aztecan, Nicaragua) *nu-* 'my', *mu-* 'your'; Mapudungun (isolate, Chile) *ñi* 'my', *mi* 'your' (**Figure 2**). Each of these patterns is frequent and densely attested in several separate language families in its own macrocontinent, but very rare elsewhere.

This geography indicates that each system has enjoyed an evolutionary advantage in its respective macrocontinent—and only there. Furthermore, the Eurasian pattern, where we have longer historical records and early attestation of languages, has demonstrably expanded over the last few millennia (Nichols, 2012a,b, 2013), with gains outnumbering losses and /m/ in particular sometimes gained in pronouns but almost never lost from them. Now, /m/, /n/, and T (especially in the form /t/) are very basic sounds, learned early by children, present in the sound systems of most languages, and easily audible, but if these factors motivated their expansion and stability in pronouns we would expect *m-T* pronoun systems to be common worldwide. Pronouns are rarely borrowed from one language to another, and abstract consonantal skeletons of words are borrowed very rarely if at all; but these factors have not inhibited spreads of pronoun forms in Eurasia and the Americas. Thus, what calls for explanation here is the distribution revealed by the geography.

What appears to favor the emergence and spread of such systems is, first, attractor status and second, a sociolinguistics that favors selection of attractors even in the resistant domain of pronouns. Why these systems are attractors is not covered here, and in any case accounting for it would amount to

explaining frequencies of elements in the pre-existing variation that selection works on, rather than describing the mechanism of selection itself (to put it in terms of Darwinian theory). Here, evidently the status of attractor becomes relevant, and selection goes to work, only in the presence of factors other than the phonetic basicness of /m/ or the grammatical basicness of pronouns. Those factors appear to center on ones enhancing the prospects for emergence and uptake of attractors.

Of those factors the one studied here is a sociolinguistics I call *linguistic symbiosis* because it involves two (or more) languages functioning as a single communicative system while remaining discrete (i.e., without forming a mixed language). Symbiosis is the essential coexistence of, and possibility of selection from, more than one language variety, where both (or all) varieties are neutrally valued, selection is bidirectional (or multidirectional), and code switching is accepted. Less technically, in symbiosis two (or more) languages function side-by-side in a society under conditions that make it possible for the languages mingle in speech and for the speakers to select from both languages in a single utterance. The extent and frequency of such mingling are much greater than in ordinary code switching—as, for example, if in discussing Peruvian cuisine I insert the term *aji amarillo* (a variety of pepper), complete with Spanish phonology, into an English sentence or perhaps put the entire phrase or sentence containing it into Spanish (as is possible and not uncommon if both the interlocutor and I know Spanish well). Symbiotic intermingling, in contrast, may be so thoroughgoing that it is difficult for a linguist to decide which of the languages an utterance is in, though the languages actually remain discrete (i.e., they do not merge to create a single mixed language, as occasionally happens under somewhat different sociolinguistic conditions: see e.g., Bakker and Mous, 1994; Meakins, 2013). The main sociolinguistic conditions that make symbiosis possible are lack of a standard or prestige language (which might favor use of one language over the other), minimal or absent language identity or other ideology linking language to other aspects of identity, acceptance of code switching, and sufficient dialect or language diversity to offer a range of options to choose from. Examples are discussed below. This sociolinguistic context facilitates selection and in particular lets attractors be selected because they are attractors and not (e.g.) because they are emblematic of a prestige language.

Propagation of selected attractors is another matter. It is evidently favored by factors that provide opportunities for lateral transmission: sufficiently dense social networks, sufficient distant social connections, and sufficient population mobility, to maintain connections and expose individuals to linguistic diversity, including the range of variation made possible by code switching and bilingualism; and sufficient population density to make possible numerous and long-range social connections and repeat contacts with the same individuals or groups. Some level of density, extent, and reliability of contacts makes it possible for some individuals to be well-connected, and this seems to be essential to the uptake and transmission of innovations (Fagyal et al., 2010). Now, sufficient population density to suppport dense and extensive social networks, in ecological and economic conditions supportive of mobility, has probably existed to any
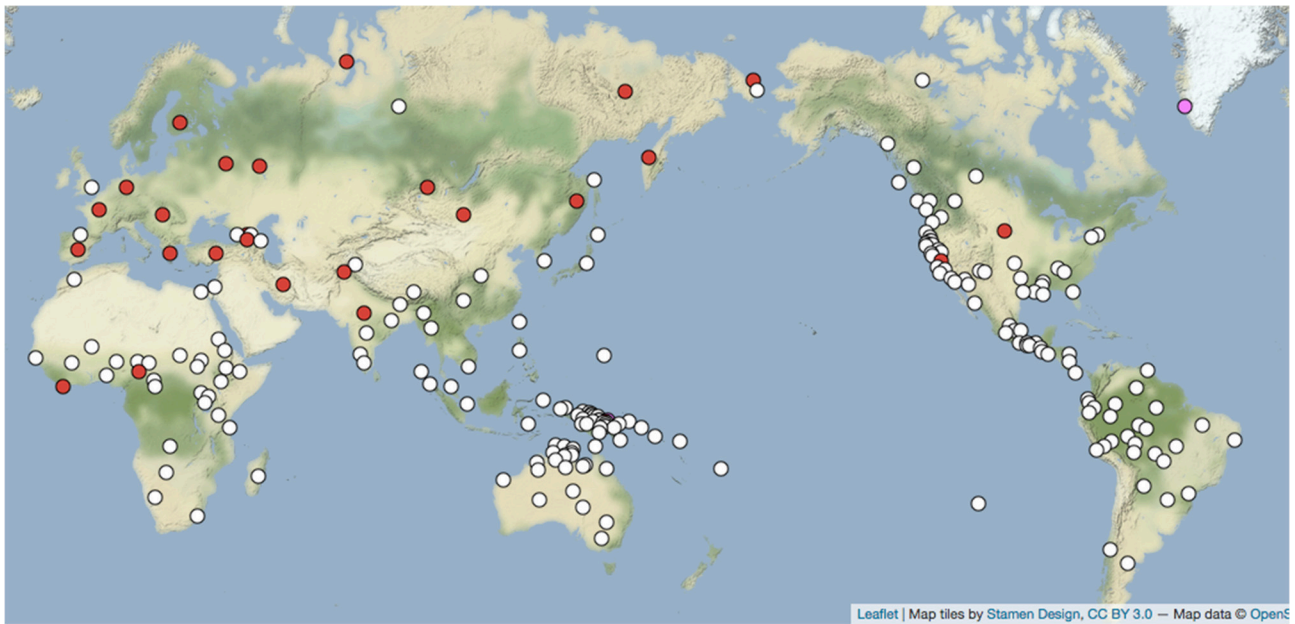
FIGURE 1 | *m-T* pronoun paradigms (*N* = 230). Red = *m-T* paradigm present; white = absent (Nichols and Peterson, 2013a,b). http://wals.info/feature/136A#2/24.8/153.6.
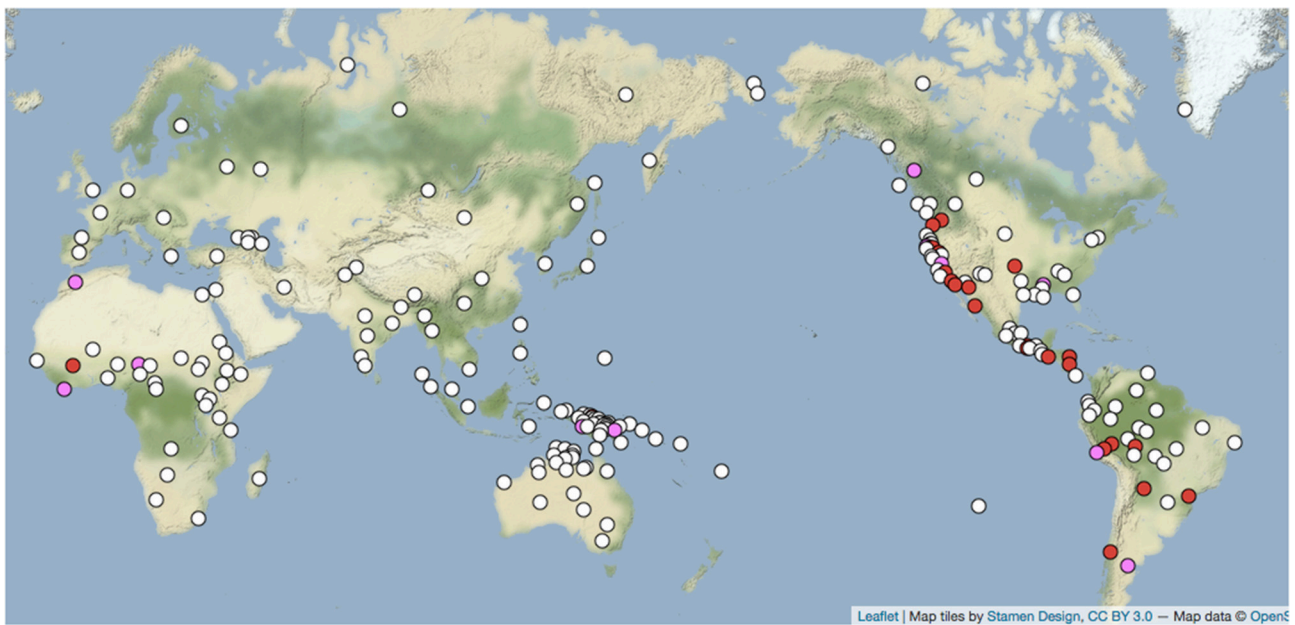
FIGURE 2 | *n-m* pronouns (*N* = 230). Red = *n-m* paradigm present; white = absent; pink = non-paradigm present (Nichols and Peterson, 2013a,b). http://wals.info/feature/137A#2/24.8/153.6.

appreciable extent only since the rise of food production. It is no accident that the *m-T* pronoun systems are thickly attested among the language families that have been involved with the rise and spread of nomadic pastoralism in Eurasia, where population growth, long-range client-patron and guest-host connections, and mobility were hallmarks of the societies and essential to

the spreads of their languages (see Anthony, 2007; Nichols and Rhodes, 2017). Prehistoric sociolinguistics is difficult to determine, but in the surviving fossil of the frontier of the Eurasian pastoral expansion, Khamnigan Mongol (Janhunen, 1990, 1991, 2005; Yu, 2011), there is easy code switching and apparently little language identity, though the languages remain

discrete and the mingling of forms in speech does not lead to language shift or mixed languages.

Linguistic symbiosis must have been an important part of the sociological and ethnic situations that have obtained at the frontiers of large language spread like those reviewed below, where new economic and social opportunities are constantly being created and and an enterprising individual can seize or create a new niche. In such undertakings clear communication is essential and the means of communication can be improvised. This situation is what Nichols and Rhodes (2017) call *frontier* conditions: an interface involving cultural, economic, and technological intermingling and offering prospects for entrepreneurship, intermediary roles, and trade management distant from the center of authority and prestige.

Sometimes a society at the frontier has taken advantage of this situation and its members have seized the roles of merchant, tinker, interpreter, diplomat, mercenary, camp follower, money changer, organizer, and/or others who mediate between the expanding culture and those beyond the frontier. Sometimes the frontier society melts into the expanding one, but sometimes its language spreads out far in advance of the expanding one. This is a *catalyst language* (Nichols and Rhodes, 2017), so called because the intermediary roles of its speakers assist or make possible the spread of the expanding language; examples include Ainu (catalyst for the Japanese Yayoi expansion), Tungusic (catalyst for the Mongolic northeastward expansion), the Mongolic family itself (catalyst for the northward expansion of Chinese empire), and several Turkic expansions (catalyst for the westward expansion of Chinese economic control) (see Janhunen, 2002, 2008, 2012). These languages have spread far from their points of origin (Ainu survived only at its own far northern frontier). These languages all bear markers of attractor spread, and it seems likely that symbiosis is a regular trait of catalyst languages.

Linguistic symbiosis overlaps in part with what Hill (2001a) calls a *distributed stance*: an outlook or attitude that tolerates variation on the part of others and generates variation in the speaker's own output. Its development is favored by contingent or unreliable access to vital resources and a combination of mobility and sparse population that causes individuals to grow up without a stable cohort of age mates and thus without a dialect identity. Hill's examples come from desert populations, where resource insecurity and high mobility are the rule. I count the elements of the distributed stance (contingent access to resources, weak or no dialect identity) as factors that contribute to symbiosis, together with the sociolinguistic properties identified here (diversity, neutral valuation of varieties). These are distinct from the factors discussed above that stabilize selected variants: dense networks, long connections, open connections, and any others that favor uptake and transmission of what would otherwise be one-off selections.

Factors that can be symptomatic of symbiosis where we have no direct evidence include archaeological, economic, and/or political-historical evidence for back-and-forth shifting of cultural or economic or political allegiance, bidirectional pattern copying (calquing, grammatical borrowing, etc.) in languages, direct or indirect evidence of catalyst function, and large-scale expansion of a language in a desert or high-latitude environment.

Below these are used as operative criteria for positing prehistoric symbiosis. The Khamnigan Mongol case mentioned above is important because it preserves linguistic symbiosis at what was the frontier of Mongol economic and linguistic expansion. It shows that identifying spreads of a certain type with symbiosis is a safe move.

METHOD AND SURVEY

The case for pronoun consonantism rests on an attractor that becomes relevant and selection that becomes operative only in the right sociolinguistic and demographic context. Though the geography supporting this account is compelling, it is circumstantial evidence. Actually testing the claims is problematic because pronoun consonantism is difficult to work with statistically: the range of options is small, essentially just 'yes' vs. 'no' (i.e., /m/ or no /m/) per language and per pronominal category; conforming languages are a minority even in those continents where they are most frequent; the pronominal context is defined as a search through options (independent pronouns, verb agreement affixes, possessive affixes, etc.), the generic T for the Eurasian second person is also a set of options, and casting about through options inflates the possibility of success. Therefore, what follows seeks evidence of sociolinguistic and sociological conditioning in a more tractable part of grammar: the causative alternation.

The causative alternation is illustrated in the verb pairs shown in **Table 1**. Each pair consists of a non-causal verb ('laugh', 'die', 'sit', etc.) and the corresponding causal ('make laugh', 'kill', 'seat', etc.), whose semantics consists of the non-causal predicate plus causation: 'frighten, scare' means 'cause to fear or be afraid' and causal (transitive) 'break' means 'cause to break or get broken'. The semantic relationship of non-causal and causal is alike for each pair, but the formal structures differ, and the point at issue here is how, grammatically and structurally, the two verbs in each pair are related. The causal can be derived from the non-causal, as in Estonian 'fear': 'frighten' or Kazakh 'break'; the non-causal can be derived as in Macedonian 'fear' or Czech and Spanish 'break'; both can be derived, as in Aymara 'break'; completely different verbs can be used, as in Norwegian, Catalan, and Russian 'fear': 'scare'; or the two forms can be identical as in German 'break' (and English *break* and many other verbs).

I used a set of 18 such pairs (Nichols et al., 2004), assembled from dictionaries and/or consultation with native speakers and language experts, surveyed across 207 languages, about half of which figure centrally here. The pairs are listed in **Table 2**. Most of the counts and graphs below use only the nine such pairs that typically have an animate undergoer (e.g., 'fear', 'angry', 'sit'), as these tend to be more stable over time.

The formal relationships between the two members of the pair can be reduced to three basics: the causal form is derived; the non-causal is derived; they have the same vs. different roots. Languages are typologized by the percent of the pairs exhibiting those three basic types, and what primarily figures here is the percent that use causativization, i.e., derivation of the causal from the non-causal (as in Estonian 'fear: scare' and Kazakh 'break' in

TABLE 1 | Some causal/non-causal verb pairs.

	Non-causal	Causal
	'fear, be afraid'	'frighten, scare'
Macedonian	**se** plaši	plaši
Russian	bojat'-**sja**	pugat'
Estonian	hirmuma	hirmu-**ta**-ma
Norwegian	frykte	skremme
Catalan	témer, tenir por	espantar, esporuguir
	'break' (intransitive)	'break' (transitive)
Czech	lomit **se**	lomit
Spanish	romper-**se**	romper
Aymara	p'aki-**si**-	p'aki-**ña**-
Kazakh	synu	syn-**dyru**
German	brechen	brechen

The suffix or other morphology deriving one from the other is boldface. Hyphens are for clarity (they are not orthographic in the languages).

TABLE 2 | Surveyed causal-noncausal verb pairs.

Animate		Inanimate	
Non-causal	Causal	Non-causal	Causal
laugh	make laugh, amuse	(come to) boil	(bring to) boil
die	kill	burn, catch fire	burn, set afire
sit	seat	break	break
eat	feed	open	open
learn	teach	be/get dry	dry (off, out)
see	show	straight	straighten (out)
be/get angry	anger, make angry	hang, dangle	hang (up)
fear, afraid	scare, frighten	turn over	turn over
hide	hide	fall	drop

Animate, inanimate = typically undergone by animate or inanimate entity.

Table 1). Of interest here is preferred causativization, i.e., above-mean percentage of pairs in which the causal is derived. (The mean for the animate set of verbs is 54%, or just under five pairs.) As shown in **Figure 3**, high and low percentages are not evenly distributed worldwide: very few pairs use causativization in Europe (blue symbols) and many in northern Asia and North America. (A sparser but essentially similar picture emerges if they are plotted as ±1 standard deviation from the mean). What predominates in Europe is decausativization, where the non-causal is derived from the causal, as in Spanish *romperse* 'break', Macedonian *se plaši*, and others. (In these two examples the verbs are reflexive, a derivational type that is common in Europe but infrequent elsewhere).

The hypothesis here is that, of the possible realizations of the causative alternation, causativization is an attractor that is selected in symbiosis. There seem to be two reasons why causativization is an attractor. First, it aligns with semantics. In a verb pair like 'sit' and 'seat', 'sit' involves only a subject and an

activity or position, while 'seat' adds an agent and semantics of causation. If 'seat' is derived from 'sit', the morphological form echoes the semantics and the cognitive complexity[1]. Second, most languages have a ready source of potential causativizing morphemes: verbs like 'make' function easily to create phrases with causative semantics (e.g., *That always makes me laugh*, where *make laugh* is well on the way to being lexicalized as a discontinuous causative verb). In very many languages the causativizing morphology is in fact a reduced form of a verb like 'make' that has become a causativizing suffix. There is no comparably ready source for decausativization, which involves removal of the agent and the agency. Reflexive pronouns derive non-causatives in many European languages, but this is an idiosyncratic construction, not common outside of Europe, with no correlation to the semantics: 'get angry' may look literally like 'make oneself angry', but that is not at all the meaning.

That causativization is associated with symbiosis was first suggested (not using the term *symbiosis*, and describing the sociolinguistics differently), in Nichols (2011). Here I draw on expanded data, improved coding, and improved understanding of the sociolinguistics (Grünthal and Nichols, in press) to give firmer results from more parts of the world. Despite these advances, this is a hypothesis-raising study, using a database originally designed for other purposes, which uses a standard sampling approach that strives for independence of languages by choosing only one per family or major branch, while what is needed for hypothesis testing is dense coverage of families. The goal here is to determine whether such further testing would be worthwhile.

RESULTS

This section reports seven case studies supporting an association of causativization with symbiosis and/or frontier conditions.

The Northeastern Caucasus

The first case study is what I call the Avar sphere in the eastern Caucasus, from the middle ages to the Russian conquest of the Caucasus in the mid nineteenth century. It involves mostly protohistorical and early historical spreads and a sociolinguistic situation that was viable until the mid-twentieth century and is still in evidence, so we are on firm ground in describing it[2]. At the time of the conquest the Avar khanate dominated the north slope of the eastern Caucasus (a.k.a. Daghestan). The Avar khanate was the continuation of the Sarir Kingdom, which arose c. 800 BCE (and changed its name to Avar on converting to Islam)[3]. Prior to the Russian conquest, the Avar khanate was an economic and

---

[1] A case of iconicity; see e.g., (Haiman, 1985) and much other work.
[2] Sources for the historical and sociolinguistic description in the next paragraphs include (Lavrov, 1953; Volkova, 1967; Wixman, 1980; Aglarov, 1988, 1994, 2002; Nichols, 2005, 2016; Karpov and Kapustina, 2011; Dobrushina, 2013). Here and below, for each section I cite sources used and a few well-known overviews, selecting from a very large literature on each topic.
[3] They adopted the name of an important pre-Hunnish nomadic society from the eastern steppe. The ethnonym had in turn been taken on by the Avars who ruled central Europe from the 6th to 8th centuries, attacking Byzantium and invading the Balkan peninsula. Apart from the ethnonym there is no connection between the Caucasian Avars and the other two groups.

FIGURE 3 | Distribution of high and low frequencies of causativization in the 9-pair verb list (*N* = 200). Black = above mean; blue = below mean (Mean = 54%).
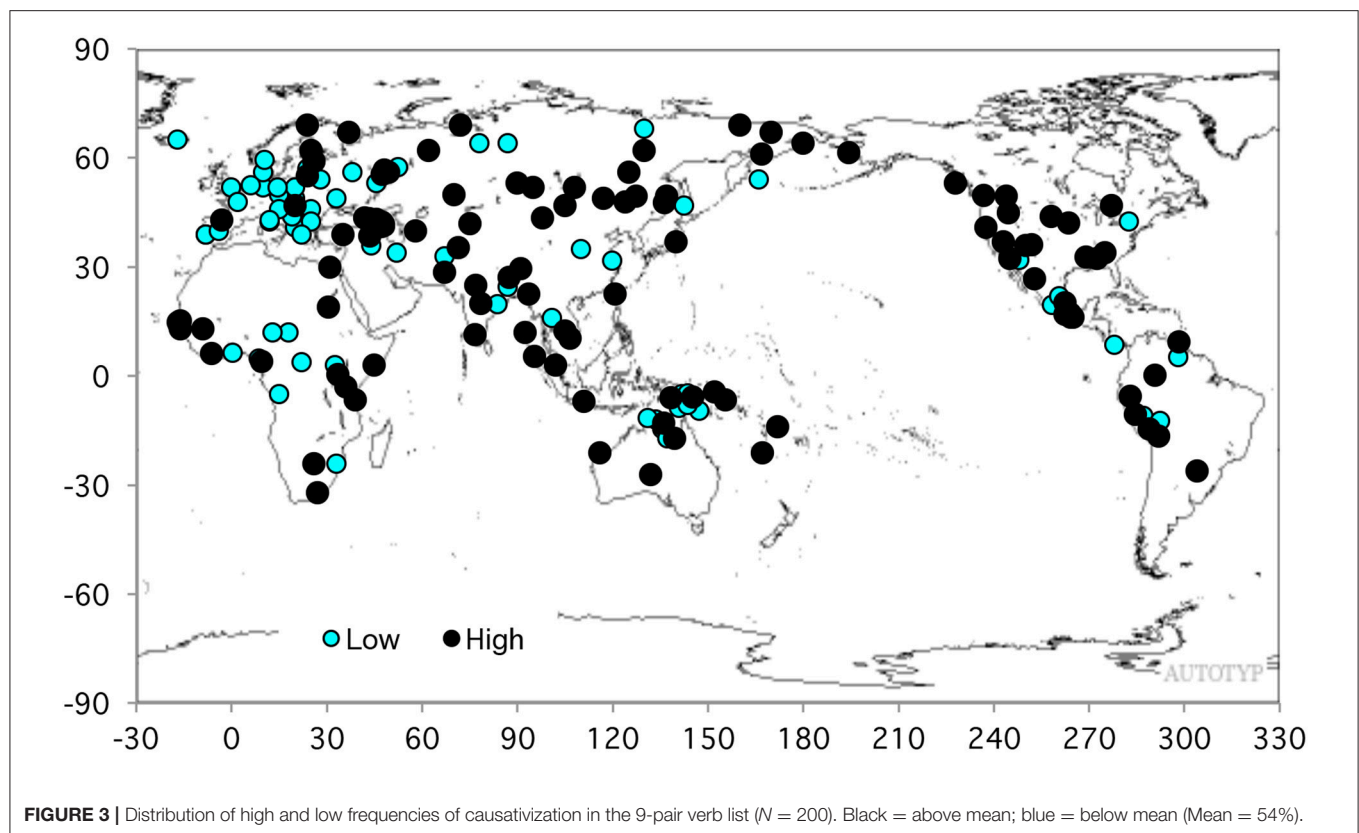
cultural power and a military confederation encompassing a large number of small highland and foothill city-states located along highland watercourses, chiefly the Avar Koisu and Andi Koisu rivers, their tributaries, and their lowland confluence in the Sulak (which flows to the Caspian Sea), whence there were connections to Silk Road ports and cities. The city-states were independent and could join or leave the confederacy at will; mostly they joined and remained, and while they were members their young men served in the Avar army, where Avar served as language of command. For millennia, since the adoption of food production, Daghestanian highland societies were half transhumant, with the working-age male population spending the winter half of the year in the lowlands tending herds in winter pastures and/or taking seasonal employment or owning businesses in lowland cities. The non-transhumant female part of the population traveled downhill regularly to the market towns or the larger lowland markets. Roads ran along river canyons, so the Avar Koisu and Andi Koisu roads funneled all such traffic to the confluence, where the Avar capital Khunzakh was strategically located in an ideal position for trade and taxation.

For these essential economic contacts highlanders needed to know foothill and lowland languages, but not vice versa; lowlanders had no need to travel uphill, rarely did so, and did not learn highland languages. As a result, the linguistic situation in Daghestan involved massive local asymmetrical vertical bilingualism and multilingualism with an overlay of Avar as an always available contact language. In mountain areas, languages generally spread uphill from the economically better-connected and more densely populated lowlands to the more isolated highlands, and the vertical bilingualism of Daghestan strengthened that tendency: Avar or the language of a market town, known to many people from towns above it, could come to be used in a higher town as well. The language family spoken in most of the eastern Caucasus is Nakh-Daghestanian, an old and much-differentiated family, and as a result of repeated uphill spreading the daughter branches, most of which are of about Romance-like or Slavic-like diversity and apparent age (so ~2,000 years), extend from lowlands to the highest inhabited levels. The archaeological age of villages, where known, is generally well over 2,000 years. This gives reason to reconstruct repeat uphill spreads of Nakh-Daghestanian branch ancestors, probably accompanying periods of economic prosperity in the lowlands, ever since the Nakh-Daghestanian dispersal several millennia ago.

The Avar language is now spoken along the Sulak, all along the Avar Koisu and beyond, spilling over the crest to northern Azerbaijan, and along the lower Andi Koisu with occasional outliers in the highlands. Those are outliers of lowland northern dialects, so the Avar dialect diversity there is not great. The diversity of Avar along the Avar Koisu is greater, but still all dialects are said to be more or less mutually intelligible. This suggests that the Avar spread began some 500 years ago; more than about 500 years generally spells loss of ready mutual intelligibility (of course all such figures are very approximate). The Andic subbranch, the closest sister to Avar, extends above Avar along the Andi Koisu, with two outliers along the Avar

Koisu. The Andic languages are closely related but generally not mutually intelligible, and this together with the more uphill position indicates that the Andic uphill spread was somewhat earlier than the Avar one. Andic place names are found Avar lands along the lower Andi Koisu and the Sulak, testifying to Avar expansion there. The Andi, a large foothill Andic group for whose language the branch is named, were economically powerful until the Russian conquest, collecting taxes along the Andi Koisu and into the Chechen lands in the west, and monopolizing the lucrative trade in the Caucasian *burka*, the felt coat worn by highland shepherds and the czarist Russian army. The Andi and Avar were rivals for political and economic power (the Andi won their last important battle against the Avars in the late seventeenth century).

The more distantly related Tsezic languages are uphill of the Andic ones along the Andi Koisu, with two outliers on upper tributaries to the Avar Koisu. They evidently represent a still earlier spread, whose lower languages have since shifted to Andic much as Andic has shifted to Avar in the lowlands.

Repeated uphill spreads would mean absorption of highland populations by language shift, with adults learning the spreading language. This should bring about decomplexification of the language, and indeed Avar and the Andic languages show considerable decomplexification compared to most other Nakh-Daghestanian languages. In addition, however, there is evidence of linguistic symbiosis. The mix of spreads described above implies oscillating dominance of Avar and ancestral Andic, depending on political and economic fortunes in the lowlands. In addition, there was no standard language and no source of linguistic prestige apart from market and inter-ethnic usefulness (to the extent that any language was prestigious it was Arabic, and that only after the conversion to Islam). Another factor favoring linguistic symbiosis was the mobility of the transhumant societies. In addition, there was little or no language ideology or identity; the foci of identity were clan, village, and in recent centuries sometimes religion. Social networks were dense but open, with many long-range contacts both uphill and downhill. The pan-Daghestanian term for a host in a guest-host relationship is *kunak,* and such connections were sought and valued, especially at long distances or when they involved a well-placed lowlander.

The interaction of Avar with local Andic and Tsezic languages was a historically documented matter of symbiosis, with free interjection of Avar words into the local language. Some such words have by now stabilized as loans but some are one-time code switching. e.g., in Hinuq (Tsezic), Avar adjectives "constitute an open class in the sense that whenever a Hinuq speaker wants to use an adjective and does not find a Hinuq term (s)he uses an Avar term" (Forker, 2013, p. 170). Pronouns in Avar, Andic, and Tsezic languages are strongly assonant, using both rhyme and alliteration, much of it innovative compared to Proto-Nakh-Daghestanian (Nichols, 2012b): examples are in **Table 3**.

These languages make extensive use of causativization. **Table 4** shows Avar verb pairs from the list above (Creissels, 2014). **Table 5** shows percentages of causativization in the Avar sphere and elsewhere in the Caucasus. Percentages decrease with distance from Avar, both within the Avar sphere and between

TABLE 3 | Avar-Andic-Tsezic pronouns.

	1sg	2sg	Inclusive	1pl	2pl
Avar	dun	mun	nił	niž	nuž
Godoberi (Andic)	den	min	iλé	išše	bitté
Hunzib (Tsezic)	də	mə		ile	miže

Nominative case only. 1sg = first person singular, 2pl = second person plural, etc.

TABLE 4 | Avar verb pairs.

'laugh', 'make laugh'	beł-ize	beł-iz-**ab**-ize
'sit', 'seat'	k'us-ize	k'us-iz-**ab**-ize
'eat', 'feed'	k'wan-aze	k'wan-az-**ab**-ize
'see', 'show'	bix-ize	bix-iz-**ab**-ize
'get angry', 'make angry'	ccin+daxx-ine	ccin+daxx-in-**ab**-ize
'fear', 'frighten'	ħinq'-ize	ħinq'-iz-**ab**-ize
'hide'	baxč-ize	baxč-ize;
		baxč-iz-**ab**-ize

Causativizing suffix bold. Hyphens (not orthographic) segment off the infinitive ending and the causative suffixx.

TABLE 5 | Eastern and central Caucasus: Proportion of the nine verb pairs that use causativization.

Avar sphere:	Avar	0.72
	Akhvakh	0.89
	Karata	0.89
	Bagwalal	0.78
	Godoberi	0.89
	Tsez	0.78
	Hinuq	0.67
	Hunzib	0.50
Nearby:	Lak	0.50
	Dargwa	0.67
	Chechen	0.67
Distant:	Ingush	0.56
	Archi	0.39
	Udi	0.56
	Tabassaran	0.25
	Rutul	0.11
	Tsakhur	0.25
	Lezgi	0.56

Languages are listed within groups in order of increasing distance from the Avar capital (in the Avar sphere this amounts to increasing altitude). Hunzib is peripheral to the Avar sphere; its winter pastures and other connections were in Georgia to the south.

groups. The languages shown cover the Avar sphere, other languages of the eastern Caucasus (Lak, Dargwa, Lezgi, Tsakhur, etc.), and languages to the west (Chechen, Ingush), and they include languages on both the north and south slopes. The conclusion is that, where symbiosis has been most common, causativization is most frequent. No other known factor accounts for the frequency of causativization within Nakh-Daghestan.

The Eastern Steppe

The eastern, or Mongolian, steppe is the band of grassland extending along the south slope of the southern Siberian mountains from the Tien-shan and Altai to north central China[4]. Here, from the rise of mining and metallurgy in the Altai area and the rise of imperial power in China, successive nomadic pastoral tribes, kingdoms, and states have formed and their languages have spread far, chiefly westward, along the steppe and in Central Asia. The spreads have generally involved conquest of rulers and language shift by much of the population, with the result that the languages that have undergone large spreads are considerably decomplexified and regularized in their grammars and lexicons. There has also been a good deal of borrowing and grammatical convergence among them: the modern Turkic, Mongolic, and Tungusic languages in particular are strikingly similar in their overall structures. In historical and protohistorical times the various expansions have created frontier conditions along the expanding periphery, and there is firm evidence of linguistic symbiosis in the surviving Khamnigan Mongol-Evenki situation mentioned above. Zgusta (2015, pp.104–164) gives evidence of frequent movement and realignment of ethnic groups along the lower Amur that appear likely to have involved symbiosis among different Tungusic languages and with unknown pre-Tungusic languages.

The known language families involved in these spreads, in chronological order of earliest importance, are eastern Iranian (Indo-European), Turkic, and Mongolic. Other, poorly attested languages are likely to have been involved in the early stages, perhaps including an ancestral Yeniseian language (the family is historically attested only along the upper and middle Yenisei, with Ket on the middle Yenisei the only survivor). The medieval and later Turkic and Mongolic spreads are historically and ethnographically well described and some of the sociolinguistics is attested or reconstructable. The two families both originated in or near today's northern Mongolia and seem to have had connections to both the Altai metallurgical center and the steppe nomadic economies. Between these two families, locally and in general along the frontier, there was some history of back-and-forth shifting, each functioning as catalyst to the other at least some of the time. Before the rise of Genghis Khan Mongolic was spreading at westward and absorbing Turkic speakers (Janhunen, 2008). During the Mongol expansion, Turkic speakers whose tribes and states had been incorporated into the Mongol empire were so much more numerous than Mongols that, although Mongolian was the language of command, it was Turkic rather than Mongolic speech that was chiefly spread across Central Asia and the central and western steppe.

The nomadic pastoral economy, which propelled the spreads, fostered mobility and contacts with other peoples and languages around the steppe periphery: hunter-gatherers in the north who traded in furs; miners and metalworkers in the Altai area;

TABLE 6 | The Turko-Mongol steppe and neighbor Tungusic: Proportion of the nine verb pairs that use causativization.

Turkic	Tatar	0.88
	Kazakh	1.00
	Turkmen	1.00
	Chuvash	1.00
	Yakut	0.88
	Uighur	0.75
	Kirgiz	0.89
	Tuvan	0.88
	Khakas	0.67
Mongolic	Khalkha	0.75
	Buriat	1.00
	Dagur*	0.67 (only six pairs found)
	Khamnigan*	0.67 (only three pairs found)
Tungusic	Manchu	0.88
	Nanai	0.78
	Udehe	0.83
	Evenki	0.63
	Even	0.50

*Within each family, languages closer (or historically closer) to the centers of expansion are listed first. *Proportions not accurate as not all of the nine pairs could be found.*

urban centers in China and Central Asia; various trade outposts. Language identity among nomads appears not to have been strong, and there were no standard or written languages and no durable prestige language. Clan and client-patron relations were primary. In addition to the decomplexification and regularization that testifies to histories of language shift, the languages of both families and also the neighboring Tungusic family to the east have pronoun systems with rhyme, alliteration, and the $m\text{-}T$ type that bespeak symbiosis. Causativization is high overall, highest in Turkic, which has the longest history of nomadic spreading, and least high in Tungusic, a family of languages spoken by settled semi-agriculturalists in northern China and Korea and spread in Siberia by reindeer herders (**Table 6**). Within each family, languages closest to the center of symbiosis have the highest percentages, supporting the correlation of symbiosis with causativization.

Uralic

The Uralic family stretches across northwestern and north central Eurasia, from western Norway beyond the Yenisei to the eastern Taimyr Peninsula, a distribution that was continuous down to about the southern limit of the northern forest zone until the relatively recent northward expansions of the Scandinavian languages and Russian[5]. Testifying to its long presence in

---

[4]Sources on the history and sociolinguistics for this section include (Krader, 1963; Barfield, 1989; Chernykh, 1992, 2009; Khazanov, 1994; Janhunen, 1996, 2008, 2012; Pulleyblank, 2000; Schönig, 2003; Anthony, 2007; Kohl, 2007; Di Cosmo et al., 2009; Hanks, 2010; Golden, 2011; Frachetti, 2012; Werner, 2014; Vovin et al., 2016; Nichols and Rhodes, 2017).

[5]Map of modern distributions: https://en.wikipedia.org/wiki/Uralic_languages#/media/File:Linguistic_map_of_the_Uralic_languages_(en).png. This is the visually clearest map I have found, but the subgroupings listed are not all correct. Current classification: http://www.helsinki.fi/~tasalmin/fu.html Map showing branch homelands: http://www.helsinki.fi/~tasalmin/Uralic.jpg. Other sources for this section: (Sinor, 1988, 1990; Napol'skix, 1997; Abondolo, 1998; Anthony, 2007; Grünthal and Petri, 2012; Holopainen, 2017).

TABLE 7 | Uralic languages: Proportion of the nine verb pairs that use causativization.

North Saami	0.78
Kildin Saami	0.83
Finnish	0.89
Estonian	0.72
Erzya Mordvin	0.50
Mari	0.61
Udmurt	0.44
Hungarian	0.67
Mansi	0.61
Khanty	0.44
Tundra Nenets	0.78

Languages are ordered from west to east (Hungarian is placed with the eastern languages where it originated).
Hungarian has a fairly high level of causativization, the reasons for which are not examined here. Hungarian has not undergone a major spread; it moved from southern Siberia to central Europe by migration, keeping its language (and apparently its ethnic and language identity) through several centuries as an enclave in a Turkic confederation and then in the Iranian-speaking western steppe population of the post-Roman centuries.

TABLE 8 | Indo-European: Proportion of the nine verb pairs that use causativization.

Latin	0.14
Albanian	0.11
Greek (modern)	0.22
W. Armenian	0.83
Germanic mean	0.08
Romance mean	0.26
Slavic mean	0.11
Baltic mean	0.53
Indo-Iranian:	
Kurdish	0.28
Ossetic	0.38
Persian	0.78
Pashto	0.78
Waigali	0.40 (Only five pairs found)
Palula	1.00
Hindi	1.00

Ordering of Indo-Iranian is west to east.

the region and the momentum of its spread, the family has representatives in the three linguistically diverse accretion zones to the south of its main range: the eastern Circum-Baltic area (Estonian and several small languages), the middle Volga (Erzya and Moksha Mordvin, Mari), and south central Siberia (Samoyedic languages in the Altai mountains, now extinct). These zones are populated by remnant languages from other prehistoric spreads. Most of the westward spread of Uralic postdates, and was probably triggered by, the Indo-Iranian expansion c. 4,000 years ago from what is now northeastern Kazakhstan (a number of early Iranian or Indo-Iranian loans entered the Proto-Finno-Ugric branch of Uralic at that time). The westernmost extension—the spread of Finnic into Finland and Saami into Scandinavia—occurred less than 1,000 years ago, before which first ancestral Saami, then early Finnic, had been adopted by agricultural people in the east Baltic area (probably Germanic- and Baltic-speaking; both of these are Indo-European branches). Spreads of North Saami within Saami (in Scandinavia) and Nenets within Samoyedic (in Siberia) are also recent and involved the spread of reindeer herding.

These spreads were at high latitudes and involved sparse and mobile populations (even the agriculturalists of southern Finland were relatively mobile and sparse, relying on slash-and-burn methods and moving to new fields from time to time). The known large spreads—Saami, Finnic, Tundra Nenets—can therefore be assumed to have involved symbiosis, and it is these large spreading languages that have the highest proportions of causativization (**Table 7**), supporting the hypothesis.

Indo-European
The Indo-European family has a long history of spreads of types that should not favor symbiosis: expansions of state and imperial languages, spreads of written languages, and spreads driven by economic, technological, and/or political advantage (the earliest Indo-European spreads must have been of these latter types: see Mallory, 1989; Mallory and Adams, 1997; Anthony, 2007).

The very earliest spreads, which brought the Anatolian languages (Hittite and its sisters) to what is now Turkey and the ancestors of at least Greek, Latin, and the Celtic languages to Europe, may have been migrations with formation of local outposts (Anthony, 2007) that only later grew by language shift, as was happening with Latin in early historical times; or they may have begun with invasion, conquest, and wholescale language and culture replacement in southeastern Europe (Parpola, 2012). The migration-and-outpost scenario could have produced occasional local cases of symbiosis, but more probably the outpost languages were economically prestigious and remained discrete. The invasion scenario is unlikely to have produced symbiosis.

What is striking about Indo-European is its low overall frequency of causativization (**Table 8**); the European cluster of low causativization in **Figure 3** is mostly Indo-European languages. For the modern languages the structural reason for this is that their most common kind of pairing derives the non-causal from the causal by reflexivization (see again **Table 1**). Reflexivization is a post-classical development: absent from Greek, beginning to occur in Latin, halfway developed in Old Church Slavic (ninth century), and evidently it spread between early Romance, Germanic, and Slavic by calquing[6].

Table 8 shows proportions of causativization in some Indo-European languages and branches. Differences within and beween European branches have no obvious cause (they have not been studied closely for this survey). Comparison across the whole family reveals three general principles. First, contact with causativizing languages can increase causativization; the clear example is Western Armenian, with Turkish and Persian contact

---

[6]For general aspects of verb root and stem structure in Indo-European see e.g., (Rix, 2001; Jasanoff, 2003; Fortson, 2010).

effects[7]. Second, light verb constructions, common in Iranian languages, lower the frequency of causativization. An example of a light verb construction is Tajik *xušk šudan* (dry become) 'dry out, dry up, get dry': *xušk kardan* (dry make) 'dry off, dry (something)'; or English *fall asleep, go to sleep: put to sleep* or *catch (on) fire: set on fire*. These consist of an element with lexical meaning (*xušk* 'dry', *(a)sleep, on fire/afire*) and an auxiliary which contributes little lexical meaning but carries tense and agreement and determines the syntactic valence of the construction. Third, causativization levels are high in the Indo-Iranian branch, especially in its eastern representatives. This branch spread rapidly across the entire steppe about 4,000 years ago, propelled by development of metallurgy and metalworking in the Ural area and military advances including chariot technology. Speakers of early Indo-Iranian came to dominate, and finally absorbed, the the western Central Asian oasis civilizations of the Bactria-Margiana Archaeological Complex (Hiebert, 1994; Witzel, 2003), and the entire branch shows contact effects from a Dravidian or Dravidian-like language (the Dravidian family is indigenous to India) usually attributed to that episode. The Indic branch shows further contact effects from Dravidian. The Dravidian languages have high proportions of causativization, and it is plausible, though far from proven, that the Indo-Iranian high causativization results from these contacts. Whether any of these contacts could have produced symbiosis is a different question. Military conquest (as across the steppe) and economic dominance (as in Central Asia and later in northwestern India) usually do not, but substrata can, and certainly the deep intermingling of Indo-European and Dravidian-like or Indic-like myth and religion in Vedic Sanskrit suggests something like symbiosis[8].

Therefore it is at least possible that the high proportions of causativization in Indo-Iranian result from symbiosis. If not symbiosis, they may result from ordinary close contact involving calquing. Western Central Asia is desert and sparsely populated–except for the oasis cities, which have large and dense populations, and were the main target of Indo-Iranian dominance. Therefore the Indo-Iranian spread to the cities was a language spread through a dense population.

Uto-Aztecan

The Uto-Aztecan family, about 5,000 years old, ranges north-south from Shoshoni in the northern U.S. Great Basin to Nahuatl varieties throughout Mexico and an outlier in Pipil (Nicaragua, a former Aztec garrison)[9]. The family probably originated in or near Mexico, i.e. in the southern part of its range, and spread northward with or in advance of the northward advance of agriculture. Much later came the Aztec imperial spread. Daughter languages are spoken mostly by agriculturalists or (in the Great

TABLE 9 | Uto-Aztecan languages: Proportion of the nine verb pairs that use causativization.

Tümpisa Shoshone	0.89	
Hopi	0.67	
Tohono O'odham	0.38	
Raramuri	0.78	
Huastec Nahuatl	0.60	0.75 (different analyses)

Languages are listed from north to south.

Basin) hunter-gatherers focusing on plant-based and especially seed-grinding subsistence. The two major spreads in the family are the spread of Nahuatl with the Aztec empire and the Spanish conquest (which used classical Nahuatl as official contact language), and the spread of the Numic branch through the Great Basin after a severe drought in the middle ages destroyed the early agricultural economy there.

A small sample of Uto-Aztecan languages (**Table 9**) gives some support to the correlation of causativization with symbiosis, with mobility and large spreads implying symbiosis. Tümpisa Shoshone, with the highest proportion of causativization, represents the highly mobile and sparse populations of the Great Basin which gave Hill (2001a) (drawing on work on Shoshoni by Wick Miller) her example of a society without stable groups of age mates and hence with minimal dialect identity. The others are settled agriculturalists; the Tohono O'odham were partly transhumant between summer and winter water sources (the transhumant population, inhabiting the driest part of the range, gave Hill her example of contingent access to resources and her documentation of variability in such populations).

Austronesian

The widespread Austronesian family originated on or near Taiwan some 6,000 years ago and spread through Island Southeast Asia and thence to near and far Oceania[10]. The spread to New Guinea and nearby islands involved coastal or offshore settlement and usually intensive contact and intermarriage as indicated by grammatical and lexical influence and genetic evidence. The spread to Micronesia and Polynesia involved colonization of previously uninhabited islands. As a result of this long history of migration the family is very large, with about 1,000 daughter languages. The eight languages in **Table 10**, representing all the Austronesian languages in my database, are a grossly inadequate sample of this diversity, but they cover the geographical range and some of the branches. They give some support to the hypothesis. High proportions might be expected in languages of Island Southeast Asia, where pre-Austronesian populations were absorbed in the early stages of spreading, populations are dense, and there is a history of statehood, which makes changing alliances and oscillating dominance plausible. In New Guinea and the nearby large islands, Austronesian languages colonized coastal areas, occupied a maritime economic niche, and

---

[7]It can also retard loss of causativization: Romani varieties in Europe and nearby have generally lost the inherited Indic causative morphology except for varieties in contact with Turkic (Adamou, 2012; E. Adamou p.c.).

[8]For the Indo-Iranian takeover of the Central Asian civilizations see (Witzel, 2003; Anthony, 2007; Frachetti, 2008; Kuzmina, 2008); for the civilization (Hiebert, 1994).

[9]Sources for this section: (Fowler, 1972; Miller, 1983; Madsen and Rhode, 1994; Hill, 2001b, 2010; Kemp et al., 2010; Golla, 2011; Merrill, 2012).

[10]See e.g., (Pawley and Ross, 1993; Friedlaender et al., 2008; Ross et al., 2008; Blust, 2009; Kirch, 2010; Donohue and Denham, 2012; Bellwood, 2017).

TABLE 10 | Austronesian languages: Proportions of the nine verb pairs that use causativization, and broad locations.

Paiwan	0.61	(Taiwan)
Malay	0.78	(Island Southeast Asia)
Acehnese	1.00	(Island Southeast Asia)
Javanese	0.60	(Island Southeast Asia)
Tolai	0.67	(Coastal New Guinea)
Tawala	0.50	(Coastal New Guinea)
Drehu	0.57	(Remote Oceania)
Samoan	0.71	(Remote Oceania)

Languages are listed by increasing distance from the homeland.

interacted and intermarried with indigenous horticulturalists. The outcome is sometimes linguistically mixed households with multilingualism beginning in childhood, and grammatical convergence, but languages that remain discrete because they are associated with descent groups. If the situation described by Ross (1996) for north coastal New Guinea is at all common, the distinction of ethnic and inter-ethnic language and the different directions of phonological and lexicosemantic influence show that the languages are ideologically distinct and not sociolinguistically neutral. Symbiosis should not occur in such situations and the proportion of causativization should not be high. In remote Oceania, where languages mostly occupy small islands that do not foster diversity and offer few day-to-day contacts with other languages, symbiosis should not be common and causativization rates should not be high. In **Table 10**, the highest proportions are indeed found in Island Southeast Asia (Malay, Acehnese) and lower proportions are found elsewhere, supporting the hypothesis or at least not undermining it, but a much larger survey and community-specific accounts of sociolinguistics are needed to draw any firm conclusions. Causativization, and specific causative morphology, are ancestral in Austronesian, and here it is the retention of an attractor state that is relevant. Retention rates are lower in places where symbiosis is unlikely to have occurred, higher where it might have occurred.

The Balkan Sprachbund

The Balkan Sprachbund, or Balkan language area, in the southern part of the Balkan Peninsula, is the exception that proves the rule. The languages of the Sprachbund are Greek, Albanian, Macedonian, Bulgarian, southeastern (Torlak) Serbian, Arumanian (Balkan Romanian), and Romani; Turkish has been present for several centuries but does not participate in the Sprachbund. The Sprachbund is a textbook case of a linguistic area involving contact, multilingualism, and grammatical convergence[11]. Causativization is low in the Balkan Sprachbund, not appreciably different from the rest of western Europe. There has been a good deal of lexical borrowing, extensive grammatical convergence, but no selection of the

[11]Overviews of the Balkan area include (Joseph, 1983; Thomason, 2001; Aronson, 2008; Friedman, 2011).

attractors covered here[12]. The evident reason is that Balkan sociolinguistics is quite different from symbiosis. There is multilingualism beginning in childhood, clear language identity, language discreteness, and low tolerance for mixing and code switching. All of the languages except for Romani are national languages with written standards that further inhibit selection and mixture (though Arumanian and Torlak Serbian are quite different from the national standards). Symbiosis and selection are not expected in this situation and they have apparently not occurred in the Balkan Sprachbund.

DISCUSSION

Non-linguistic causation, in the domain studied here, is evidently for real, but it is not a simple cause-and-effect matter. We need a three-factor model. First, alignment with event-structure semantics and the ready availability of sources of causativizing morphemes make causativization a potential attractor. Second, the sociolinguistics of symbiosis lets selection operate. Third, the right combination of environmental and sociolinguistic conditions lets selected variants be propagated and take root. The environmental factors include deserts and high latitudes, and it should be emphasized again that the actual cause is not these geophysical environments but the sparse populations they host.

The *m-T* and *n-m* pronoun patterns used as introductory illustration have striking geographical distributions: well attested in one macrocontinent and rare elsewhere. Causativization is less black-and-white, found to appreciable extents everywhere except Europe, and it is more frequent worldwide. Some of the difference may be in how the two are measured (causativization is sought over a larger wordlist than the basic first and second person pronouns), but the main factor must be ease of selection: borrowing of pronouns is generally inhibited, but pattern copying of verb derivational structure is more readily tolerated (as shown by accommodation of derivational types to those of neighboring languages, discussed for Western Armenian and Romani).

Language families vary in their mean frequencies of causativization, and most of that variation reflects not the non-linguistic causes described here but relatively stable family traits. Therefore the effect of symbiosis and the relevant environmental factors is to raise or lower proportions of causativization relative to family means. There is no absolute threshold above which symbiosis can be confidently posited and below which it cannot.

Symbiosis is a product of intense contact, but not all intense contact produces symbiosis. The Balkan Sprachbund is the clearest case of intense contact without symbiosis. Other areas known to have language identity, linguistic discreteness,

[12]A prominent thread in Balkanist literature describes such Balkan traits as loss of case inflection and some affixal tense-mood inflection and their replacement by clitics and particles as an increase in analyticity and thereby in transparency, a change that also favors convergence by making grammatical formatives easily calquable (e.g., Lindstedt, 2000). This is a form of simplification and a favored outcome of contact-induced change, but favored outcomes is a broader notion than attractor as defined here.

and grammatical but not lexical convergence include northern Australia and much of Amazonia, where societies and languages are smaller and languages are mostly unwritten but the sociolinguistics and striking combination of shared grammar and discrete lexicons are also present. Another kind of contact situation without symbiosis is asymmetrical dominance, where one language is more widely used or valued than another (for reasons such as political dominance, national language used in education vs. minority language restricted to home use, economic usefulness, inter-ethnic language, educational policy, etc.), a situation that often leads to language shift and drives the non-dominant language into extinction. In the great variety of language contact scenarios and sociolinguistic situations, symbiosis is not particularly common, but the results presented here show that it does occur and can be identified with reasonable reliability, even prehistorically.

Such are the non-linguistic causes that nudge languages toward greater use of causativization. Given these promising results, work about to begin will survey more families, more of their daughter languages, and more structural variables, and will cover sociolinguistic, ethnographic, and demographic factors in more depth.

AUTHOR CONTRIBUTIONS

JN designed the study, did some of the data collection and analysis, carried out most of the analysis, and wrote the article.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abondolo, D. (ed.). (1998). *The Uralic Languages*. London; New York, NY: Routledge.

Adamou, E. (ed.). (2012). "Verb morphologies in contact: Evidence from the Balkan area," in *Morphologies in Contact*, eds T. S. Martine Vanhove, A. Urdze, and H. Otsuka (Berlin: De Gruyter), 143–162.

Aglarov, M. A. (1988). *Sel"skaja obshchina v Nagornom Dagestane v XVII-nachale XIX v*. Moscow: Nauka.

Aglarov, M. A. (1994). "Andi," in *Encyclopedia of World Cultures, VI: Russia and Eurasia/China*, eds P. Friedrich and N. Diamond (Boston: G. K. Hall & Co), 23–27.

Aglarov, M. A. (2002). *Andijcy: Istoriko-ètnografičeskoe issledovanie*. Maxachkala: Jupiter.

Anthony, D. W. (2007). *The Horse, the Wheel, and Language: How Bronze Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton, NJ: Princeton University Press.

Aronson, H. I. (2008). *The Balkan Linguistic League, "Orientalism," and Linguistic Typology*. Ann Arbor, MI: Beech Stave.

Bakker, P., and Mous, M. (eds.). (1994). *Mixed Languages: 15 Case Studies in Language Intertwining*. Amsterdam: IFOTT.

Barfield, T. (1989). *The Perilous Frontier: Nomadic Empires and China, 221 BC to AD 1757*. London: Blackwell.

Bellwood, P. S. (2017). *First Islanders: Prehistory and Human Migration in Island Southeast Asia*. Hoboken: Wiley.

Blust, R. A. (2009). *The Austronesian Languages. Canberra: Research School of Pacific and Asian Studies*, Australian National University. Available online at: http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf

Chernykh, E. N. (1992). *Ancient Metallurgy in the USSR: The Early Metal Age*. New York, NY: Cambridge University Press.

Chernykh, E. N. (2009). "Formation of the Eurasian steppe belt cultures: Viewed through the lens of archaeometallurgy and radiocarbon dating," in *Social Complexity in Prehistoric Eurasia: Monuments, Metals, and Mobility*, eds B. K. Hanks and K. M. Linduff (Cambridge: Cambridge University Press), 115–145.

Creissels, D. (2014). P-lability and radical P alignment. *Linguistics* 52, 911–944. doi: 10.1515/ling-2014-0012

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: Benjamins.

Di Cosmo, N., Fank, A. J., and Golden, P. B. (eds). (2009). *The Cambridge History of Inner Asia: The Chinggisid Age*. Cambridge: Cambridge University Press.

Dobrushina, N. (2013). How to study multilingualism of the past: Investigating traditional contact situations in Daghestan. *J. Sociolinguist.* 17, 376–393. doi: 10.1111/josl.12041

Donohue, M., and Denham, T. (2012). Lapita and proto-oceanic: thinking outside the pot? *J. Pac. History* 47, 443–457. doi: 10.1080/00223344.2012.742609

Fagyal, Z., Samarth, S., Escobar, A. M., Gasser, L., and Lakkaraju, K. (2010). Centers and peripheries: Network role in language change. *Lingua* 120. 2061–2079. doi: 10.1016/j.lingua.2010.02.001

Forker, D. (2013). *A Grammar of Hinuq*. Berlin: de Gruyter Mouton.

Fortson, B. W. I. V. (2010). *Indo-European Language and Culture: An introduction*. Malden, MA; Oxford: Blackwell.

Fowler, C. S. (1972). "Some ecological clues to Proto-Numic homelands," in *Desert Research Institute Publications in the Social Sciences, 8. Great Basin Cultural Ecology: A Symposium* (Reno: Desert Research Institute), 105–121.

Frachetti, M. D. (2008). *Pastoralist Landscapes and Social Interaction in Bronze Age Eurasia*. Berkeley, CA; Los Angeles, CA: University of California Press.

Frachetti, M. D. (2012). Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* 53, 2–38. doi: 10.1086/663692

Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., Kidd, K. K., Kidd, J. R., Chambers, G. K., et al. (2008). The genetic structure of Pacific Islanders. *PLoS Genet.* 4:e19. doi: 10.1371/journal.pgen.0040019

Friedman, V. A. (2011). The Balkan languages and Balkan linguistics. *Annu. Rev. Anthropol.* 40, 275–291. doi: 10.1146/annurev-anthro-081309-145932

Golden, P. (2011). *Central Asia in World History*. Oxford: Oxford University Press.

Golla, V. (2011). *California Indian Languages*. Berkeley, CA; Los Angeles, CA: University of California Press.

Grünthal, R., and Petri, K. (2012). *eds.A Linguistic Map of Prehistoric Northern Europe*. Helsinki: Société Finno-Ougrienne.

Grünthal, R., and Nichols, J. (in press). Transitivizing/detransitivizing typology and language family history. *Lingua Posnaniensis*.

Haiman, J. (1985). *Natural syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.

Hanks, B. (2010). Archaeology of the eurasian steppes and mongolia. *Annu. Rev. Anthropol.* 39, 469–486. doi: 10.1146/annurev.anthro.012809.105110

Hiebert, F. T. (1994). *Origins of the Bronze Age Oasis Civilization in Central Asia*. Cambridge, MA: Peabody Museum of Archaeology and Ethnology, Harvard University.

Hill, J. H. (2001a). "Languages on the land," in *Language, Archaeology, and History*, ed J. Terrell (Westport, CT: Bergin & Garvey), 257–282.

Hill, J. H. (2001b). Proto-Uto-aztecan: a community of cultivators in central America? *Am. Anthropol.* 103, 913–934. doi: 10.1525/aa.2001.103.4.913

Hill, J. H. (2010). New evidence for a Mesoamerican homeland for Proto-Uto-Aztecan. *Proc. Natl. Acad. Sci. U.S.A.* 107, E33. doi: 10.1073/pnas.0914473107

Holopainen, S. (2017). *Indo-Iranian loanwords in Uralic.* Ph.D dissertation, University of Helsinki.

Janhunen, J. (1990). *Material on Manchurian Khamnigan Mongol.* Helsinki: Finno-Ugrian Society.

Janhunen, J. (1991). *Material on Manchurian Khamnigan Evenki.* Helsinki: Finno-Ugric Society.

Janhunen, J. (1996). *Manchuria: An Ethnic History.* Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, J. (2002). On the chronology of the Ainu ethnic complex. *Bull. Hokkaido Mus. Northern Peop.* 11, 1–20.

Janhunen, J. (2005). *Khamnigan Mongol.* Munich: Lincom Europa.

Janhunen, J. (2008). "Mongolic as an expansive language family," in *Past and Present Dynamics: The Great Mongolian State*, ed T. Kurebito, Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa, 127–137.

Janhunen, J. (2012). "The expansion of Tungusic as an ethnic and linguistic process," in *Recent Advances in Tungusic Linguistics*, eds A. L. Malchukov and L. J. Whaley (Harrassowitz: Wiesbaden), 5–16.

Jasanoff, J. H. (2003). *Hittite and the Indo-European Verb.* Oxford: Oxford University Press.

Joseph, B. (1983). *The Synchrony and Diachrony of the Balkan Infinitive: A Study in Areal, General, and Historical Linguistics.* Cambridge: Cambridge University Press.

Karpov, J. J., and Kapustina, E. L. (2011). *Gorcy posle gor: Migracionnye processy v Dagestane v XX-nachale XXI vv.: ix social"nye i etnokul"turnye posledstvija i perspektivy.* St. Petersburg: Rossijskaja AN, Muzej antropologii i etnografii.

Kemp, B. M., González-Oliver, A., Malhi, R. S., Monroe, C., Schroeder, K. B., McDonough, J., et al. (2010). Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6759–6764. doi: 10.1073/pnas.0905753107

Khazanov, A. M. (1994). *Nomads and the Outside World.* Madison, WI: University of Wisconsin Press.

Kirch, P. V. (2010). The peopling of the Pacific: A holistic anthropological perspective. *Annu. Rev. Anthropol.* 39, 131–148. doi: 10.1146/annurev.anthro.012809.104936

Kohl, P. L. (eds.). (2007). *The Making of Bronze Age Eurasia.* Camridge: Cambridge University Press.

Krader, L. (1963). *Social Organization of the Mongol-Turkic Pastoral Nomads.* The Hague: Mouton.

Kuzmina, E. E. (2008). *The Prehistory of the Silk Road.* Philadelphia, PA: University of Pennsylvania Press.

Lavrov, L. I. (1953). Nekotorye itogi raboty Dagestanskoj èkspedicii 1950-52 gg. *Kratkie soobščenija Instituta Ètnografii* 19, 3–7.

Lindstedt, J. (2000). "Linguistic balkanization: contact-induced change by mutual reinforcement," in *Languages in Contact*, eds D. G. Gilbers, J. Nerbonne, and J. Schaeken (Amsterdam; Atlanta: Rodopi), 231–246.

Madsen, D., and Rhode, D. (1994). *Across the West: Human Population Movement and the Expansion of the Numa.* Salt Lake City, UT: University of Utah Press.

Mallory, J. P. (1989). *In Search of the Indo-Europeans: Language, Archaeology, and Myth.* New York, NY: Thames & Hudson.

Mallory, J. P., and Adams, D. Q. (1997). *Encyclopedia of Indo-European Culture.* London: Fitzroy Dearborn.

Meakins, F. (2013). "Mixed Languages," in *Mixed Languages: A Comprehensive Guide*, eds P. Bakker and Y. Matras (Berlin: Mouton de Gruyter), 159–228.

Merrill, W. L. (2012). The historical linguistics of Uto-Aztecan agriculture. *Anthropol. Linguistics* 54, 203–260.

Miller, W. R. (1983). "Uto-Aztecan languages," in *Handbook of North American Indians,* Vol. 10, ed A. Ortiz (Washington, DC: Smithsonian Institution), 113–124.

Milroy, J., and Milroy, L. (1985). Linguistic change, social network, and speaker innovation. *J. Linguist.* 21, 339–384.

Milroy, L., and Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic mode. *Lang. Soc.* 21, 1–6.

Napol'skix, V. V. (1997). *Vvedenie v istoričeskuju uralistiku.* Izhevsk: RAN: Ural"skoe otdelenie: Udmurtskij istitut istorii, jazyka i literatury.

Nichols, J. (2005). The origin of the Chechen and Ingush: a study in alpine linguistic and ethnic geography. *Anthropol. Linguist.* 46, 129–155.

Nichols, J. (2011). "Causativization and contact in Nakh-Daghestanian," in *Proceedings of the 37th Annual Meeting: Special Session on Languages of the Caucasus*, eds C. Cathcart, S. Kang, and C. S. Sandy (Berkeley, CA: Berkeley Linguistics Society), 68–80.

Nichols, J. (2012a). "Selection for m: T pronominals in Eurasia," in *Copies vs. Cognates in Bound Morphology*, eds L. Johanson and M. Robbeets (Leiden: Brill), 47–70.

Nichols, J. (2012b). "The history of an attractor state: Adventitious m in Nakh-Daghestanian," in *Per Urales ad Orientem: Iter Polyphonicum Multilingue (Festschrift for Juha Janhunen)*, eds T. Hyytiäinen, L. Jalava, J. Saarikivi, and E. Sandman (Helsinki: SUST), 261–278.

Nichols, J. (2013). "The origin and evolution of case-suppletive pronouns: Eurasian evidence," in *Languages across Boundaries: Studies in Memory of Anna Siewierska*, eds D. Bakker and M. Haspelmath (Berlin: De Gruyter Mouton), 313–345.

Nichols, J. (2016). "Complex edges, transparent frontiers: Grammatical complexity and language spreads," in *Complexity, Isolation, and Variation*, eds R. Baechler and G. Seiler (Berlin: de Gruyter), 117–137.

Nichols, J., and Peterson, D. A. (2013a). *M-T pronouns.* eds M. Dryer and M. Haspelmath, *The World Atlas of Language Structures.* Available online at: http://wals.info/chapter/136

Nichols, J., and Peterson, D. A. (2013b). *N-M Pronouns.* eds M. Dryer and M. Haspelmath, *The World Atlas of Language Structures.* Available online at: http://wals.info/chapter/137

Nichols, J., Peterson, D. A., and Barnes, J. (2004). Transitivizing and detransitivizing languages. *Linguist. Typolo.* 8,m 2149–2211.

Nichols, J., and Rhodes, R. A. (2017). *Vectors of Language Spread at the Central Steppe Periphery: Finno-Ugric as Catalyst Language.* British Archaeological Reports (International Series), special issue edited by Rune Iverson & Guus Kroonen.

Parpola, A. (eds.). (2012). *Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology: revised and integrated "total" correlations.* Grünthal and Kallio. 119–184.

Pawley, A. K., and Ross, M. D. (1993). Austronesian historical linguistics and culture history. *Annu. Rev. Anthropol.* 22, 425–459.

Pulleyblank, E. G. (2000). "Tribal confederations of uncertain identity," in *History of the Turkic Peoples in the Pre-Islamic Period*, ed H. R. Roemer (Berlin: Klaus Schwarz Verlag), 52–101.

Rix, H. (eds.). (2001). *Lexikon der Indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen.* Wiesbaden: Reichert.

Ross, M. D. (1996). "Contact-induced change and the comparative method: cases from Papua New Guinea," in *The Comparative Method Reviewed*, eds M. Durie and M. D. Ross (New York, NY: Oxford University Press).

Ross, M. D., Pawley, A. K., and Osmond, M. (eds.). (2008). *The Lexicon of Proto Oceanic.* Canberra, ACT: Research School of Pacific and Asian Studies, Australian National University.

Schönig, C. (2003). "Turko-Mongolic Relations" in *The Mongolic Languages*, ed J. Janhunen (London: Routledge), 403–419.

Sinor, D. (ed.). (1990). *The Cambridge History of Early Inner Asia.* Cambridge: Cambridge University Press.

Sinor, D. (ed.). (1988). *The Uralic Languages: Description, History, and Foreign Influences.* Leiden: Brill.

Thomason, S. G. (2001). *Language Contact: An Introduction.* Washington, DC: Georgetown University Press.

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Structure and Complexity.* Oxford: Oxford University Press.

Volkova, N. G. (1967). Voprosy dvujazyčija na Severnom Kavkaze. *Sovetskaja etnografija* 1967, 27–40.

Vovin, A., Vajda, E., and de la Vaissière, É. (2016). Who were the *Kjet (羯) and what language did they speak? *J. Asiatiq.* 304, 125–144.

Werner, H. (2014). *Die Jenissejer unter den frühen Völkern Zentralasiens.* Munich: Lincom Europa.

Witzel, M. (2003). *Linguistic Evidence for Cultural Exchange in Prehistoric Western Central Asia. (Sino-Platonic Papers 129.)* Philadelphia, PA: Department of East Asian Language and Civilization, University of Pennsylvania.

Wixman, R. (1980). *Language Aspects of Ethnic Patterns and Processes in the North Caucasus.* Chicago, IL: University of Chicago.

Yu, W. (2011). *A study of the Mongol Khamnigan Spoken in Northeastern Mongolia.* Seoul: Seoul National University Press.

Zgusta, R. (2015). *Peoples of Northeast Asia through Time: Precolonial Ethnic and Cultural Processes along the Coast between Hokkaido and the Bering Strait.* Leiden: Brill.

Conflict of Interest Statement: 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.

Population Size and the Rate of Language Evolution: A Test Across Indo-European, Austronesian, and Bantu Languages

Simon J. Greenhill [1,2]*, Xia Hua [1,3], Caela F. Welsh [3], Hilde Schneemann [1,3] and Lindell Bromham [1,3]

[1] ARC Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT, Australia, [2] Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History (MPG), Jena, Germany, [3] Research School of Biology, Macroevolution and Macroecology, Australian National University, Canberra, ACT, Australia

What role does speaker population size play in shaping rates of language evolution? There has been little consensus on the expected relationship between rates and patterns of language change and speaker population size, with some predicting faster rates of change in smaller populations, and others expecting greater change in larger populations. The growth of comparative databases has allowed population size effects to be investigated across a wide range of language groups, with mixed results. One recent study of a group of Polynesian languages revealed greater rates of word gain in larger populations and greater rates of word loss in smaller populations. However, that test was restricted to 20 closely related languages from small Oceanic islands. Here, we test if this pattern is a general feature of language evolution across a larger and more diverse sample of languages from both continental and island populations. We analyzed comparative language data for 153 pairs of closely-related sister languages from three of the world's largest language families: Austronesian, Indo-European, and Niger-Congo. We find some evidence that rates of word loss are significantly greater in smaller languages for the Indo-European comparisons, but we find no significant patterns in the other two language families. These results suggest either that the influence of population size on rates and patterns of language evolution is not universal, or that it is sufficiently weak that it may be overwhelmed by other influences in some cases. Further investigation, for a greater number of language comparisons and a wider range of language features, may determine which of these explanations holds true.

Keywords: language evolution, language phylogenies, computational historical linguistics, demography, population size, Galton's problem, phylogenetic independence

INTRODUCTION

The role of speaker population size in shaping patterns and rates of language and cultural evolution has been much discussed, but few generalities have been agreed upon. It has been suggested that larger populations should have higher rates of language change, because populations containing more individuals provide more opportunity for innovations to arise (Richerson et al., 2009; Kline and Boyd, 2010; Baldini, 2015). Large populations might also be less prone to

random sampling effects that can cause elements of language and culture to be lost (Shennan, 2001; Henrich, 2004; Kline and Boyd, 2010; Collard et al., 2013) and they may have less stringent norm enforcement allowing them to change faster (Bowern, 2010; Trudgill, 2011). Larger populations might also have more robust transmission systems: having more people to learn from might increase fidelity of information transition (Derex et al., 2013), possibly because learners in large populations have a large set of potential models to learn from (Henrich, 2004; Kline and Boyd, 2010). Exposure to more people may make learning more robust, potentially allowing retention of a wider range of linguistic diversity (Trudgill, 2004; Hay and Bauer, 2007; Atkinson, 2011; Wichmann et al., 2011; Derex et al., 2013), although this effect is not universally supported (Caldwell and Millen, 2010; Read, 2012).

Other researchers have proposed that rates of change should be fastest in small populations due to the more rapid diffusion of new features (Nettle, 1999). Languages spoken by small speaker populations might be able to develop and retain greater linguistic complexity (Nettle, 2012). Smaller populations may have greater tolerance of diversity (Milroy and Milroy, 1985, 1992) and more malleable linguistic representations (Lev-Ari, 2017) which could speed up rates of change. Further, it has been suggested that the rate of language change may be accelerated by serial founder effects as new languages are started from relative small populations (Atkinson et al., 2008), which could increase the rate of loss of language elements from the ancestral language (Trudgill, 2004; Atkinson, 2011). Small speaker populations may also be more influenced by language contact through trade and marriage across groups, which might increase rates of language change (Bowern, 2010).

In contrast, other studies have found little or no significant effect of population size on the rate of language change or phoneme inventory size (Wichmann and Holman, 2009; Moran et al., 2012). If languages evolve in a purely stochastic manner, analogous to neutral molecular evolution, then rates of change might be independent of population size (Neiman, 1995; Shennan and Wilkinson, 2001; Bentley et al., 2004). The controversial claim that the average rate of word turnover is essentially the same in all languages, has led to much-disputed attempts to date language diversification by assuming a uniform rate of change over time (for examples of contributions to this debate see: Swadesh, 1952, 1955; Hoijer, 1956; Rea, 1958; Bergsland and Vogt, 1962; Sankoff, 1970; Blust, 2000). A similar effect has been suggested for cultural evolution because, for a variety of cultural traits from Neolithic pottery motifs to modern American pop songs, the frequency of variants matches the predictions of a purely stochastic model such that the rate of change is reasonably regular (Bentley et al., 2007).

So, despite many studies on a wide range of languages and language features, there is no consensus on whether population size has a consistent influence on patterns and rates of linguistic evolution (Bowern, 2010; Greenhill, 2014). The lack of a consistently predictable influence of population size on language change might indicate that it is not a universally important factor in rates of language change. Alternatively, the inconsistent patterns might also be due to complicated patterns

of change. For example, if rates of word gain show different relationships with population size than rates of word loss, then overall rates of change may show no consistent pattern, and the patterns uncovered in any study might depend on the mode of measuring language change (Bromham et al., 2015a). The diversity of conclusions in published studies could also arise from the diversity of languages studied, data types analyzed, or methodological approaches.

Testing these hypotheses has been challenging for several reasons. Most studies analyzing rates of language change have focused on features within one language (e.g., Johnson, 1976), or relied on simulations (e.g., Nettle, 1999), making it difficult to draw general conclusions about language change. Comparative studies of language change also need a way of overcoming the problem of statistical non-independence due to relatedness. Since languages evolve and diversify from shared ancestors, closely related languages are likely to be more similar to each other in many ways. This similarity by descent means that any association between the two traits might simply be due to the co-occurrence of the traits in a common ancestor, even if there is no functional connection between the two. Therefore, statistical tests cannot treat each language as an independent piece of evidence about the relationship between population size and the patterns of language evolution. This methodological problem, often referred to as *Galton's problem*, can confound attempts to find relationships between language and demographic factors (Moran et al., 2012; Roberts and Winters, 2013).

Our aim in this paper is to examine the influence of one aspect of demography (size of speaker population) on one aspect of language evolution (the gain and loss of words from basic vocabulary). Specifically, we wish to test whether the association between population size and rates of word gain and loss noted in a study of 10 pairs of Polynesian languages reflects a general pattern. The study of Polynesian languages compared the gain and loss of cognate terms for basic vocabulary and demonstrated greater rates of word gain in larger populations and greater rates of word loss in smaller populations (Bromham et al., 2015a). In many ways, Polynesia represents a perfect "laboratory" of language evolution, with a recent, well-characterized history of colonization of previously uninhabited islands (Goodenough, 1957). Most Polynesian languages are restricted to clearly-defined groups of islands, and the population size of speakers is closely correlated with the area inhabited (Bromham et al., 2015a). As they are the product of a recent human expansion (Spriggs, 2010), Polynesian cultures, and languages share many similarities (Pawley, 1967) and are largely found in similar environments (Kirch and Green, 1987). While these features make Polynesia an ideal case study in language evolution, it also makes it difficult to extrapolate from the patterns observed in Polynesia to general patterns of language evolution. Do languages spoken in other parts of the world by much larger groups of people with wider continental distributions show similar patterns?

To test the generality of the relationship between population size and rates of word gain and loss, we chose 153 pairs of closely related sister languages from three of the largest language families, Austronesian, Indo-European, and Niger-Congo (Bantu subfamily). The languages in our analysis are from a wide

geographic area, from the North Atlantic to the South Pacific (**Figure 1**). These language pairs span a huge range of speaker population sizes, from Perai to Aputai spoken on the island of Wetar in the Maluku province of Indonesia (spoken by 280 and 150 people, respectively), to Sambaa and Bondei spoken in the mountain regions of Northern Tanzania (664,000 and 50,000 people), to German and Luxembourgish in continental Europe (spoken by 69,800,000[1] and 266,000 people respectively). For each of these families, we used published linguistic databases of basic vocabulary to evaluate relative rates of word gain and loss, using a technique that explicitly accounts for non-independence due to the relatedness of the languages.

MATERIALS AND METHODS

Language Families

We analyzed data from three of the largest language families, Austronesian, Indo-European, and Niger-Congo (Bantu subfamily). These language groups span a large range of population sizes, a wide geographic area and varied cultures and histories, which allows us to test the generality of the influence of population size on rates of language change (**Figure 1**).

The Austronesian language family is the world's second largest, containing 1,274 languages spoken across a wide range of islands as well as on continental landmasses, from Madagascar to Southeast Asia and the Pacific (Hammarström et al., 2016). There are 10 major Austronesian sub-groups, nine of which contain only 20 languages in total, and are spoken by indigenous Formosan people in Taiwan (Blust, 2013). The other languages form the Malayo-Polynesian group, which began diversifying around 4,000 to 4,500 years ago in a series of expansions across the Pacific Ocean (Gray et al., 2009; Hung et al., 2011; Spriggs, 2011; Amano et al., 2013; Ko et al., 2014; Blust, 2015). Austronesian societies include hunter-gatherer groups (e.g., the Mikea in Madagascar), agriculturalists (e.g., the Saisiyat in Taiwan), and complex socially-stratified societies such as in Java or Bali (Geertz, 1959; Jay, 1969). Austronesian languages vary greatly in their range and degree of isolation (Gavin and Sibanda, 2012), from remote Pacific islands containing a single indigenous language, to the diverse larger islands and landmasses of Southeast Asia and Near Oceania where many different languages may come into contact.

The Indo-European language family contains 581 languages in 8–10 sub-families, including many of the languages of Europe (e.g., English, Spanish, Portuguese, Russian), as well as many spoken in the Middle East and India (e.g., Bengali, Farsi, Hindi, Punjabi). The origin of the family is debated: while some place the origin in the Russian Steppes 5,000 years ago (Anthony and Ringe, 2015; Chang et al., 2015; Haak et al., 2015), others date it to Anatolia 8,000 years ago (Renfrew, 1987; Gray and Atkinson, 2003; Gray et al., 2011; Bouckaert et al., 2012). However, the uncertainty concerning the origin of the family does not affect our analysis of closely related sister pairs.

---

[1]The current population of Germany is ∼82 million speakers, but Lewis et al. (2015) cites a 2012 European Commission report for Standard German which indicates 69.8 million native speakers.

The Niger-Congo languages comprise the world's largest language family with 1,430 languages spoken across sub-Saharan Africa (Hammarström et al., 2016). The Bantu languages (550 languages), one of the major subgroups of Niger-Congo, are thought to have originated between 4,000 and 5,000 years ago in west central Africa, perhaps near the Nigerian-Cameroon border, and expanded south through the rainforest (Berniell-Lee et al., 2009; Montano et al., 2011; Pakendorf et al., 2011; de Filippo et al., 2012; Currie et al., 2013; Li et al., 2014; Grollemund et al., 2015).

Language Data

There are many different ways of investigating language change, for example considering changes to lexicon, morphology, phonology, or syntax (Bowern and Evans, 2014). Here we consider one particular form of language evolution, the gain, and loss of word variants from basic vocabulary, as it allows us to make comparable measures of rate of language change across different languages (Bromham et al., 2015a). Basic vocabulary consists of a common set of concepts found in all languages, such as "hand," "mother," or "water," for which the common word forms have been recorded in different languages—sometimes referred to as a Swadesh list (Swadesh, 1955).

We used published databases of the different words (lexemes) used for a defined set of basic concepts (semantic categories). Using curated databases ensures that word forms are recorded in a comparable format for the different languages within a family. Each of the databases identifies cognate sets: forms which exhibit some systematic degree of similarity and are identified as derived from a common ancestor (Durie and Ross, 1996; Bowern and Evans, 2014). For example, the semantic category "tree" is represented by different words in different Indo-European languages. In some languages, the words for "tree or wood" reflect the same homologous cognate class derived from the common proto-Indo-European *deru-o-* (Derksen, 2008), including ʾδҽṇdrǫ (Greek), dⁱɛrⁱɛvɔ (Russian), and English *tree* (via Old English, *trēow*). In contrast, the Italic languages have adopted a new lexeme reflected in forms like Latin *arbor*, French *arbre*, Italian *albero* and Spanish *árbol*. Homologous forms are not just look-alikes but are identified using the linguistic comparative method to determine systematic sound correspondences and phonological innovations (Paul, 1880; Bloomfield, 1933; Durie and Ross, 1996; Bowern and Evans, 2014). We can use these patterns of homology to identify the presence of words shared by descent, the loss of shared cognates from related languages, and also to identify cases of gain of new words that have not been inherited from a common ancestor.

For the Austronesian languages we used the Austronesian Basic Vocabulary Database (ABVD, Greenhill et al., 2008) which contains wordlists for 210 semantic categories from 1,278 languages. For the Indo-European languages, we used the Indo-European Lexical Cognacy Database (IELex, Bouckaert et al., 2012), which contains wordlists for 225 semantic categories from 163 languages. Basic vocabulary for 100 words from 409 Bantu languages were provided by Grollemund et al. (2015) in a phylogenetic dataset that records a single variant per semantic category for each language. The wordlists in these three databases are not identical as they have been modified to contain region
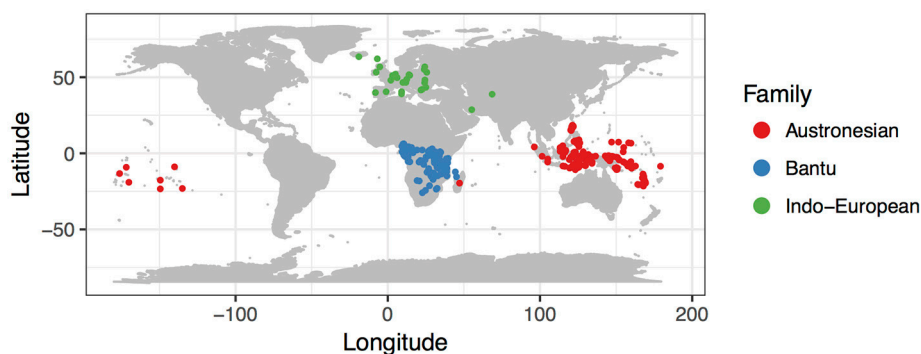
FIGURE 1 | Map of languages included in this study. Each point represents the mid-point of the area occupied by one of the languages included in our study (see **Tables 1–3**).

specific words, but the lists do overlap substantially as they are based on standard Swadesh lists (Swadesh, 1952).

Language Pairs

To control for relatedness between languages and avoid Galton's problem, we use a simple and robust method of selecting phylogenetically independent sister pairs. Sister pairs are each other's closest relatives on a phylogeny that form a pair of tips connected by their most recent common ancestor. This means that any difference between the two sister languages has arisen since that last common ancestor, and changes in one language are independent of changes in its sister language. Therefore we can ask questions such as: when two languages evolve from a common stock, does the language with the smaller population acquire new words at a greater or lesser rate than the larger language? If we select sister pairs that are each other's closest relatives, such that they share a more recent common ancestor with each other than either shares with any other language in the analysis, then the pairs are said to be phylogenetically independent (Felsenstein, 1985; Harvey and Pagel, 1991), because any differences between the pair has evolved since their common ancestor, and is not a result of their shared inheritance. Selecting phylogenetically independent sister pairs is like running an experiment over and over again, taking one language, splitting it in two, and seeing which one evolves faster (Bromham, 2016). Given sufficient independent comparisons we can use statistical analysis to look for consistent patterns between the features of languages and their rate of change, by comparing them to their sister languages.

The sister pairs approach has advantages over whole tree phylogenetic methods that use every branch in a phylogeny as a datapoint in an analysis. Using only the tips of the phylogeny avoids the need to infer ancestral states at increasing depths down the phylogeny in order to correlate past states with rates of change inferred from the internal branches of the tree. Using only tip branches also avoids the problem of non-independence between ancestor and descendant lineages within the phylogeny, as each branch is likely to be more similar in many traits to its immediate neighbors than it is to other more distantly related branches.

Phylogenetically independent pairs of languages were chosen from published phylogenies and checked for consistency with language taxonomy based on linguistic comparative data. We did not include creoles as they are hybrid languages with a high degree of borrowing and may have different patterns of change to other related languages (Thomason and Kaufman, 1988; Blasi et al., 2017). We did not include extinct or ancient languages, as their lexical documentation may not be as complete as for extant languages, and their speaker population sizes may also be less well established. We included only well-attested sister pairs in our analysis. We began by selecting sister pairs from the published phylogenies (Gray et al., 2009; Bouckaert et al., 2012; Grollemund et al., 2015; Hammarström et al., 2016), then checked the relationship between pairs in the Ethnologue (Lewis et al., 2015). We discarded any pairs where the classification in the Ethnologue was at odds with pairs identified from the phylogeny. We also used phylogenetic support measures from published phylogenies as a guide to selecting well-attested sister pairs, rejecting any pairs with less than 80% posterior probability in the published phylogeny.

Contemporary speaker population size was obtained from the Ethnologue (Lewis et al., 2015) using the *in area* speaker population where given, rather than the total global number of speakers. Languages with insufficient linguistic, temporal or population data were excluded. Thus, this is not an exhaustive list of all sister languages for these language families, but a conservative selection which fits all relevant criteria for this study. This selection process resulted in 81 pairs of Austronesian languages (**Table 1**), 14 pairs of Indo-European languages (**Table 2**), and 58 pairs of Bantu languages (**Table 3**).

Language pairs that have a shorter period of divergence will have larger uncertainty in the estimates of their rates of language change (Welch and Waxman, 2008; Hua et al., 2015), so we use estimated branch lengths between sister languages to correct for this effect. We extracted branch lengths from the published language phylogenies (Gray et al., 2009; Bouckaert et al., 2012; Grollemund et al., 2015) which are estimated using phylogenetic dating methods from their total datasets combined with historical

TABLE 1 | Sister pairs of languages from the Austronesian language family, showing the taxon label, the ISO-639-3 language identification code, the number of gains, losses, and total changes, population size, and branch-length.

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Branch length
1	Agta	agt	50	32	82	780	138.91
	Gaddang	gad	54	34	88	30,000	
2	AmbaiYapen	amk	112	36	148	10,100	777.07
	WindesiWandamen	wad	117	12	129	5,000	
3	AmbrymSouthEast	tvk	74	45	119	3,700	0.07
	PaameseSouth	pma	51	31	82	6,000	
4	Anakalang	akg	12	23	35	16,000	828.02
	Wanukaka	wnk	23	36	59	10,000	
5	Aputai	apx	14	16	30	150	111.68
	Perai	wet	12	14	26	280	
6	As	asz	86	26	112	230	1905.88
	BigaMisool	xmt	85	25	110	1,250	
7	Atoni	aoz	124	46	170	700,000	1224.40
	RotiTermanu_D	twu	97	18	115	30,000	
8	AttaPamplona	att	26	18	44	1,000	0.00
	Ibanag	ibg	34	28	62	500,000	
9	Avava	tmb	57	67	124	700	552.95
	Neveei	vnm	44	42	86	500	
10	Bali	ban	106	58	164	3,330,000	1897.90
	Sasak	sas	73	59	132	2,100,000	
11	Baree	pmf	80	36	116	137,000	9.22
	Mori	xmz	104	48	152	14,000	
12	Belait	beg	72	34	106	1,000	1107.19
	BerawanLongTerawan	zbw	85	43	128	1,000	
13	Bintulu	bny	70	38	108	4,200	2335.48
	MelanauMukah	mel	68	40	108	113,000	
14	Bobot	bty	47	21	68	4,500	971.12
	Bonfia	bnf	50	17	67	1,000	
15	Bonerate	bna	27	13	40	9,500	0.00
	Popalia	bhq	27	12	39	130,000	
16	BontokGuinaang	bnc	56	34	90	40,700	0.00
	KankanayNorthern	xnn	37	33	70	70,000	
17	BugineseSoppeng_D	bug	80	50	130	5,000,000	2102.00
	TaeSToraja	rob	58	41	99	340,000	
18	Bugotu	bgt	107	51	158	4,050	0.20
	Nggela	nlg	72	30	102	11,900	
19	Bukat	bvk	100	47	147	400	1102.31
	Lahanan	lhn	71	25	96	350	
20	Buli	bzq	114	19	133	2,520	1578.20
	Giman	gzn	152	41	193	2,900	
21	BuruNamroleBay	mhs	110	38	148	33,000	2158.07
	Soboyo	tlv	121	53	174	4,520	
22	Bwaidoga	bwd	52	14	66	6,500	4.10
	Diodio	ddi	57	29	86	2,180	
23	Cebuano	ceb	31	44	75	15,800,000	553.03
	Surigaonon	sgd	70	42	112	400,000	
24	ChekeHolo	mrn	94	61	155	10,800	313.81
	KilokakaYsabel	jaj	34	23	57	10	
25	Dai	dij	52	18	70	820	0.01

(Continued)

TABLE 1 | Continued

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Branch length
	NorthBabar	bcd	50	17	67	1,000	
26	Dehu	dhv	190	22	212	13,000	1722.11
	Nengone	nen	185	30	215	8,720	
27	Dobuan	dob	73	38	111	10,000	667.39
	Molima	mox	81	35	116	4,010	
28	Emae	mmw	4	23	27	400	0.00
	UveaWest	uve	2	24	26	2,200	
29	Gapapaiwa	pwg	76	18	94	3,000	756.15
	Ubir	ubr	101	44	145	2,560	
30	Geser	ges	63	23	86	36,500	476.20
	Watubela	wah	71	36	107	4,000	
31	GhariGuadalcanal	gri	39	31	70	12,100	0.01
	Tolo	tlr	33	33	66	12,500	
32	GorontaloHulondalo	gor	96	50	146	1,000,000	0.12
	Kaidipang	kzp	71	22	93	26,600	
33	HituAmbon	htu	64	27	91	16,000	531.14
	Paulohi	plh	73	33	106	50	
34	HoavaNewGeorgia	hoa	61	41	102	460	400.12
	MarovoNewGeorgia	mvo	67	54	121	8,090	
35	Imroing	imr	31	24	55	560	327.51
	TelaMasbuar	tvm	25	16	41	1,050	
36	Inibaloi	ibl	35	33	68	111,000	117.06
	KallahanKayapaProper	kak	22	20	42	15,000	
37	ItnegBinongan	itb	34	40	74	7,500	0.01
	KalingaGuinaangLubuagan_D	knb	29	36	65	30,000	
38	Jawe	jaz	109	24	133	990	0.00
	Nelemwa	nee	118	26	144	1,090	
39	Kalagan	kqe	33	38	71	70,000	0.00
	Mansaka	msk	25	31	56	57,800	
40	Kapampangan	pam	74	41	115	1,900,000	1165.44
	SambalBotolan	sbl	108	56	164	32,900	
41	Kapingamarangi	kpg	4	18	22	3,000	226.89
	Nukuoro	nkr	3	16	19	860	
42	Kedang	ksx	106	37	143	30,000	1219.42
	Lamaholot	slp	93	33	126	180,000	
43	Kemak	kem	65	16	81	72,000	866.01
	Mambai	mgm	80	27	107	131,000	
44	Kerinci	kvr	56	33	89	260,000	188.09
	Minangkabau	min	29	37	66	5,530,000	
45	Komering	kge	74	37	111	470,000	1899.99
	Lampung	ljp	45	29	74	827,000	
46	KoronadalBlaan	bpr	10	11	21	150,000	415.53
	SaranganiBlaan	bps	4	5	9	90,800	
47	Kuanua	ksd	111	31	142	61,000	652.24
	LungaLungaMinigir	vmg	83	21	104	600	
48	KwaraaeSolomonIslands	kwf	43	33	76	32,400	197.90
	Toambaita	mlu	47	49	96	12,600	
49	Leipon	lek	42	22	64	650	840.70
	Loniu	los	43	20	63	460	
50	Lenakel	tnl	34	25	59	11,500	0.00
	TannaSouthwest	nwi	26	13	39	4,500	

(Continued)

TABLE 1 | Continued

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Branch length
51	Levei	tlx	62	16	78	1,600	1480.51
	Likum	lib	56	13	69	80	
52	Lou	loj	74	36	110	1,000	2.12
	Nauna	ncn	64	25	89	100	
53	Luangiua	ojv	6	14	20	2,370	189.54
	Sikaiana	sky	3	17	20	730	
54	Maanyan	mhy	74	22	96	150,000	1100.00
	MerinaMalagasy	plt	119	54	173	7,520,000	
55	Manam	mva	94	36	130	7,950	171.10
	Wogeo	woc	87	34	121	1,620	
56	Mangareva	mrv	1	28	29	600	670.85
	Marquesan	mrq	23	33	56	5,400	
57	ManobollianenKibudtungan_D	mbi	22	34	56	14,600	125.89
	WBukidnonManobo	mbb	23	31	54	15,000	
58	ManoboKalamansigCotabatoParil_D	mta	47	48	95	30,000	306.23
	ManoboSaranganiKayaponga_D	mbs	33	34	67	58,000	
59	Masiwang	bnf	17	4	21	1,000	0.00
	Werinama	bty	19	7	26	4,500	
60	Matukar	mjk	51	16	67	430	556.52
	Megiar	tbc	49	18	67	40,000	
61	Modang	mxd	90	24	114	15,300	339.52
	PunanKelai	sge	83	21	104	2,000	
62	Mokilese	mkj	15	9	24	1,500	1232.98
	Ponapean	pon	34	29	63	31,350	
63	Mortlockese	mrl	2	6	8	5,900	156.44
	Satawalese	stw	1	7	8	460	
64	Mota	mtt	87	33	120	900	933.62
	Mwotlap	mlv	68	42	110	1,800	
65	Naman	lzl	52	42	94	15	415.28
	Tape	mrs	70	74	144	15	
66	Ngadha	nxg	77	26	103	60,000	162.76
	Soa	ssq	73	36	109	10,000	
67	NgaiborSAru	txn	100	19	119	7,910	1319.07
	UjirNAru	udj	89	8	97	1,030	
68	Nguna	llp	57	24	81	9,500	2179.01
	SouthEfate	erk	62	39	101	6,000	
69	Niue	niu	12	52	64	2,030	0.00
	UveaEast	wls	7	25	32	9,620	
70	PeteraraMaewo	mwo	45	44	89	1,400	1667.29
	Raga	lml	47	38	85	6,500	
71	Rurutuan	aut	38	19	57	3,000	31.67
	TahitianModern	tah	27	33	60	68,260	
72	Saliba	sbe	82	29	111	2,500	0.00
	Suau	swp	48	24	72	6,800	
73	SangilSaraganiIslands	snl	42	41	83	15,000	497.32
	SangirTabukang_D	sxn	20	19	39	255,000	
74	Seimat	ssg	98	39	137	1,000	2128.97
	Wuvulu	wuv	100	35	135	1,000	
75	Serili	sve	27	14	41	330	480.25
	SouthEastBabar	vbb	21	10	31	4,460	

(Continued)

TABLE 1 | Continued

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Branch length
76	SubanonSiocon	suc	47	17	64	125,000	415.21
	SubanunSindangan	syb	50	23	73	140,000	
77	SyeErromangan	erg	53	15	68	1,900	1828.80
	Ura	uur	61	28	89	6	
78	Taiof	sps	88	39	127	1,400	26.49
	Teop	tio	129	40	169	5,000	
79	Tigak	tgc	63	39	102	6,000	558.85
	TungagTungakLavongai	lcm	123	22	145	12,000	
80	Tokelau	tkl	12	45	57	1,410	1428.51
	Tuvalu	tvl	4	21	25	10,700	
81	VaghuaChoiseul	tva	63	35	98	1,960	0.01
	Varisi	vrs	40	20	60	5,160	

and archeological information (**Tables 1**, **3**). Because the relative height of the ancestral node of any given pair will be determined not only by the differences between the pair but also by rates of change estimated on the rest of the phylogeny, it should be at least partially independent of the number of gains and losses between members of any given pair. Branch lengths were only used for the Welch & Waxman analysis (see below).

Comparing Rates of Language Change

We use comparisons of words from basic vocabulary between pairs of closely-related languages to identify instances of gain and loss of words. We identified patterns of word gain and loss by recording instances where a cognate form within a given semantic category was present in one language in a sister pair but not found in its sister language (Bromham et al., 2015a). A cognate class is a set of words identified as derived from a common ancestor, and therefore the presence of a cognate class in one language of a pair, and in other languages within the family, implies the presence of that cognate class in the common ancestral language of the pair. This method differs from approaches where the net dissimilarity between lists of terms is compared (Wichmann and Holman, 2009). Instead we use only those words that show a pattern of occurrence that is informative for determining differences in rates of gain and loss of words (Bromham et al., 2015a).

If a word form found in one sister language has a cognate in other languages in the language family, then it is likely to have been inherited from the common ancestor. This implies that the absence of that cognate form in the other sister language must be due to its loss after divergence from the common ancestor of the pair (**Figure 2**). If one of the sister languages has a unique word form that has no recognized cognates in any other language in the family, then it presumably represents a gain of a new word since it split from its sister language. Therefore we can identify instances of word gain and loss in both members of a related pair of languages. Any such changes that have occurred in one sister pair of languages can be considered to have happened independently from changes in other sister pair of languages, so

these comparisons can be treated as statistically independent data points (Bromham et al., 2015a).

Our analysis only includes cognate classes showing rates-informative patterns that allow us to localize a word gain or loss to only one member of a sister pair (**Figure 2**). There are two rates-informative patterns. Presence of a cognate class in one member of the pair but not the other indicates a loss of the shared ancestral cognate form from one sister language after divergence from the common ancestor. Presence of a novel form in one member of the pair that has no known cognates in any other member of the language family indicates the gain of a new word in one sister language after divergence from the common ancestor. We did not consider cognate forms that are present in both members of a sister pair because they have both inherited those forms from their common ancestor, and neither has lost that cognate, so those cognates are non-informative for rates of gain and loss. Similarly, we did not count any cognate class that is absent from both members of a sister pair, on the assumption that it was not present in their common ancestor.

We do not include any identified loan words in the analysis, so any cognate terms shared by two languages should be present in the language due to inheritance from a common ancestor, rather than borrowing (horizontal transfer) from another language. The addition of a new word does not necessarily involve the loss of an existing word as languages can have multiple lexemes for one category, therefore each recorded gain, or loss of a lexeme was counted as a separate event, regardless of semantic category. Any lexemes that were recorded as "doubtful" or "exclude" in the databases were excluded from our analysis. Any semantic categories that did not contain entries for both languages in the pair were also excluded as we are unable to ascertain if this absence is a true absence or simply missing data.

This counting procedure will in some cases count semantic shifts as a change (e.g., Danish *træ* "tree" is cognate with proto-Indo-European *dóru* but has shifted to also mean "wood"). Due to the nature of these datasets (cognate classes coded *within* a limited number of semantic categories), we cannot quantify semantic shift, which may include gain, or loss of meaning from unrecorded semantic categories. Cognates that change

TABLE 2 | Sister pairs of languages from the Indo-European language family, showing the taxon label, the ISO-639-3 language identification code, the number of gains, losses, and total changes, population size, and branch-length.

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Branch length
1	Persian_List	pes	16	36	52	45,000,000	788.52
	Tadzik	tgk	39	26	65	6,380,000	
2	Romanian_List	ron	41	19	60	19,900,000	727.95
	Vlach	rup	31	44	75	50,000	
3	Sardinian_C	sro	12	22	34	500,000	615.19
	Sardinian_N	src	20	29	49	500,000	
4	Ladin	lld	13	18	31	31,000	649.30
	Romansh	roh	20	33	53	40,000	
5	French	fra	2	11	13	60,000,000	522.68
	Walloon	wln	20	26	46	600,000	
6	Portuguese_ST	por	36	24	60	10,000,000	337.65
	Spanish	spa	19	36	55	38,400,000	
7	Irish_A	gle	40	25	65	138,000	563.10
	Scots_Gaelic	gla	47	25	72	58,700	
8	Dutch_List	nld	7	17	24	15,700,000	208.55
	Flemish	vls	5	22	27	1,070,000	
9	German_ST	deu	7	14	21	69,800,000	641.05
	Luxembourgish	ltz	17	30	47	266,000	
10	Faroese	fao	9	14	23	66,000	777.56
	Icelandic_ST	isl	7	27	34	230,000	
11	Bulgarian	bul	19	44	63	7,020,000	712.58
	Macedonian	mkd	32	14	46	1,340,000	
12	Lusatian_L	dsb	4	8	12	6,670	54.80
	Lusatian_U	hsb	1	5	6	13,300	
13	Byelorussian	bel	15	45	60	2,220,000	535.34
	Ukrainian	ukr	42	26	68	32,000,000	
14	Latvian	lav	68	46	114	1,470,000	1359.36
	Lithuanian_ST	lit	61	40	101	2,800,000	

meaning and undergo semantic shifts into a new category in the word list might appear as the gain of a new cognate into the recipient semantic category. If there is a subsequent change of meaning away from the original semantic category, then we would count this as loss of a cognate from the original semantic category. While this represents a somewhat different kind of change from the origin, replacement and loss of lexical items, it is still indicative of language change. In this way, we may include changes in both form and meaning. One of the ways that the population size hypothesis might affect language change is through altering semantics.

The total number of gains, losses, and non-informative results were counted for all available semantic categories for each pair of languages. The raw counts were standardized by the total number of comparisons made between the pairs (gains + losses + non informative + excluded) to allow for comparisons to be made between languages. We have developed a Python package, *RateCounter* (https://github.com/SimonGreenhill/RateCounter), to extract this rate information from common phylogenetic file formats.

Statistical Analysis

We applied two statistical analyses to test for any consistent relationship between population size and rates of word gain and loss. One analysis is Poisson regression (Bromham et al., 2015b;

Hua et al., 2015), which assumes that gain and loss counts follow a Poisson process, and rates of word gain and loss are linear functions of population size on a log-log scale (which confines rates to positive values). The regression coefficient between population size and rate of word gain and loss was estimated by accounting for the phylogenetic structure of the data and using a model with stable population size, origination of new language by fission, and negligible founder effect—the simplest population model tests from a previous study (Bromham et al., 2015a). We also tested an alternative model that incorporates population growth, to reflect recent population expansion, however this model provided a poor fit to the data and would not converge for most datasets. Therefore we applied the simplest model because it has the least number of parameters and assumptions and does not require divergence dates. To assess the model fit, we used likelihood ratio tests to compare each model to null models which assume no effect of population size on rates of language evolution. The effect size was calculated as the pseudo R^2 measures for the Poisson regression (**Table 1**).

In addition, we performed an analysis that first uses the Welch & Waxman test to remove pairs where the divergence between the sister languages is too recent to obtain reliable measures of rates of word gain and loss (Welch and Waxman, 2008). This is done by progressively removing pairs until there is no negative relationship between the absolute value of the standardized difference in the counts of gains and losses between sister languages and the square root of divergence time (Welch and Waxman, 2008), here represented by branch length from the published phylogeny (**Tables 1–3**). This analysis asks whether the difference in population size between each pair predicts the difference in the gain and loss rate, while accounting for the differences in divergence times between the pairs. So the difference in the gain and loss rate needs to be standardized by divergence times. Since the quantity of data for each language pair may vary, we also need to standardize the differences in the gain and loss rate by the amount of available data. We calculate the standardized difference as the difference in the counts of gains and losses between sister languages divided by their average counts of gains and losses and by the square root of branch length (following Bromham et al., 2015a). We removed any pairs for which the standardized difference was not a reliable estimate of difference in gains or losses rate, for example due to too recent a divergence or insufficient differences between the languages. following the procedure of Welch and Waxman (2008). After removing pairs with unreliable estimates, the analysis then applies least squares regression of the standardized differences between the remaining sister language pairs against their differences in log-transformed population sizes divided by the square root of branch length (Bromham et al., 2015a).

RESULTS

The Poisson regression of population size and rates of change in the Indo-European language family (14 pairs) suggests that languages with smaller speaker population sizes had significantly

TABLE 3 | Sister pairs of languages from the Bantu language sub-family, showing the taxon label, the ISO-639-3 language identification code, the number of gains, losses, and total changes, population size, and branch-length.

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Time
1	A15C_Akossi	bss	1	7	8	100,000	479.35
	A15C_Mkaa	bqz	3	9	12	30,000	
2	A24_Duala	dua	0	11	11	87,700	684.16
	A27_Malimba	mzd	5	16	21	2,230	
3	A32C_Batanga	bnm	0	9	9	9,000	572.43
	A34_Benga	bng	2	11	13	3,900	
4	A41_Barombi-Kang	bbi	3	11	14	3,000	526.00
	A42_Abo	abb	0	8	8	12,000	
5	A44_Tunen	tvu	8	23	31	35,300	1226.99
	A46_Nomaande	lem	12	27	39	6,000	
6	A62B_Mmala	mmu	0	1	1	8,000	317.48
	A62C_Libie	ekm	4	5	9	6,400	
7	A841_Badwe	ozm	1	3	4	40,000	149.27
	A84_Njem	njy	0	2	2	4,400	
8	A91_Kwakum	kwu	12	25	37	10,000	1193.38
	A93_Kako	kkj	8	21	29	100,000	
9	B201_Ndasa	nda	0	2	2	4,530	182.77
	B24_Wumbvu	wum	2	4	6	18,300	
10	B252_Mahongwe	mhb	1	10	11	8,000	433.10
	B25_Kota	koq	2	11	13	25,000	
11	B301_Viya	gev	4	23	27	50	1263.89
	B305_Vove	buw	1	20	21	4,000	
12	B304_Pinzi	pic	1	7	8	1,000	251.89
	B32_Kande	kbs	2	8	10	500	
13	B52_Nzebi	nzb	1	7	8	120,000	350.62
	B53_Tsaangi_Poungi	tsa	2	8	10	13,600	
14	Bamun_Grassfields	bax	7	7	14	420,000	536.22
	Mungaka_Grassfields	mhk	6	6	12	50,100	
15	C142_Mondongo	bui	1	8	9	4,000	313.36
	C412_Libobi	bmg	2	9	11	20,000	
16	C37_Ebudza	bja	8	16	24	226,000	1116.06
	C42_Ebwela	bwl	12	20	32	8,400	
17	C71_Tetela	tll	6	19	25	750,000	930.34
	C76_Ombo	oml	4	17	21	8,400	
18	C83_Bushong	buf	0	10	10	155,000	751.44
	C85_Wongo	won	2	12	14	12,700	
19	D201_Liko	lik	10	31	41	60,000	1176.87
	D21_Baali	bcp	11	32	43	42,000	
20	D305_Nyanga-li	nyc	4	4	8	48,000	583.87
	D43_Nyanga	nyj	4	4	8	150,000	
21	D333_Ndaaka	ndk	3	8	11	25,000	467.68
	D334_Mbo	zmw	5	10	15	11,000	
22	E72a_Giryama	nyf	2	14	16	944,000	600.19
	E73_Digo	dig	5	17	22	313,000	
23	E74a_Dawida	dav	8	23	31	274,000	1081.89
	G39_Saghala	tga	10	25	35	79,000	
24	F12_Bende	bdp	10	29	39	27,000	1126.01
	F23_Sumbwa	suw	1	20	21	191,000	
25	F24_Kimbu	kiv	4	18	22	78,000	762.10
	F31_Nyiramba	nim	7	21	28	455,000	

(Continued)

TABLE 3 | Continued

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Time
26	G11_Gogo	gog	2	26	28	1,440,000	813.11
	G12_Kagulu	kki	1	25	26	241,000	
27	G23_Sambaa	ksb	3	14	17	664,000	363.63
	G24_Bondei	bou	2	13	15	50,000	
28	G35_Luguru	ruf	6	21	27	692,000	469.00
	G36_Kami	kcu	1	16	17	16,400	
29	G44D_Maore	swb	4	6	10	92,800	262.27
	G44b_Ndzwani	wni	1	3	4	264,000	
30	G61_Sangu	sbp	1	20	21	75,000	611.71
	G66_Wanji	wbi	6	25	31	28,000	
31	G62_Hehe	heh	3	13	16	805,000	491.97
	G63_Bena	bez	6	16	22	670,000	
32	H16a_Kisikongo_2013	kwy	1	12	13	537,000	695.08
	H16a_Kisolongo_DRC_2012	kng	2	13	15	3,000,000	
33	JD64_Shubi	suj	0	5	5	153,000	288.52
	JD65_Hangaza	han	2	7	9	150,000	
34	JD66_Kiha	haq	3	11	14	990,000	483.82
	JD67_Kivinza	vin	2	10	12	10,000	
35	JE11_Runyoro	nyo	3	10	13	667,000	358.91
	JE12_Rutooro	ttj	5	12	17	488,000	
36	JE13_Runyankore	nyn	0	6	6	2,330,000	342.41
	JE14_Rukiga	cgg	3	9	12	1,580,000	
37	JE21_Runyambo	now	1	9	10	400,000	404.52
	JE22_Haya	hay	1	9	10	1,300,000	
38	JE25_Jita	jit	5	10	15	205,000	494.21
	JE25_Kilegi	reg	3	8	11	86,000	
39	JE31_Lumasaaba	myx	6	9	15	1,120,000	544.30
	JE31c_Bukusu	bxk	9	12	21	1,433,000	
40	K332_Rumanyo	diu	2	11	13	10,200	783.09
	K33_Kwangali	kwn	3	12	15	73,100	
41	Kom_Grassfields	bkm	1	2	3	233,000	407.67
	Oku_Grassfields	oku	5	6	11	87,000	
42	L31a_Luba-Kasai	lua	2	12	14	6,300,000	1144.85
	L32_Kanyok	kny	6	16	22	200,000	
43	L35_Sanga	sng	1	8	9	431,000	570.02
	L41_Kaonde	kqn	0	7	7	206,000	
44	M11_Pimbwe	piw	1	7	8	29,000	429.35
	M12_Lungwa	rnw	1	7	8	18,000	
45	M21_Ndali	ndh	7	23	30	150,000	734.25
	M31_Nyakyusa	nyy	7	23	30	805,000	
46	M21_Wanda	wbh	1	4	5	24,000	203.84
	M22_Namwanga	mwn	0	3	3	140,000	
47	M24_Malila	mgq	2	19	21	65,000	414.18
	M25_Safwa	sbk	4	21	25	158,000	
48	M52_Lala	leb	1	4	5	353,000	293.42
	M54_Lamba	lam	1	4	5	201,000	
49	M61_Lenje	leh	2	7	9	128,000	643.12
	M62_Soli	sby	9	14	23	34,100	
50	Moghamo_Grassfields	mgo	9	9	18	183,000	715.68
	Njen_Grassfields	njj	6	6	12	1,800	
51	N11_Manda	mgs	1	18	19	22,000	671.25

(Continued)

TABLE 3 | Continued

Pair	Taxon	ISO-639-3	Gain	Loss	Total	Population	Time
	N12_Ngoni	ngo	3	20	23	170,000	
52	N13_Matengo	mgv	5	16	21	150,000	545.66
	N14_Mpoto	mpa	0	11	11	80,000	
53	N31_Chewa	nya	6	18	24	7,000,000	755.02
	N42_Kunda	kdn	1	13	14	145,000	
54	P21_Yao	yao	10	16	26	2,200,000	598.01
	P22_Mwera	mwe	5	11	16	469,000	
55	P31G_Ikorovere	mgh	6	6	12	963,000	390.78
	P31_Emakhua	vmw	2	2	4	3,090,000	
56	S11_Shona	sna	4	13	17	10,700,000	858.44
	S16_Kalanga	kck	6	15	21	700,000	
57	S311_Shekgalagari	xkv	8	14	22	40,000	557.18
	S31_Tswana	tsn	5	11	16	1,070,000	
58	S51_Tshwa	tsc	2	6	8	1,160,000	276.76
	S53_Tsonga	tso	1	5	6	2,280,000	

Language identification codes following Guthrie's scheme are prepended to the taxon label.

FIGURE 2 | Method for determining word gains and losses. If a cognate form is found in one member of a sister pair and in another language in the family, it must have been lost from the other sister language. A lexeme that has no cognates in any other language in the family, including its sister language, is considered to have been gained since they split from their shared common ancestor.

TABLE 4 | Results of Poisson regression on Population size and rate of language change in pairs of Austronesian, Indo-European languages, and Bantu languages.

	N	Mean	SE	Statistic	*P*-value	R^2
AUSTRONESIAN						
Gain	81	0.000	0.017	0.07	0.791	0.000
Loss	81	0.001	0.024	0.13	0.718	0.001
INDO-EUROPEAN						
Gain	14	−0.042	0.062	2.18	0.140	0.035
Loss	14	−0.095	0.058	**12.82**	**0.000**	0.216
BANTU						
Gain	58	−0.000	0.086	0.01	0.911	0.000
Loss	58	−0.000	0.047	0.00	0.951	0.000

N: number of language pairs; Mean, estimated regression coefficient for the relationship between population size and rates of language change; SE, standard error for the regression coefficient; Statistic, likelihood ratio; P-value, results significant at 0.05 shown in bold; R^2, pseudo R^2 for Poisson regression.

higher rates of word loss (**Table 4**, **Figure 3**). Least squares regression also suggests a significant negative relationship between contrasts in population size and contrasts in the rate of word loss (coefficient = −0.13, $P = 0.05$, $R^2 = 0.22$). However, this result is no longer significant when a single shallow pair, Upper and Lower Sorbian (Lusatian_U and Lusatian_L) are removed following the Welch & Waxman test (**Table 5**, **Figure 4**).

We found no evidence of a significant association between rate of word gain and population size in the Indo-European language pairs, nor in gains or losses for the Austronesian and Bantu data (**Tables 4**, **5**, **Figure 4**). One possible explanation for the observation of a significant relationship between rate of language change and population size only in the Indo-European languages is that we expect this dataset to have relatively higher power to detect differences in rates of change. Although the Indo-European dataset has many fewer pairs than the Austronesian or Bantu datasets, the Indo-European word list contains more cognates per category: that is, there are more synonymous lexemes per word (see **Table 6**). The test we use to detect rate differences is broadly based on the Tajima test (Tajima, 1993), the power of which is dependent on the number of variable *sites*, which are columns in DNA alignments in which the sequences being compared differ from each other (Bromham et al., 2000). It may be that the more synonyms recorded per lexical category, the more likely we will record a true gain and less likely we will record a false loss (i.e., a synonym is used less frequently

FIGURE 3 | Histograms of observed and expected numbers of word losses in 14 Indo-European language pairs. Plotted distributions show the expected probability of having a certain number of losses for each language, by fitting Poisson regression to all datapoints. Vertical lines show the observed numbers of losses in each language. The language with the larger speaker population size is colored blue while the language with smaller population size is colored red. The analysis reveals a pattern of a smaller population having a faster rate of word loss, with blue line left to red line particularly when difference in population size is large.

TABLE 5 | Results of least squares regression after Welch & Waxman test on Population size and rate of language change in pairs of Austronesian, Indo-European, and Bantu languages.

	N	Mean	SE	Statistic	P-value	R^2
AUSTRONESIAN LANGUAGES						
Gain	59	0.041	0.024	3.06	0.086	0.034
Loss	59	0.032	0.021	2.31	0.135	0.022
INDO-EUROPEAN LANGUAGES						
Gain	13	−0.047	0.073	0.42	0.532	−0.051
Loss	13	−0.084	0.053	2.52	0.141	0.112
BANTU LANGUAGES						
Gain	47	−0.027	0.074	0.13	0.718	−0.019
Loss	41	0.003	0.018	0.02	0.886	−0.025

N, number of language pairs after removing shallow pairs in regression; Mean, estimated regression coefficient for the relationship between population size and rates of language change; SE, standard error for the regression coefficient; Statistic, F-statistic for least square regression; P-value, results are considered significant at 0.05 level; R^2, adjusted R^2 for least square regression.

whereby only the most frequent word was entered for each lexical category. This means that cognates may be retained in lineages even if not recorded, if there are in less frequent usage than a more predominant form. A gain, in this case, may represent the rise in frequency of one cognate over alternatives, therefore may not involve the loss of an alternative form. Given the differences in the nature of the recorded data, we do not know whether the lack of significant relationships for the Bantu and Austronesian data is due to lack of a consistent association between population size and rates of word gain and loss in these language groups, or due to biases in counts of word gain and loss and thus insufficient power to detect rate differences for these datasets.

DISCUSSION

Languages evolve, creating patterns of descent and relatedness reminiscent of biological species. Because of this, tools from evolutionary biology are being increasingly applied to studying language change (Levinson and Gray, 2012; Gavin et al., 2013; Bromham, 2017). However, we cannot assume that the mechanisms underlying change, or the observed patterns and rates of change, will be the same for both languages and biological lineages.

Evolutionary theory makes clear predictions about the relationship between population size and rates and patterns of genetic change. Selection is more efficient in large populations, so deleterious mutations should be removed more effectively, and advantageous mutations should more rapidly go to fixation. However, in smaller populations, random sampling effects can have a comparatively greater impact on the frequency of genetic variants, so that positively selected mutations may be reduced in frequency by chance, and may thus occasionally be lost rather than going to fixation. Conversely, in small populations, slightly deleterious changes may increase in frequency by chance, and

in a language but not completely lost). This may be a particular problem for the Bantu dataset which has the fewest synonyms as it was collected following Swadesh's (1952, 1955) approach

FIGURE 4 | Contrasts in the number of word gains and word losses against contrasts in population size. Each data point represents a language pair, for Austronesian and Indo-European language families and the Bantu subfamily of the Niger-Congo language family. Red data points are language pairs that have reliable estimates for word gain and loss rates according to Welch & Waxman test.

TABLE 6 | Overall statistics for the three cognate datasets showing the language group, source publication, word list size, average number of cognates per language (±standard deviation) and average number of synonyms per lexical entry across languages (±standard deviation).

Family	Data source	Word list	Cognates	Synonyms
Austronesian	Greenhill et al., 2008	210	198.91 (31.25)	0.95 (0.15)
Indo-European	Bouckaert et al., 2012	207	223.46 (20.95)	1.08 (0.10)
Bantu	Grollemund et al., 2015	100	91.17 (12.29)	0.91 (0.12)

change if the process of substitution is by horizontal spread of variants through the population, rather than by inheritance (Reali and Griffiths, 2010). So, unlike adaptive genetic change in biological populations, it is possible that smaller speaker populations might have a greater rate of adoption of innovations because it is easier for new words to diffuse to all speakers and replace all other variants (Nettle, 1999). It is therefore difficult to predict whether smaller or larger speaker populations should have greater rates of language change, whether patterns should be the same or different for both gains and losses of language elements, and whether we expect similar patterns across all language families or more idiosyncratic associations, particular to given language groups.

Our analysis suggests that, as for Polynesian languages, smaller Indo-European languages have greater rates of word loss from basic vocabulary. This result is consistent with the claim that smaller populations are at greater risk of loss of language elements, and other aspects of culture, due to effects of incomplete sampling of variants over generations. However, we note that the relatively small sample size for this dataset complicates the interpretation of this result. Least squares regression after Welch & Waxman test has the same false positive rate but has much less power than Poisson regression when sample size is small (~ten or fewer pairs, Hua et al., 2015). This makes it difficult to interpret the inconsistent results of these two analyses, as they may be due to their difference in the statistical power. Hence, the negative relationship between rates of loss and population size for Indo-European languages would benefit from additional investigation. We do not find evidence for a negative relationship between population size and word loss rates in the Austronesian and Bantu groups. This finding suggests that either these datasets contain too few language variants to have sufficient power to detect rate differences, or that the increased loss rate in small populations is not a universal phenomenon, or that it is a relatively weak force in some language groups and thus may be overwhelmed by other social, linguistic or demographic factors.

One factor that may be playing a role in the uncertainty in our results, and in the wider debate in general, is that measuring speech community size is notoriously difficult. How exactly does one delimit a speech community (Crystal, 2008) and what degree of proficiency in a language is sufficient to be part of the community (Bloomfield, 1933)? This task is made harder as there are few national censuses that collect detailed speaker statistics. Further, speaker population size can change rapidly with many modern world languages (especially

may eventually drift to fixation, leading to the loss of other variants at that locus (Charlesworth, 2009; Lanfear et al., 2014).

In contrast, the effects of population size on language evolution are not as straightforward to predict, and many alternative hypotheses have been suggested. Large populations of organisms generate more mutations per generation because there are more genomes in the population that can undergo change. Languages with large speaker populations might be expected to generate more innovations (Kline and Boyd, 2010; Collard et al., 2013), however unlike genetic mutation, the processes that create new language variants are not well understood, and may occur by a wider range of mechanisms. Unlike mutation, which is random with respect to utility, introduction of new language variants can be guided by perceived need, and can be regulated by social convention or top-down rules (see Bromham, 2017). Similarly, rates of language change may show different patterns to genetic

the Indo-European languages) experiencing rapid growth over the last few hundred years (Crystal, 2008), while others have experienced catastrophic declines (Bowern, 2010). For the same reasons, the difficulty of obtaining accurate population estimates is also a problem in biology. Furthermore, the relevant parameter for genetic change—the effective population size—is difficult to estimate directly, even when accurate census information is available (Wang et al., 2016). Likewise, there may be an important role played by population and network density—tight-knit networks may inhibit change, while loosely integrated speech communities (regardless of their size), may facilitate change (Granovetter, 1973; Milroy and Milroy, 1992). One way forward here is perhaps to simulate rates of change over a range of population sizes and network topologies (c.f. Reali et al., 2018).

Despite the obvious challenges in obtaining an accurate measure of speaker population size, several previous studies have reported that empirical estimates of population size do correlate with aspects of language change (Hay and Bauer, 2007; Lupyan and Dale, 2010; Bromham et al., 2015a). Therefore, either census population size, as reported in databases such as the Ethnologue, are sufficiently accurate reflections of speaker population size that they are able to reveal significant patterns of language change, or census population size is reflecting some aspect of languages that is connected to change. In either case, the reported relationships with speaker population size invite further investigation.

We can draw two conclusions from these results. Firstly, we provide some evidence that rates of language change can be affected by demographic factors. Even if the effect is not universal, the finding of significant associations between population size and patterns of linguistic change in some languages urges caution for any analysis of language evolution that makes an assumption of uniform rates of change. These results also potentially provide a window on processes of language change in these lineages, providing further impetus to investigate the effect of number of speakers on patterns of language transmission and loss. A more detailed study of language change for a larger number of comparisons might clarify the relationship between population size and word loss rates, particularly within the Indo-European language family.

Secondly, we have shown that the significant patterns of language change identified in a previous study are not a universal phenomenon. Unlike the study of Polynesian languages, we did not find any significant relationships between word gain rate and population size, and the association between loss rates and population size was not evident for all language families analyzed. The lack of universal relationships suggests that it may be difficult to draw general conclusions about the influence of demographic factors on patterns and rates of language change. Many other factors have been proposed to influence rates of language change (Greenhill, 2014) including population density, social structure (Nettle, 1999; Labov, 2007; Ke et al., 2008; Trudgill, 2011), degree of contact, and connectedness with other languages (Matras, 2009; Bowern, 2010), degree of language diffusion within a speech community (Wichmann et al., 2008), degree of bilingualism or multilingualism (Lupyan and Dale, 2010; Bentz and Winter, 2013), language group diversity (Atkinson et al., 2008) and environmental factors such as habitat heterogeneity

and latitude (Bowern, 2010; Blust, 2013; Amano et al., 2014). These factors might mediate or overwhelm the effect of speaker population size.

We find no evidence to support the hypothesis that uptake of new words should be faster in small populations, which is based on the assumption that new words can diffuse more efficiently through a smaller speaker population than a larger one (Nettle, 1999). Nor do we find support for the suggestion that large, widespread languages have a tendency to lose linguistic features a greater rate (Lupyan and Dale, 2010). However, this latter hypothesis is predominantly expected to explain loss of complex linguistic morphology (such as case systems), which may be harder for non-native speakers to learn, rather than basic vocabulary studied here which may be comparatively easier for second language learners to acquire (but see Kempe and Brooks, 2018). Further, our results cannot be interpreted as confirmation of previous studies that suggest there is no effect of population size on rates (Wichmann and Holman, 2009). The detection of significant patterns in rates of lexical change with population size variation in the Polynesian and Indo-European languages, but the failure to identify similar patterns in the Bantu and Austronesian data, suggests that patterns of rates may need to be investigated on a case-by-case basis.

The failure to find a consistent association between population size and rate of change for languages means that analogies drawn between biological and linguistic evolution must be carefully considered to make sure that they are appropriate for linguistic evolution (Bowern and Evans, 2014). For example, patterns of human migration can leave similar traces on both genetic and linguistic diversity (Hurles et al., 2003; Hunley et al., 2007, 2008; Longobardi et al., 2015), but even though the patterns are the same, the underlying mechanisms may not be identical. The observation of decreasing phoneme inventories along chains of human migration has been attributed to serial founder effects (Trudgill, 2004; Atkinson, 2011). While founder effect is likely to influence genetic variability, because a small number of colonists cannot carry all of the genetic variation of the parent population, it might not have the same effect on language variants, as the founding population may use all the main variants in basic vocabulary. Similarly, while a correlation between lineage diversity and rate of change has been reported for both genetic and linguistic evolution (Pagel et al., 2006; Atkinson et al., 2008; Lanfear et al., 2010; Bromham et al., 2015a), it may not reflect a shared mechanism: while formation of new languages may drive higher rates of word turnover, speciation itself is unlikely to drive faster mutation rates in molecular evolution. Our results suggest that the population size effects may be another example of a pattern that is superficially similar between linguistic and biological evolution, yet may be driven by different mechanisms.

However, although the processes underlying language change and genetic change may be different, many of the same analytical tools can be used in the study of both biological and language evolution (see Bromham, 2017). This point was well recognized by early promoters of cross-disciplinary dialogue between evolutionary biology and historical linguistics (Morpugo Davies, 1975), such as Charles Darwin, August Schleicher, and Charles Lyell (Lyell, 1863; Schleicher, 1869; Darwin, 1871). For example,

Schleicher's analogy between borrowing from a foreign language and biological cross-breeding did not imply the same mechanism for both, yet both have the effect of confounding attempts to represent evolutionary history as a bifurcating phylogeny (List et al., 2014). Yet the same solutions may apply to both processes, regardless of their mechanistic origin, such as representation of relationships as a network rather than a tree. Similarly, the shared problem of phylogenetic non-independence due to shared inheritance applies to both languages and species despite the many differences in mode of evolutionary change. While some solutions may be more readily applied to cross-species analysis, due to the availability of phylogenies for many groups, other solutions can be applied more readily to both languages and species, even in the absence of a phylogeny. We demonstrate here that sister pairs analysis is a viable solution to Galton's problem, and it can be applied using information from widely available language taxonomies.

CONCLUSION

Our results show that some of the variation of rates of lexical change in languages can, in some cases, be attributable to differences in speaker population size. Significant correlations between population size and rate of word loss were identified for Indo-European languages, but not for Austronesian and Bantu languages. One possible explanation for the negative relationship between speaker population size and loss rates is that language evolution shares similar mechanisms with genetic evolution, because both show patterns of greater rates of loss of variation in small populations. However, the lack of significant relationships between word gain and loss in two other large language groups— Austronesian and Bantu—warns that we cannot reliably predict variation in rates of linguistic evolution by extrapolation from general principles. By demonstrating that differences can exist in rates of change even between closely related languages, our results caution against assuming uniform rates of change across all languages, and suggest that in some cases the rates of change may be consistently influenced by demographic factors.

AUTHOR CONTRIBUTIONS

LB, CW, XH, and SG: Conceived the project and wrote the paper; SG, CW, and HS: Collected data; XH: Analyzed data.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Amano, N., Piper, P. J., Hung, H.-C., and Bellwood, P. (2013). Introduced domestic animals in the neolithic and metal age of the Philippines: evidence from Nagsabaran, Northern Luzon. *J. Island Coast. Archaeol.* 8, 317–335. doi: 10.1080/15564894.2013.781084

Amano, T., Sandel, B., Eager, H., Bulteau, E., Svenning, J. C., Dalsgaard, B., et al. (2014). Global distribution and drivers of language extinction risk. *Proc. R. Soc. B Biol. Sci.* 281:20141574. doi: 10.1098/rspb.2014.1574

Anthony, D. W., and Ringe, D. (2015). The Indo-European homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* 1, 199–219. doi: 10.1146/annurev-linguist-030514-124812

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–349. doi: 10.1126/science.1199295

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., and Pagel, M. (2008). Languages evolve in punctuational bursts. *Science* 319:588. doi: 10.1126/science.1149683

Baldini, R. (2015). Revisiting the effect of population size on cumulative cultural evolution. *J. Cogn. Cult.* 15, 320–336. doi: 10.1163/15685373-12342153

Bentley, R. A., Hahn, M. W., and Shennan, S. J. (2004). Random drift and culture change. *Proc. R. Soc. Lond. Series B Biol. Sci.* 271, 1443–1450. doi: 10.1098/rspb.2004.2746

Bentley, R. A., Lipo, C. P., Herzog, H. A., and Hahn, M. W. (2007). Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* 28, 151–158. doi: 10.1016/j.evolhumbehav.2006.10.002

Bentz, C., and Winter, B. (2013). Languages with more second language learners tend to lose the nominal case. *Lang. Dyn. Change* 3, 1–27. doi: 10.1163/22105832-13030105

Bergsland, K., and Vogt, H. (1962). On the validity of glottochronology. *Curr. Anthropol.* 3, 115–153. doi: 10.1086/200264

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., et al. (2009). Genetic and demographic implications of the bantu expansion: insights from human paternal lineages. *Mol. Biol. Evol.* 26, 1581–1589. doi: 10.1093/molbev/msp069

Blasi, D., Michaelis, S. M., and Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nat. Hum. Behav.* 1, 723–729. doi: 10.1038/s41562-017-0192-4

Bloomfield, L. (1933). *Language.* London: Allen & Unwin.

Blust, R. (2000). "Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages," in *Time Depth in Historical Linguistics*, eds C. Renfrew, A. McMahon, and L. Trask (Cambridge: McDonald Institute for Archaeological Research), 311–331.

Blust, R. (2013). *The Austronesian Languages.* Canberra, ACT: Asia-Pacific Linguistics.

Blust, R. (2015). "Southeast Asian islands and oceania: Austronesian linguistic history,"in *The Global Prehistory of Human Migration*, ed P. Bellwood (New Jersey, NJ: Wiley-Blackwell), 276–283.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., et al. (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337, 957–960. doi: 10.1126/science. 1219669

Bowern, C. (2010). Correlates of language change in hunter-gatherer and other 'small' languages. *Lang. Linguist. Compass* 4, 665–679. doi: 10.1111/j.1749-818X.2010.00220.x

Bowern, C., and Evans, B. (2014). "Foundations of the new historical linguistics," in *The Routledge Handbook of Historical Linguistics*, eds C. Bowern and B. Evans (London: Routledge), 1–42.

Bromham, L. (2016). Testing hypotheses in macroevolution. *Stud. Hist. Philos. Sci. A* 55, 47–59. doi: 10.1016/j.shpsa.2015.08.013

Bromham, L. (2017). Curiously the same: swapping tools between linguistics and evolutionary biology. *Biol. Philos.* 32, 855–886. doi: 10.1007/s10539-017-9594-y

Bromham, L., Rambaut, A., Hendy, M. D., and Penny, D. (2000). The power of relative rates tests depends on the data. *J. Mol. Evol.* 50, 296–301. doi: 10.1007/s002399910034

Bromham, L., Hua, X., Fitzpatrick, T. G., and Greenhill, S. J. (2015a). Rate of language evolution is affected by population size. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2097–2102. doi: 10.1073/pnas.1419704112

Bromham, L., Hua, X., Lanfear, R., and Cowman, P. (2015b). Exploring the relationships between mutation rates, life history, genome size, environment and species richness in flowering plants. *Am. Nat.* 185, 507–524. doi: 10.1086/680052

Caldwell, C. A., and Millen, A. E. (2010). Human cumulative culture in the laboratory: effects of (micro) population size. *Learn. Behav.* 38, 310–318. doi: 10.3758/LB.38.3.310

Chang, W., Cathcart, C., Hall, D., and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91, 194–244. doi: 10.1353/lan.2015.0005

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature* 10, 195–205. doi: 10.1038/nrg2526

Collard, M., Ruttle, A., Buchanan, B., and O'Brien, M. J. (2013). Population size and cultural evolution in nonindustrial food-producing societies. *PLoS ONE* 8:e72628. doi: 10.1371/journal.pone.0072628

Crystal, D. (2008). Two thousand million? *Engl. Today* 24, 3–6. doi: 10.1017/S0266078408000023

Currie, T. E., Meade, A., Guillon, M., and Mace, R. (2013). Cultural phylogeography of the Bantu languages of sub-Saharan Africa. *Proc. R. Soc. B Biol. Sci.* 280:20130695. doi: 10.1098/rspb.2013.0695

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex.* London: John Murray.

de Filippo, C., Bostoen, K., Stoneking, M., and Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc. R. Soc. B Biol. Sci.* 279, 3256–3263. doi: 10.1098/rspb.2012.0318

Derex, M., Beugin, M. P., Godelle, B., and Raymond, M. (2013). Experimental evidence for the influence of group size on cultural complexity. *Nature* 503, 389–391. doi: 10.1038/nature12774

Derksen, R. (2008). *Etymological Dictionary of the Slavic Inherited Lexicon.* Leiden: Brill.

Durie, M., and Ross, M. D. (1996). *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* Oxford: Oxford University Press.

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* 125, 1–15. doi: 10.1086/284325

Gavin, M. C., Botero, C. A., Bowern, C., Colwell, R. K., Dunn, M., Dunn, R. R., et al. (2013). Toward a mechanistic understanding of linguistic diversity. *Bioscience* 63, 524–535. doi: 10.1525/bio.2013.63.7.6

Gavin, M. C., and Sibanda, N. (2012). The island biogeography of languages. *Glob. Ecol. Biogeogr.* 21, 958–967. doi: 10.1111/j.1466-8238.2011.00744.x

Geertz, C. (1959). Form and variation in balinese village structure. *Am. Anthropol.* 61, 991–1012. doi: 10.1525/aa.1959.61.6.02a00060

Goodenough, W. H. (1957). Oceania and the problem of controls in the study of cultural and human evolution. *J. Polyn. Soc.* 66, 146–155.

Granovetter, M. S. (1973). The strength of weak ties. *Am. J. Sociol.* 78, 1360–1380. doi: 10.1086/225469

Gray, R. D., and Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439. doi: 10.1038/nature02029

Gray, R. D., Atkinson, Q. D., and Greenhill, S. J. (2011). Language evolution and human history: what a difference a date makes. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1090–1100. doi: 10.1098/rstb.2010.0378

Gray, R. D., Drummond, A. J., and Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science* 323, 479–483. doi: 10.1126/science.1166858

Greenhill, S. J. (2014). "Demographic correlates of language diversity," in *Routledge Handbook of Historical Linguistics,* eds C. Bowern and B. Evans (London: Routledge), 555–578.

Greenhill, S. J., Blust, R., and Gray, R. D. (2008). The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evol. Bioinformat.* 4, 271–283. doi: 10.4137/EBO.S893

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., and Pagel, M. (2015). Bantu expansion shows that habitat alters the route and

pace of human dispersals. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13296–13301. doi: 10.1073/pnas.1503793112

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. doi: 10.1038/nature14317

Hammarström, H., Forkel, R., Haspelmath, M., and Bank, S. (2016). *Glottolog. 2.7.* Available online at: http://glottolog.org (Accessed August 05, 2017).

Harvey, P. H., and Pagel, M. (1991). *The Comparative Method in Evolutionary Biology.* Oxford: Oxford University Press.

Hay, J., and Bauer, L. (2007). Phoneme inventory size and population size. *Language* 83, 388–400. doi: 10.1353/lan.2007.0071

Henrich, J. (2004). Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the tasmanian case. *Am. Antiq.* 69, 197–214. doi: 10.2307/4128416

Hoijer, H. (1956). Lexicostatistics: a critique. *Language* 32, 49–60. doi: 10.2307/410652

Hua, X., Cowman, P., Warren, D., and Bromham, L. (2015). Longevity is linked to mitochondrial mutation rates in rockfish: a test using poisson regression. *Mol. Biol. Evol.* 32, 2633–2645. doi: 10.1093/molbev/msv137

Hung, H.-C., Carson, M. T., Bellwood, P., Campos, F. Z., Piper, P. J., Dizon, E., et al. (2011). The first settlement of remote Oceania: the Philippines to the Marianas. *Antiquity* 85, 909–926. doi: 10.1017/S0003598X00068393

Hunley, K. L., Cabana, G. S., Merriwether, D. A., and Long, J. C. (2007). A formal test of linguistic and genetic coevolution in native central and South America. *Am. J. Phys. Anthropol.* 132, 622–631. doi: 10.1002/ajpa.20542

Hunley, K., Dunn, M., Lindström, E., Reesink, G., Terrill, A., Healy, M. E., et al. (2008). Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genet.* 4:e1000239. doi: 10.1371/journal.pgen.1000239

Hurles, M. E., Matisoo-Smith, E., Gray, R. D., and Penny, D. (2003). Untangling oceanic settlement: the edge of the knowable. *Trends Ecol. Evol.* 18, 531–540. doi: 10.1016/S0169-5347(03)00245-3

Jay, R. R. (1969). *Javanese Villagers: Social Relations in Rural Modjokuto.* Cambridge, MA: MIT Press.

Johnson, L. (1976). A rate of change index for language. *Lang. Soc.* 5, 165–172. doi: 10.1017/S0047404500007004

Ke, J. Y., Gong, T., and Wang, W. S. Y. (2008). Language change in social networks. *Commun. Comput. Phys.* 3, 935–949.

Kempe, V., and Brooks, P. J. (2018). Linking adult second language learning and diachronic change: a cautionary note. *Front. Psychol.* 9:480. doi: 10.3389/fpsyg.2018.00480

Kirch, P. V., and Green, R. C. (1987). History, phylogeny, and evolution in Polynesia. *Curr. Anthropol.* 28, 431–456. doi: 10.1086/203547

Kline, M. A., and Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Proc. R. Soc. B Biol. Sci.* 277, 2559–2564. doi: 10.1098/rspb.2010.0452

Ko, A. M., Chen, C. Y., Fu, Q., Delfin, F., Li, M., Chiu, H. L., et al. (2014). Early austronesians: into and out of Taiwan. *Am. J. Hum. Genet.* 94, 426–436. doi: 10.1016/j.ajhg.2014.02.003

Labov, W. (2007). Transmission and diffusion. *Language* 83, 344–387. doi: 10.1353/lan.2007.0082

Lanfear, R., Ho, S. Y. W., Love, D., and Bromham, L. (2010). Mutation rate influences diversification rate in birds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20423–20428. doi: 10.1073/pnas.1007888107

Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends Ecol. Evol.* 29, 33–41. doi: 10.1016/j.tree.2013.09.009

Lev-Ari, S. (2017). Talking to fewer people leads to having more malleable linguistic representations. *PLoS ONE* 12:e0183593. doi: 10.1371/journal.pone.0183593

Levinson, S. C., and Gray, R. D. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends Cogn. Sci.* 16, 167–173. doi: 10.1016/j.tics.2012.01.007

Lewis, P. M., Simons, G. F., and Fennig, C. D. (2015). *Ethnologue: Languages of the World.* Dallas, TX: SIL International.

Li, S., Schlebusch, C., and Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc. R. Soc. B Biol. Sci.* 281:20141448. doi: 10.1098/rspb.2014.1448

List, J. M., Nelson-Sathi, S., Geisler, H., and Martin, W. (2014). Networks of lexical borrowing and lateral gene transfer in language

and genome evolution. *Bioessays* 36, 141–150. doi: 10.1002/bies.201 300096

Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., et al. (2015). Across language families: genome diversity mirrors linguistic variation within Europe. *Am. J. Phys. Anthropol.* 157, 630–640. doi: 10.1002/ajpa.22758

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Lyell, C. (1863). *The Geological Evidence of the Antiquity of Man*. London: John Murray.

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

Milroy, J., and Milroy, L. (1985). Linguistic change, social network and speaker innovation. *J. Linguist.* 21, 339–384. doi: 10.1017/S0022226700010306

Milroy, L., and Milroy, J. (1992). Social network and social class: toward an integrated sociolinguistic model. *Lang. Soc.* 21, 1–26. doi: 10.1017/S0047404500015013

Montano, V., Ferri, G., Marcari, V., Batini, C., Anyaele, O., Destro-Bisol, G., et al. (2011). The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. *Mol. Ecol.* 20, 2693–2708. doi: 10.1111/j.1365-294X.2011.05130.x

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language* 88, 877–893. doi: 10.1353/lan.2012.0087

Morpurgo Davies, A. (1975). "Language classification in the nineteenth century," in *Current Trends in Linguistics*, Vol. 13, ed T. A. Sebeok (The Hague: Mouton), 607–716.

Neiman, F. D. (1995). Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *Am. Antiq.* 60, 7–36. doi: 10.2307/282074

Nettle, D. (1999). Is the rate of linguistic change constant? *Lingua* 108, 119–136. doi: 10.1016/S0024-3841(98)00047-3

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1829–1836. doi: 10.1098/rstb.2011.0216

Pagel, M., Venditti, C., and Meade, A. (2006). Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314, 119–121. doi: 10.1126/science.1129647

Pakendorf, B., de Filippo, C., and Bostoen, K. (2011). Molecular Perspectives on the Bantu expansion: a synthesis. *Lang. Dyn. Change* 1, 50–88. doi: 10.1163/221058211X570349

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Pawley, A. (1967). The relationships of Polynesian outlier languages. *J. Polyn. Soc.* 76, 259–296.

Rea, J. A. (1958). Concerning the validity of lexicostatistics. *Int. J. Am. Linguist.* 24, 145–150. doi: 10.1086/464447

Read, D. (2012). Population size does not predict artifact complexity: analysis of data from Tasmania, Arctic hunter-gatherers, and Oceania fishing groups. *Hum. Comp. Syst. UCLA.* Available online at: https://escholarship.org/uc/item/61n4303q#author

Reali, F., and Griffiths, T. L. (2010). Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proc. R. Soc. B Biol. Sci.* 277, 429–436. doi: 10.1098/rspb.2009.1513

Reali, F., Chater, N., and Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: how population size affects language. *Proc. R. Soc. B Biol. Sci.* 285:20172586. doi: 10.1098/rspb.2017.2586

Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Cape.

Richerson, P. J., Boyd, R., and Bettinger, R. L. (2009). Cultural innovations and demographic change. *Hum. Biol.* 81, 211–235. doi: 10.3378/027.081.0306

Roberts, S., and Winters, J. (2013). Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLoS ONE* 8:e70902. doi: 10.1371/journal.pone.0070902

Sankoff, D. (1970). On the rate of replacement of word-meaning relationships. *Language* 46, 564–569. doi: 10.2307/412307

Schleicher, A. (1869). *Darwinism Tested by the Science of Language*. London: John Camden Hotton.

Shennan, S. (2001). Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* 11, 5–16. doi: 10.1017/S0959774301000014

Shennan, S. J., and Wilkinson, J. R. (2001). Ceramic style change and neutral evolution: a case study from Neolithic Europe. *Am. Antiq.* 577–593. doi: 10.2307/2694174

Spriggs, M. (2010). "I was so much older then, I'm younger than that now': why the dates keep changing for the spread of Austronesian languages a journey through Austronesian and Papuan linguistic and cultural space papers," in *Honour of Andrew Pawley*, eds J. Bowden, N. P. Himmelmann, and M. D. Ross (Canberra, ACT: Pacific Linguistics), 113–140.

Spriggs, M. (2011). Archaeology and the Austronesian expansion: where are we now? *Antiquity* 85, 510–528. doi: 10.1017/S0003598X000 67910

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proc. Am. Philos. Soc.* 96, 452–463.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *Int. J. Am. Linguist.* 21, 121–137. doi: 10.1086/464321

Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135, 599–607.

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

Trudgill, P. (2004). Linguistic and social typology: the Austronesian migration and phoneme inventories. *Linguist. Typol.* 8, 305–320. doi: 10.1515/lity.2004.8. 3.305

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Wang, J., Santiago, E., and Caballero, A. (2016). Prediction and estimation of effective population size. *Heredity* 117, 193–206. doi: 10.1038/hdy.20 16.43

Welch, J. J., and Waxman, D. (2008). Calculating independent contrasts for the comparative study of substitution rates. *J. Theor. Biol.* 251, 667–678. doi: 10.1016/j.jtbi.2007.12.015

Wichmann, S., and Holman, E. W. (2009). Population size and rates of language change. *Hum. Biol.* 81, 259–274. doi: 10.3378/027.081.0308

Wichmann, S., Rama, T., and Holman, E. W. (2011). Phonological diversity, word length, and population sizes across languages: the ASJP evidence. *Linguist. Typol.* 15, 177–197. doi: 10.1515/lity.20 11.013

Wichmann, S., Stauffer, D., Schulze, C., and Holman, E. W. (2008). Do language change rates depend on population size? *Adv. Comp. Syst.* 11, 357–369. doi: 10.1142/S0219525908001684

Check for
updates

Language Structures May Adapt to the Sociolinguistic Environment, but It Matters What and How You Count: A Typological Study of Verbal and Nominal Complexity

Kaius Sinnemäki[1] and Francesca Di Garbo[2]*

[1] Department of Languages, University of Helsinki, Helsinki, Finland, [2] Department of Linguistics, Stockholm University, Stockholm, Sweden

In this article we evaluate claims that language structure adapts to sociolinguistic environment. We present the results of two typological case studies examining the effects of the number of native (=L1) speakers and the proportion of adult second language (=L2) learners on language structure. Data from more than 300 languages suggest that testing the effect of population size and proportion of adult L2 learners on features of verbal and nominal complexity produces conflicting results on different grammatical features. The results show that verbal inflectional synthesis adapts to the sociolinguistic environment but the number of genders does not. The results also suggest that modeling population size together with proportion of L2 improves model fit compared to modeling them independently of one another. We thus argue that surveying population size alone may be insufficient to detect possible adaptation of linguistic structure to the sociolinguistic environment. Rather, other features, such as proportion of L2 speakers, prestige and social network density, should be studied, and if demographic numeric data are used, they should not be used in isolation but rather in competition with other sociolinguistic features. We also suggest that not all types of language structures within a given grammatical domain are equally sensitive to the effect of sociolinguistic variables, and that more exploratory studies are needed before we can arrive at a reliable set of grammatical features that may be potentially most (and least) adaptive to social structures.

Keywords: inflectional synthesis, grammatical gender, language complexity, population size, second language learning, sociolinguistic environment, language typology, adaptation

1. INTRODUCTION

Recent research suggests that linguistic structures adapt to the sociocultural environment in which languages are spoken (Ladd et al., 2015). Since languages are acquired and used in different social contexts, those contexts may bias acquisition and usage: linguistic structures become adapted to these social niches and this, over time, may be reflected in typological distributions (Lupyan and Dale, 2010; Sinnemäki, 2014). Central ideas in this approach have been:

i. Small communities with dense social networks, few adult learners, and a great deal of shared knowledge favor linguistic structures that are difficult for adults (e.g., irregularity in inflectional paradigms);

ii. Large communities with loose social networks, more adult learners, and less shared knowledge favor more regular and easier to learn linguistic structures (e.g., regularity in inflectional paradigms, transparency; Trudgill, 2011b).

These ideas have also been tested empirically, with focus on the relationship between language complexity and community size. However, the results have been conflicting. For instance, the number of cases seems to correlate inversely with the number of native speakers (Lupyan and Dale, 2010), but according to Bentz and Winter (2013) it correlates only with the proportion of L2 speakers in the community, and not with overall community size.

In this paper, we review a number of studies on language complexity, population size, and linguistic adaptation and contrast these findings with two empirical studies of our own. Study 1 focuses on verbal inflectional synthesis and Study 2 on grammatical gender. We take these features as instances of morphological complexity in the verbal and nominal domain, respectively. With respect to verbal inflectional synthesis, we find that only the number of L1 speakers has a significant effect on verbal complexity when the sociolinguistic features are modeled independently of one another. We also find that the proportion of L2 speakers has a significant effect on verbal complexity when modeled together with the number of L1 speakers in one and the same model. This suggests that the features may conspire in shaping language structure. With respect to grammatical gender, we find no significant effect of the sociolinguistic variables under study on the number of gender distinctions, both when the demographic predictors are considered independently and when they are modeled together. We also observe a confounding effect of data coding structure on the patterns detected by our models.

We highlight the discrepancies between the results and discuss the factors that could motivate them. Moreover, we argue that in order to establish more solid results on linguistic adaptation, demographic features must be studied in competition with each other and further combined with in depth studies of sociolinguistic and sociohistorical profiles. We also suggest that not all variables that describe crosslinguistic variation in a given domain of grammar may be equally suited to investigate how and whether this domain adapts to sociolinguistic structures. Selecting the right typological variables to test adaptive responses of language structures to social structures is thus crucial to studies in this field and requires going beyond existing typological databases.

2. BACKGROUND

One of the main tenets of functional-typological linguistics is the idea that language structures are shaped by properties of human cognition as well as by the dynamics of social interaction (Beckner et al., 2009). The mechanisms by which languages adapt to their contexts of use are also considered to be the driving force of language variation and change (Givón, 2009; Bybee, 2010; von Mengden and Coussé, 2014). During the last decades a new trend of studies has developed within the language sciences, which tests these assumptions empirically by investigating the relationship between typological, sociolinguistic, and environmental variables based both on micro-level qualitative investigations (Kusters, 2003; DeLancey, 2014), and large scale quantitative studies (e.g., Lupyan and Dale, 2010; Dediu and Cysouw, 2013; Everett et al., 2015; for more references, see the review by Ladd et al., 2015). Phonemic inventory size, tone, degree of inflectional synthesis, inflectional morphology, lexical diversity, and lexical stability are some of the domains of language variation investigated so far within this approach, and in connection with an array of sociolinguistic and environmental factors such as population size, proportion of L2 speakers, number of neighboring languages, and humidity. In this paper, we study linguistic adaptation from the perspective of one domain of linguistic variation, morphological complexity, as measured through verbal inflectional synthesis and number of grammatical gender distinctions. We explore typological variation in these areas of grammar in relationship with demographic data on first language (=L1) and second language (=L2) speakers. We first introduce the grammatical phenomena under investigation. We then review a number of studies that have looked at the interactions between these features and the sociolinguistic variables under study.

2.1. Morphological Complexity and Verbal Inflectional Synthesis

Morphological complexity, taken as a measure of the degree of grammatical elaboration and internal structuring of words, has traditionally attracted much attention in typology. Since the nineteenth century languages were classified in three holistic morphological types: isolating (or analytic), fusional (or inflective), and agglutinative. It was believed that one parameter of typological variation, morphology, had predictive scope on the overall appearance of entire languages. This one-dimensional, holistic approach has been later rejected in typology, and, starting with the work of Sapir (1921), alternative classifications that break morphological typology into multiple and mutually interacting parameters have been proposed (see Plank, 1998, 1999, for more detailed review of the discussion). These more recent classifications cover multiple dimensions of variation, such as the internal complexity of the word (analytic vs. synthetic), the nature of morpheme boundaries (agglutination vs. fusion), and the extent to which several roots may be combined into one and the same word (incorporation).

In recent crosslinguistic research degree of inflectional synthesis has been especially a subject of interest. This label is used to refer to the number of morphemes or morphological categories that are realized in a word. Inflection is here defined as "those categories of morphology that are *regularly responsive to the grammatical environment* in which they are expressed" (Bickel and Nichols, 2007, p. 169). The main difference to

derivation is that inflection is responsive to the grammatical (that is, morphological or syntactic) environment, whereas derivation is responsive to the lexical environment but not to the grammatical environment. For instance, in English the number of the subject is reflected in the morphological choices of agreement on the verb in sentences such as *the waiter likes ice cream* vs. *kids like ice cream*. In these examples agreement determines morphological choices based on the syntactic environment, whereas the choice of a derivational category, as in *waiter* vs. *waitress*, is entirely a lexical matter.

If a grammatical category, such as person, is expressed inflectionally as in the word *like-s*, the construction is said to be synthetic but if the category is expressed through a separate word, as in *will do*, the construction is said to be analytic (Bickel and Nichols, 2013).

In analytic constructions the relationship between the elements is syntactic and not morphological and the elements do not make up a grammatical word. It is well-known that grammatical and phonological criteria of wordhood do not coincide cross-linguistically (Dixon and Aikhenvald, 2002; Haspelmath, 2011). In the *AUTOTYP* database, which we use as our data source for verbal inflectional synthesis (see also section 3.2), this challenge has been solved by focusing on grammatical words. Synthesis is a matter of grammatical words but it is independent of phonological binding and therefore grammatical words can be composed of phonologically distinct words (Bickel and Nichols, 2007, 2013). The crucial issue here is that if a phonologically distinct word cannot be used alone without the verb and also in different orderings, then that word is part of the same grammatical word with the verb. Bickel and Nichols (2013) give the example of the tense marker *làay* in Hakha Lai (Tibeto-Burman). This marker is an independent phonological word as it bears tone and contains two moras, but it cannot be used independently of the verb and it always occurs in the same position relative to the verb, as in (1).

(1) Hakha-Lai (Tibeto-Burman; Bickel and Nichols, 2013)

    A-nii   làay.
    3SG-laugh FUT
    'She/he will laugh.'

Together with the verb, the tense marker *làay* in Hakha Lai is an example of grammatical word.

The notion of word-level semantic density has also been used in the literature to refer to degree of inflectional synthesis (Bickel and Nichols, 2007, p. 188–193). Vietnamese is a language with very low semantic density of words, since words generally consist of only a single morpheme, as in (2). More toward the other end of the synthesis/semantic density scale are languages with very complex word structure, such as Turkish, illustrated in (3), which may attach up to ten or more inflectional and derivational morphemes into one and the same grammatical word.

(2) Vietnamese (Austro-Asiatic; Thompson, 1987, p. 207)

    Tôi sẽ   đi.
    1SG FUT go
    'I will go.'

(3) Turkish (Turkic; Göksel and Kerslake, 2005, p. 74)

    Döğ-üş-tür-t-ül-me-yebil-iyor-muş-sunuz-dur.
    beat-RECP-CAUS-CAUS-PASS-NEG-PSB-IPFV-EVID.COP-
    2PL-GM
    'It is presumably the case that you sometimes were not made to fight.'

As shown in (3), morphological words in highly synthetic languages may sometimes correspond to a whole sentence in other languages.

Languages with a degree of inflectional synthesis comparable to Turkish are rather common around the world. Comparative data in the domain of verbal inflection suggests that almost half (44%; *n* = 145) of the world's languages have the same or higher degree of synthesis than Turkish (Bickel and Nichols, 2013). This distribution suggests that high word-internal complexity is not particularly difficult for children to acquire and for native speakers to use. Evidence from language acquisition supports this conclusion. By the age of two Turkish children fully master the nominal inflectional system and most of the verbal inflectional system as well (Slobin, 2005). Children also acquire inflectional cues equally or even faster than alternative cues, such as word order or prosody (Slobin and Bever, 1982). From the point of view of adult language use, high degree of synthesis should also pose no problems, whether in production or comprehension (see Kusters, 2003, p. 46–52 and references). However, compared to native speakers, adult learners are overall less sensitive to morphological structure during language processing in their L2 (Clahsen et al., 2010). Morphologically complex words have higher informational complexity and thus higher processing cost in word recognition (Moscoso del Prado Martín et al., 2004). Verbal inflection in particular poses major problems to adult learners but much less so to child learners (see Parodi et al., 2004, p. 670, and references there). This difficulty that adults have in learning and using complex inflection is related to a more general pattern supported by neurocognitive evidence: learning grammar in procedural memory creates more problems for adult learners than for L1 learners while acquiring lexical knowledge in declarative memory poses fewer such problems for adults (Ullman, 2005). This learning bias toward declarative memory means that adult learners prefer lexical strategies and periphrastic constructions over grammatical strategies, especially at low levels of exposure.

2.2. Morphological Complexity and Grammatical Gender

Grammatical gender is one of the possible strategies that languages use to partition nouns into classes. Typically, these classifications may at least partially rest on semantic distinctions based on natural gender (as in the sex-based systems of the Romance languages), or on other parameters, such as animacy, size, or shape (as in the non-sex-based systems of the Bantu languages).

The most important definitional property of grammatical gender systems is that the encoding of grammaticalized classificatory distinctions is *displaced*. It does not only (or

not necessarily) occur on nouns, but *must* appear on those words that are engaged in a syntactic relation with nouns. In languages with grammatical gender, attributive modifiers, predicates, and pronouns are the word classes that most typically carry gender marking through their inflectional morphology. The syntactic relation between nouns and carriers of gender marking is traditionally called *agreement*. Within typological literature, nouns are referred to as *controllers* of the agreement relationship because their gender controls the type of marking encoded through agreement. Conversely, those words whose inflectional morphology varies in agreement with the gender of a noun are labeled *targets* of the agreement relationship. Dahl (2004) regards grammatical gender as one of the most typical instances of *mature* grammatical phenomena in language: gender systems are long-lived features of language families and they usually presuppose intricate, non-trivial processes of grammaticalization.

In Italian (Indo-European, Romance) nouns are assigned to one of two genders: the masculine and the feminine. For at least a portion of the nominal lexicon (humans and higher animates), gender assignment is predicted based on sex. Displaced gender marking occurs on attributive modifiers, some of the pronouns, and past participles. Example (4) illustrates gender marking in Italian, both within and outside the noun phrase.

(4) Italian (Indo-European; constructed example)

    a. La   macchina è stat-a   consegnat-a
       DEF.F.SG car.F.SG is been-F.SG delivered-F.SG
       ieri
       yesterday
       'The car has been delivered yesterday'

    b. Il   sole   è tramontat-o
       DEF.M.SG sun.M.SG is set-M.SG
       'The sun has set.'

As is praxis within typology we use the label grammatical gender not just to refer to systems of noun classification of the Italian type, that is, based on natural gender and on two to three distinctions, but also to those systems that are typically found in many African and some Papuan languages, and that are often labeled *noun classes*. These systems may have up to almost 20 different agreement classes which are not always clearly motivated semantically. In Mufian (Torricelli; spoken in the East Sepik region of New Guinea), different suffixes on the noun and adjective as well as prefixes on the verb stand for different noun classes; **Table 1** shows a selection of these.

Grammatical gender, as defined above, can be associated with morphological complexity in two ways. Syntagmatically, gender marking is distributed over an utterance through agreement patterns, and several entities within that utterance may thus redundantly point to the gender of the controller noun. Paradigmatically, each word class that is sensitive to gender inflections typically displays as many forms as the number of genders to be distinguished. For instance, the Italian definite article has two forms distinguishing masculine and feminine

TABLE 1 | A selected set of noun classes in Mufian (Alungum et al., 1978, p. 93).

Class	Example singular/plural	Gloss	Noun suffix singular/plural	Verb prefix singular/plural
1	bol / bongof	"pig"	-l / -ngof	l- / f-
2	éngél / angof	"name"	-ngél / ngof	g- / f-
3	nalof / nalelef	"tooth"	-f / -lef	f- / f-
5	batéwin / batéwis	"child"	-n / -s	n- / s-
17	kos / kos	"course"	-s / -s	s- / s-

gender both in the singular and in the plural (for a total of four distinct forms). In this paper, we do not look at these dimensions directly, but focus instead on the number of gender distinctions in a language. This is estimated based on the number of distinguishable agreement patterns, and thus at least indirectly relates to paradigmatic complexity, that is, to the number of subdistinctions available in a linguistic category (see Moravcsik and Wirth, 1986).

Corbett (2013a) identifies the presence of a gender system in 112 out of 257 sampled languages. The distribution of grammatical gender in the languages of his sample is rather skewed, both geographically and genealogically, which reflects an actual tendency in the overall distribution of gender systems. Gender systems are very common in some areas of the world, such as Africa and Eurasia, but rather rare in others, such as North America. This geographical bias is directly connected to a genealogical bias. The presence of grammatical gender is often a distinctive, stable feature of individual language families, whose members do usually also cluster geographically. Moreover, the presence of grammatical gender across language families is reinforced by areal contiguity. Even though geographically biased, the pervasive distribution of grammatical gender within individual language families and coherent linguistic areas suggests that under normal circumstances of language transmission gender systems are easily acquired and mastered by children and native speakers.

This is indeed confirmed in the literature. Studies of L1 acquisition of grammatical gender, focusing on different L1s and different types of gender systems, show that children are generally able to master at least aspects of the gender system of their native language by the age of three. They are usually better at relying on phonological rather than semantic cues for gender assignment, and the frequency of individual nouns in every-day speech affects how much they use a given gender marking pattern (for language specific studies of the acquisition of grammatical gender see, for instance, Suzman, 1980; Mulford, 1985; Mills, 1986; Desmuth, 2000; Eichler et al., 2013; Gagliardi and Lidz, 2014). Similarly, studies of language processing and comprehension show that gender marking plays an important role in processes of semantic and syntactic disambinguation in adult native speaker usage (see, for instance, Gunter and Friederici, 2000; Barber and Carreiras, 2005). Even though unproblematic in L1 acquisition and native speaker usage, grammatical gender is a challenge for non-native adult learners, and exactly for the same reasons that we mentioned in the case of verbal inflectional synthesis. Mastering gender marking presupposes the acquisition of complex patterns

of inflection, which L2 speakers tend to struggle with, and thus to avoid[1].

2.3. Does Morphology Adapt to Social Structure?

Processing difficulties that language users face are one of the driving factors behind language change if, following a usage-based approach to language, we assume that preferences in language use become conventionalized over time (e.g., Sinnemäki, 2014). It has been recently suggested that the processing difficulties that adults face in learning and using an L2 may end up having an effect on the (evolution of the) grammar of the native speakers as well (see e.g., Lupyan and Dale 2010; Bentz and Winter 2013; and references there). The magnitude of this effect crucially depends on the proportion of non-native speakers in the speech population. The larger the proportion of non-native speakers, the more their presence is likely to have an impact on the grammars of L1 users.

Maitz and Németh (2014) compare three types of German varieties against four indicators of morphosyntactic complexity (degree of synthesis being one of them), and to the effect that these varieties represent three distinct sociolinguistic and sociohistorical profiles: one highly standardized contact variety (Standard German), two high contact varieties (Kiche Duits and Unserdeutsch), and one low contact L1 variety (Cimbrian). The results show significant differences between the two types of high contact varieties, on the one hand, and the low contact L1 variety, on the other, with respect to all four parameters of morphosyntactic complexity. The impact of L2 learning on the evolution of ancient language varieties has been also studied. For instance, Skelton (2017) demonstrates that peculiar features of the Ancient Greek dialect of Pamphilia (at the phonological, morphological, syntactic, and lexical level) can be explained as the result of massive influence from Anatolian speakers, who represented the majority of the population in the area and spoke Greek as L2.

Verbal inflectional synthesis and grammatical gender have been shown to be sensitive to the effect of massive L2 learning. For instance, drawing on historical and contemporary data from Quechuan, Swahili, Arabic, and Scandinavian, Kusters (2003) shows that those language varieties which, throughout their history, were characterized by high proportions of adult non-native learners have simpler verbal morphology than their closest cognates with little or no history of exposure to non-native learners. Trudgill (1999, 2001) and McWhorter (2001, 2007) also argue that high contact language varieties, characterized by a significant increase in number of adult learners at some point throughout their history, are likely to lose grammatical gender. Examples of this would be, for instance, Persian, which has lost the gender system preserved by other Iranian languages, or many pidgin and creole languages, which tend to be devoid of grammatical gender irrespectively of the presence of this feature in their lexifiers and/or substrata. Similarly, Kusters (2003) also

shows that gender agreement on verbs tends to simplify as a result of increased language contact. In all these cases loss of gender has been typically explained with the fact that gender marking is substantially afunctional from the point of view of effective communication and thus likely to be weakened/lost in non-native speaker usage. However, recent research by Blasi et al. (2017) on the dynamics of language transmission under creole emergence shows that creole languages do not exhibit any systematic structural simplification with respect to the two gender-related variables that the study accounts for, adjectival adnominal agreement and presence of gender distinctions on personal pronouns. Instead, both variables seem to be sensitive to ancestry, that is, they align with the structural type attested in either the lexifier or the substratum, and do not seem to be directly linked with the sociohistorical background the sampled languages share with other creoles. Whether some aspect of gender may adapt to sociolinguistic environment is thus a matter of current debate and open to exploration from different angles (see also section 3.3.3).

While research on linguistic adaptation in the domain of morphology has largely focused on non-native acquisition as a trigger of simplification (e.g., Kusters, 2003), evidence for the complexification of verbal morphology in the absence of large-scale non-native acquisition has also been provided. DeLancey (2014) showed that two Tibeto-Burman languages spoken in North East India, Boro (Boro-Garo branch of Tibeto-Burman) and Lai (Kuki-Chin branch of Tibeto-Burman) have different morphological profiles and are spoken in very different sociolinguistic environments. Boro, which has very little verbal morphology, is spoken by a large, widely distributed community in the Assam plains where there has historically been, and still is, much interaction with speakers of other languages. Lai, on the contrary, has developed new synthetic verbal morphology not present in proto-Tibeto-Burman and it is spoken in small relatively isolated hill communities in the mountain range which follows the India-Myanmar border. Trudgill (2017) argues that languages with polysynthetic morphology, that is, those with a very high ratio of morphemes per words and possibly also noun incorporation, tend to be spoken by small communities, with fewer than 10,000 speakers. These communities are also relatively isolated and have rather dense social network structure.

Recent quantitative typological research provides further evidence that population structure has an impact on language structures. Lupyan and Dale (2010) modeled the relationship between morphological complexity (measured on the basis of a set of 28 variables taken from the *World Atlas of Language Structures*), and the (log) number of native speakers with generalized linear modeling. In their study, speech community size was taken as a proxy for the degree of adult L2 learning in the community, under the assumption that languages with larger populations are more likely to engage in contact with other speech communities, and to be learned non-natively. The results of the study indicated that smaller languages tend to have higher degrees of morphological complexity than larger languages. This applied across geographical areas and language families, but also within language families. However, speech community size in itself is not the only predictor of change in

---

[1]Naturally, however, a number of factors may interfere with the success rate of non-native acquisition of gender, such as the presence of a gender system in the L1, typological similarities between L1 and L2, age of acquisition, motivation.

language structures, and other sociolinguistic factors may need to be taken into account as well. This point has been made by Trudgill (2011a) in relation to phoneme inventory size and later empirically confirmed by Moran et al. (2012), who show that there is no statistical evidence for a correlation between phoneme inventory size and speech community size (see section 3.3.3).

While Lupyan and Dale (2010) used log number of speakers as a proxy for the degree of adult L2 learning in a given speech community, Bentz and Winter (2013) propose to evaluate the effects of adult L2 learning more directly, by taking into account the proportion of adult L2 learners in a given speech community (the speech community comprised of both native and non-native speakers) and assessing whether this has any effect on the number of grammaticalized case distinctions in a language. Although the sample used by Bentz and Winter (2013) is not particularly large ($n = 66$ languages), their data suggest that there is a strong inverse relationship between the number of cases and the proportion of adult non-native learners in the community: high proportion of adult non-native learners correlates with low number of cases and low proportion of adult non-native learners correlates with high number of cases. To emphasize the importance of measuring the proportion of non-native learners, they also show that, in their data set, population size (native + non-native speakers) has no effect on the number of cases (Bentz and Winter, 2013, p. 11).

In Study 1 we attempt to replicate the results of Lupyan and Dale (2010) by focusing on one dimension of their morphological complexity metric, notably the degree of inflectional synthesis on the verb. The data in their study is based on the chapter by Bickel and Nichols (2013) in *WALS*, which is in turn based on the *AUTOTYP* database. The original *AUTOTYP* data set contains a much more detailed analysis of inflectional synthesis than what was later included in *WALS*. The *WALS* format required authors to keep the number of levels limited for each variable and this means that variable levels are conflated in many chapters, including the one on verbal inflectional synthesis where, for instance, synthesis degrees 6 and 7 are conflated into one category "6-7." This kind of conflation inevitably leads to loss of information, which we attempt to avoid in this paper by using the original and now expanded data of the *AUTOTYP* database (Bickel et al., 2017). The data set has information on inflectional synthesis in 309 languages. With respect to sociolinguistic variables, while, as mentioned above, Lupyan and Dale (2010) worked only with data on population size, in our study we consider both the number of L1 speakers as well as the proportions of L2 speakers. This choice of features models more closely the hypothesis put forward in sociolinguistic typology that the size and structure of a speech community, on the one hand, and the degree of language contact, on the other, should be taken into account simultaneously but also independently of each other (e.g., Trudgill, 2011a).

Dahl (unpublished) tests linguistic adaptation by looking at the relationship between the three *WALS* features devoted to grammatical gender[2] and number of speakers. The results suggest

that no consistent relationship can be found between any of the gender features and the number of speakers a language has (a weak positive correlation is however detected between non-sex-based gender systems and population data). In Study 2, we attempt to replicate these findings with a larger data set ($n = 345$). Differently from Dahl (unpublished) we focus only on one gender feature, the number of gender values, and consider only nominal gender, thus excluding pronominal gender systems, such as the one attested in English, from the data set. With respect to sociolinguistic variables, as in Study 1, we consider the log number of L1 speakers and the proportions of L2 speakers both in isolation and in combination with each other.

3. TYPOLOGICAL CASE STUDIES

We contrast the findings of the earlier research surveyed above with two empirical case studies of our own. The first study deals with the degree of inflectional synthesis on the verb, a common metric of complexity in cross-linguistic research (e.g., Kusters, 2003; Shosted, 2006; de Groot, 2008; Nichols, 2009; Kettunen, 2014). The second study deals with the number of grammatical genders in a language. Recent research regards the number of gender distinctions as one of the three main dimensions of complexity variation in gender systems (Audring, 2014, 2017; Di Garbo, 2016). Both degree of inflectional synthesis on the verb and number of gender distinctions can be interpreted straightforwardly from the perspective of language complexity as the number of parts in a system. The two case studies are presented independently in sections 3.2 and 3.3.

3.1. Materials and Methods: Demographic Data

In order to investigate whether there are general patterns in how language structure adapts to social structure, we focus on demographic data. We correlate the linguistic phenomena under study with two sociolinguistic variables, the number of native speakers and the proportion of non-native speakers in the community. In this section we discuss the structure of these demographic data and their problems. The data and sources are provided in the **Supplementary Material**.

When defining the sociolinguistic features we largely follow Lupyan and Dale (2010) and Bentz and Winter (2013). We define the number of L1 speakers as the current number of speakers and the data is largely taken from the 19th edition of the *Ethnologue* (Lewis et al., 2016), which lists the number of speakers for all currently spoken languages in the database. To better scale the number of native speakers in both small and large languages, we take the base-10 logarithm of the number of L1 speakers (cf. Lupyan and Dale, 2010). The *Ethnologue* lists the number of speakers for a particular country and separately in all countries and in some cases also the size of the ethnic population. The latter may be helpful and indicative of the relative size of the population before the number of speakers began to drastically decline as, for instance, in North America (e.g., Nichols, 2009). Here we use the number of speakers in all countries. One problem with the number of speakers is that changes in the speech community

---

[2] These are: "Number of gender values" (Corbett, 2013a), "Sex-based and Non-Sex-Based Gender" (Corbett, 2013b), and "Systems of gender assignment" (Corbett, 2013c).

can sometimes be very quick, whereas changes in grammar are generally slower (cf. Sinnemäki, 2009). For this reason, it is unclear whether the current size of a speech community (or even the current size of the corresponding ethnic community) would reflect the situation at the time of writing the grammar or at the time in which the grammatical structures that are now captured in grammatical descriptions were developed. Numbers of native speakers should thus be conceived of as mere estimations, even when based on the most recent census.

The proportion of L2 speakers in the community is defined here as the proportion of non-native speakers in the whole speech community, where the size of the whole speech community includes both native and non-native speakers [that is, as $L2/(L1 + L2)$] (Bentz and Winter, 2013). This measure is meant to estimate the likelihood that the grammar is affected by the presence of a particular proportion of population speaking the language as an L2. Some researchers have used a cut-off point for the proportion of non-native speakers. For instance, Kusters (2003, p. 41) defined his type 2 communities as those in which more than half of the speech community were adult L2 learners. On the other hand, a reviewer suggested that maybe there is some cut-off point after which the population size is large enough to act as a buffer against effects from the L2 population. While this is an interesting suggestion, there is some evidence actually to the contrary. McWhorter (2007) argues that especially the languages of large empires tend to be susceptible to simplifying effects from a large L2 population. Wray and Grace (2007) even suppose that bigger languages have more contact with surrounding languages. This latter point is not supported by our data, which instead suggests that there is some tendency for large languages to have lower proportions of L2 speakers, as indicated by the negative correlation (albeit not consistently significant) between log number of L1 speakers and the proportion of L2 speakers below. We return to this briefly in section 3.2.2. Overall, in the spirit of Bentz and Winter (2013), we hypothesize that the proportion of L2 speakers is best seen as a continuum, since there are no clear, theoretically motivated cut-off points between the two endpoints.

A reviewer also suggested that perhaps the raw number of L2 speakers would be a better predictor than the proportion of L2 speakers. Since the number of L1 speakers is used in counting the proportion of L2 speakers, this might increase collinearity owing to the mathematical interconnectedness between the number of L1 speakers and the proportion of L2 speakers. We do not think that using the number of L1 speakers in counting the proportion of L2 speakers is a problem to us. Log number of L1 speakers did not correlate significantly with the proportion of L2 speakers when semi-speakers were excluded ($r = -0.147$; $df = 63$; $p = 0.24$), only when they were included ($r = -0.374$; $df = 71$; $p = 0.001$) and it is the former measure that is our primary estimate for the proportion of L2 speakers (more on semi-speakers below).

There are also some problems related to the availability and reliability of the data that need to be addressed. While the data for the number of speakers are readily available in the *Ethnologue*, data on L2 speakers are available only for a small proportion of languages in the *Ethnologue*. Alternative sources are sporadic and poorly representative of the world's languages. In our sample

this meant that we were able to obtain estimates for L2 data for roughly 70 languages.

The L2 data is problematic for two more reasons. One is that there is a range of speaker types that have been identified in the literature and not necessarily all sources use the same typology of speaker types. Grinevald (2003), for instance, divides speakers into 1. native speakers, 2. semi-speakers, 3. terminal speakers, and 4. rememberers. Native speakers are fluent, semi-speakers range from near-fluent to limited L2 speakers, terminal speakers are the last speakers of a dying language, and rememberers are speakers who have lost much of their earlier fluency in the language. In this classification most L2 speakers would be classified as semi-speakers. But it is not always clear what is counted as "L2 speaker." Sources that focus more on language acquisition or database-building make a difference between native speakers and L2 speakers, but they do not necessarily distinguish L2 speakers from semi-speakers. Yet sometimes this distinction is made, as is done in the *Ethnologue*, which distinguishes L2 speakers from semi-speakers. The latter is possibly reserved as a characteristic speaker-type in situations of language endangerment in which the last fluent speakers are the elders of the community who do not accept the younger generation's error prone talk (cf. Thomason, 2015). But this is not quite clear from the *Ethnologue*, since the figures for L2 speakers are defined for all non-native speakers irrespective of their level of competence in the target language. These issues lead to possible problems in the comparability of the numbers reported in the sources. For the purpose of this paper we assume that the problems are not too great.

The second problem with the L2 data concerns the often poor quality of the data. The compilers of the *Ethnologue* are well aware of this and report in "Ethnologue Global Dataset" that they originally "refrained from including these data due to" problems with adequacy of the data[3]. However, they finally published the data because the customer demand was very high. Although the data is continually updated, estimating the number of L2 speakers is very difficult and involves a considerable amount of guesswork. For instance, the number of L2 speakers for Bengali, the main language of Bangladesh, was estimated to be at 140 million speakers in the 17th edition of the *Ethnologue*, published in 2014. This many L2 speakers constitute 56% of the whole speech community of Bengali in Bangladesh (including native and non-native speakers). However, the latest 20th edition of the *Ethnologue* (published in 2017) reports that there are 19.2 million L2 speakers of Bengali in Bangladesh, which is not more than 9.7% of the Bengali-speaking population in Bangladesh. In a similar way, the number of L2 speakers of Russian was about 30 million in the 2010 census (cf. the 19th edition of the *Ethnologue*), but according to Arefyev (2012) (via the 20th edition of the *Ethnologue*) the number of L2 speakers of Russian is closer to 113 million. Our point is not to criticize the data in the *Ethnologue*, because of all available language databases that contain information on speech community size this is still the largest and most reliable source. Rather, we argue that any

---

[3]The Ethnologue Global Dataset is available at https://www.ethnologue.com/data-consulting.

database that aims to collect information on this type of figures would run into the same problems. When facing such degree of uncertainty with the data, one possibility is to average the reported figures (e.g., Bentz and Winter, 2013). We decided not to use averages but to take the data from the sources that were most recent or that we evaluated as the most reliable.

In order to explore whether the number of semi-speakers, as usually reported for small endangered languages, had an effect on the results, we conducted the statistical models by including the number of semi-speakers in the L2 data, but also report results about the models in which the L2 figures did not include the number of semi-speakers. The fact that semi-speakers have low competence of the target language may suggest that they may use simplified language with transfer effects from the native language. This kind of pidginization has been hypothesized to influence language structures in the target language. However, a high number of semi-speakers may not necessarily be indicative of the kind of sociolinguistic situation that has been hypothesized as having an influence on language structure. For instance, the situation of many North American Indian communities is such that the elders speak the language which the younger generation learns only as a L2. The elders may not accept the language of the younger generation, who may in turn feel inferior because of their bad knowledge of the language. This suggests that, in these and similar contexts, it is unlikely that the language use of the semi-speakers would simplify the language of the whole community.

3.2. Study 1: Morphological Complexity of the Verb

3.2.1. Materials and Methods

The data for inflectional synthesis come from the *AUTOTYP* database, thus we follow its definition of the phenomenon. The database contains information on the degree of inflectional synthesis of verbs but not of other parts of speech. Here we provide succinct description of the definitions but guide the reader to Bickel and Nichols (2007, 2013) for further details (see also section 2.1). The material for inflectional synthesis is provided in the **Supplementary Material**.

According to Bickel and Nichols (2013) the degree of synthesis measures the number of morphological categories expressed per word in a maximally inflected verb form. The notion of maximally inflected word form refers to the fact that verbs can vary in terms of their synthesis within a language: the English past tense is marked with an affix *-ed* and the future tense with a separate word *will* so the past tense is more synthetic than the future tense in English. The data set codes the most synthetic verb forms in each sample language and registers the maximal number of categories per verb. For English this approach counts two categories, namely agreement (third person in present tense) and tense (past tense *-ed*). The counted categories do not have to coincide in the same verb form in language use, and often they do not.

Our hypothesis is that an inverse relationship exists between degree of inflectional synthesis on the verb and demographic factors. To assess this relationship we constructed generalized

TABLE 2 | Model names and predictors in case study 1.

Model name	Predictor(s)
SYNTHESIS.L1	log number of L1 speakers
SYNTHESIS.L2	proportion of L2 speakers (excluding semi-speakers)
SYNTHESIS.L2+	proportion of L2 speakers (including semi-speakers)
SYNTHESIS.ALL	log number of L1 speakers and proportion of L2 speakers (including semi-speakers)

linear mixed effects models (GLMMs) using the package `glmmADMB` (Fournier et al., 2012; Skaug et al., 2016) in R (R Core Team, 2017). Mixed models have been recently applied and discussed in language typology by Sinnemäki (unpublished) and Jaeger et al. (2011). We used `glmmADMB` instead of the more popular `lme4` package because the maximal models (see below) converged better with the former and because `glmmADMB` also offers ways of dealing with zero-inflated variables (see section 3.3.1)[4]. In addition, in models involving the number of L1 speakers the L1 population sizes were set to 50 when the actual number of L1 speakers was 50 or less. In doing so we follow Lupyan and Dale (2010). They do not explain why they manipulated the number of speakers in this way but the reason might be that for such small speech communities the numbers of speakers may be very unreliable.

We constructed four models for this case study. The model designs are similar except for the predictors; the model names and their predictors are listed in **Table 2**. In all of the models the degree of inflectional synthesis was the response and the random structure was the same: *AUTOTYP* stocks were used as a grouping factor for genealogical affiliation and the 24 areas of *AUTOTYP* as the grouping factor for geographical areas. Stocks are the highest level in the genealogical taxonomy of *AUTOTYP*, roughly corresponding to language families in *WALS*. We prefer the *AUTOTYP* stocks to the *WALS* families because they are generally more conservative and do not posit problematic higher level families such as Altaic. The 24 areas of *AUTOTYP* consist of areas such as California, Europe, and Southeast Asia as well as 21 additional areas that are roughly parallel in size. **Figure 1** illustrates these areas on a world map.

Our models are maximal in that they include all the theoretically motivated random intercepts and slopes. In the light of recent debates, maximal models are preferred in mixed models since especially models without random slopes are susceptible to produce spurious results (Schielzeth and Forstmeier, 2009; Barr et al., 2013). However, models containing random slopes may lead to overfitting and the random effect variances being zero or approaching zero. To improve our models we tested whether some of the random slopes could be removed. For mixed models *p*-values can be derived by using maximum likelihood ratio tests, which can be applied for both fixed and random effects. To evaluate the *p*-values of effects we compared the likelihood ratio

---

[4]In `glmmADMB` parameters are estimated by maximum likelihood ratio using Laplace approximation. We improved this Laplace approximation by using importance sampling, providing the argument `impSamp` with values >0; (Skaug and Fournier, 2006).
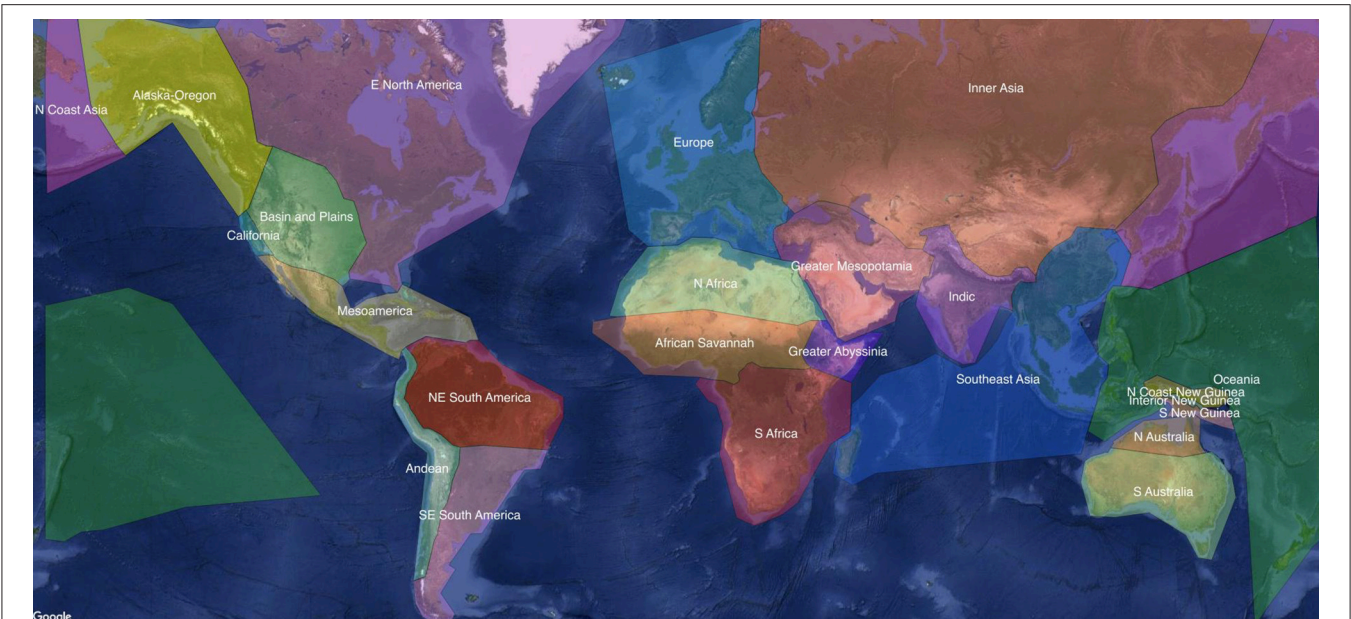
FIGURE 1 | The 24 areas of the *AUTOTYP* on a world map (Bickel et al., 2017; used under CC-BY 4.0 license).

TABLE 3 | Dispersion ratio and deviance from 1 for models in case study 1.

Model name	Dispersion ratio	Estimation of deviance
SYNTHESIS.L1	0.84	$\chi^2_{(311)} = 261.5; p = 0.98$
SYNTHESIS.L2	1.01	$\chi^2_{(57)} = 57.5; p = 0.46$
SYNTHESIS.L2+	0.85	$\chi^2_{(65)} = 55.4; p = 0.79$
SYNTHESIS.ALL	0.96	$\chi^2_{(64)} = 61.3; p = 0.57$

of a model with the variable of interest to that of a simpler model without the variable of interest (e.g., Baayen et al. 2008; Barr et al. 2013).

The degree of inflectional synthesis is discrete count data, ranging from 0 to 14, and therefore we used Poisson regression to model the data. Poisson distribution assumes that the sample mean is identical with the sample variance. The dispersion ratios in all the models were not significantly different from 1 (see **Table 3**), which means that the assumption of Poisson regression about identical sample mean and variance was met.

3.2.2. Results

The sample contains data on log number of native speakers and the degree of verbal inflectional synthesis in 309 languages. It was possible to get data on the proportion of L2 speakers in 65 languages and for an additional 8 languages on the number of semi-speakers. The histogram distribution of degree of inflectional synthesis is provided in **Figure 2**. The degree of inflectional synthesis is roughly normally distributed around a mean of six inflectional categories per verb. The areal distribution of the sample languages and their degree of inflectional synthesis is provided in **Figure 3**.
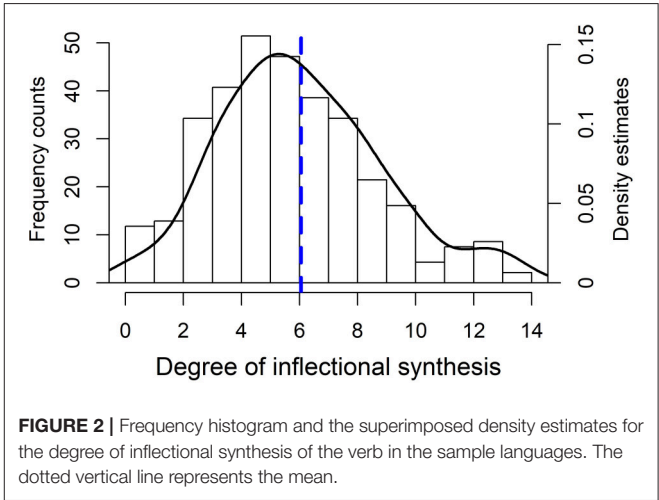
FIGURE 2 | Frequency histogram and the superimposed density estimates for the degree of inflectional synthesis of the verb in the sample languages. The dotted vertical line represents the mean.

The distribution of the demographic factors is shown in **Figure 4**. In the sample the median size of L1 populations was 14,100, which is much larger than the total median (7,000) for all spoken languages in the *Ethnologue*. This difference is possibly due to the fact that larger languages tend to be also better described than smaller languages. The median proportion of L2 speakers was 18% and that of semi-speakers 58%. The reason why the proportion of semi-speakers tends to be higher than that of L2 speakers is that the data for semi-speakers comes from small languages in North America with the kind of sociolinguistic situation we described in section 3.1.

According to the mixed logistic regression of the maximal model of SYNTHESIS.L1, log number of L1 speakers had a significant negative effect on the degree of inflectional synthesis
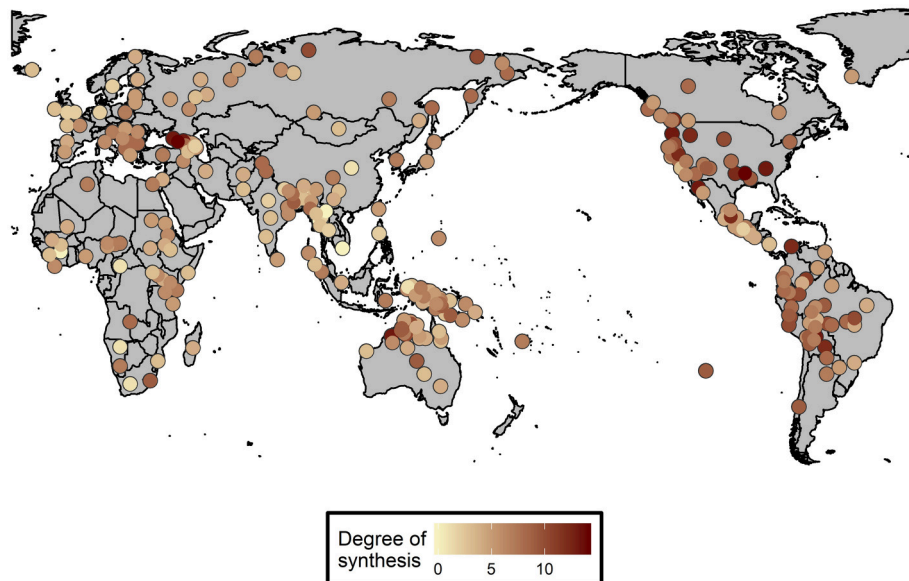
FIGURE 3 | Degree of inflectional synthesis and its areal distribution on a world map.
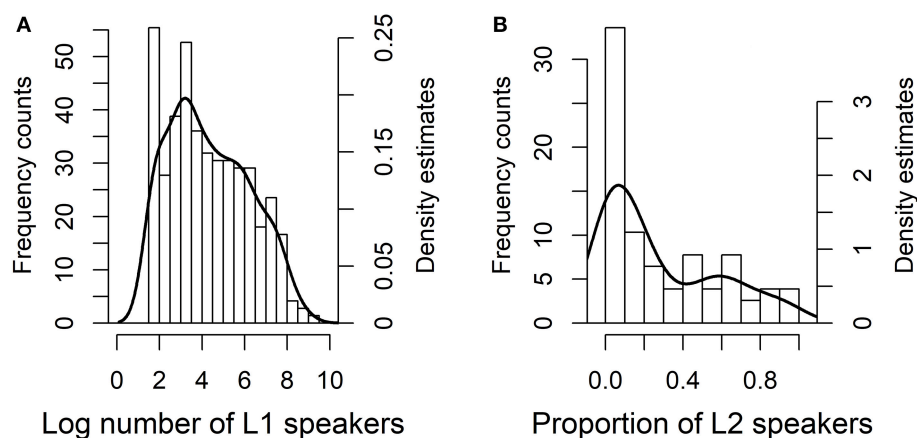
FIGURE 4 | Frequency histograms and superimposed density estimates for the independent demographic variables in case study 1, for log number of native speakers on the left **(A)** and for the proportion of L2 speakers (including semi-speakers) on the right **(B)**.

[$\log(\lambda) = -0.077 \pm 0.018$; $\chi^2(1) = 17.5$; $p = 0.000028$]. However, while this maximal model converged the random effect variances for the slopes (both Stocks and Area) were very close to zero (see **Figure 5**), which suggests that the random slopes may be superfluous. The maximum likelihood ratio tests confirm that both slopes may be removed from the model [random slope over Stocks: $\chi^2(1) = 0.33$; $p = 0.57$; random slope over Area: $\chi^2(1) = 0.89$; $p = 0.35$], which leaves us with a random intercept model.

According to the reduced model log number of L1 speakers had a significant negative effect on the degree of inflectional synthesis [$\log(\lambda) = -0.079 \pm 0.018$; $\chi^2(1) = 17.9$; $p = 0.000023$]. The negative coefficient and the significant p-value suggest that the hypothesis is confirmed. But because the estimate

is rather small, the size of the speech community seems to have only a small impact on the degree of inflectional synthesis. Because in Poisson regression it is the log of the expected counts that is modeled, the coefficients can be transformed via inverse logarithm to better understand them. The coefficient for log of L1 speakers was -0.077 and its inverse logarithm is 0.926. This means that as the population size becomes 10 times larger (we used log10) the language will have on average 7.4% fewer inflectional categories per verb conditioned by the random effect structure.

To further assess the models' goodness of fit, we used Akaike Information Content (AIC) or its small sample equivalent AICc which is corrected for bias (Burnham and Anderson, 2002). AIC can be used to evaluate the importance of a predictor by
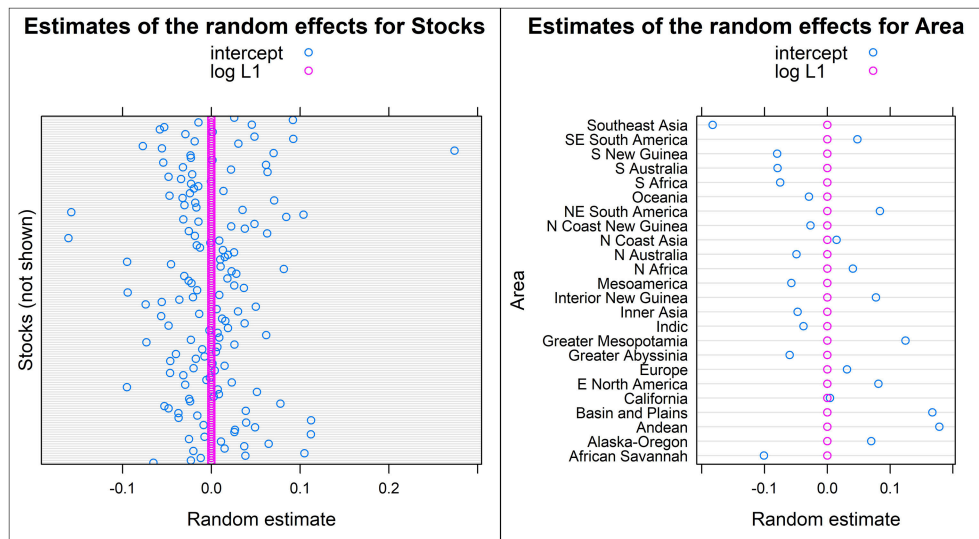
FIGURE 5 | Random effect variation of model SYNTHESIS.L1. The left panel shows the estimates for the random intercept and random slope over Stocks and the panel on the right shows the estimates for the random intercept and random slope over Area.

considering to what extent adding the fixed effect reduces AIC. Lower values of AIC improve the model's fit and, therefore, the larger the reduction in AIC is, the more important the predictor (e.g., Baayen, 2013). As a rough guideline, if the difference in AIC between the models is <2, the models fit the data roughly equally well, that is, there is no significant difference between the models; if the difference is between 4 and 7 there is much less support for the model with the higher AIC value, that is, the AIC difference can be considered important; if the difference is 10 or greater, there is basically no support for the model with the higher AIC value (Burnham and Anderson, 2002, p. 70–71). We compared the AIC values in the model which contained the log number of L1 speakers ($AIC = 1459.9$) to a model that contained only the random intercepts ($AIC = 1475.8$). Adding the fixed effect reduced the AIC by 15.9. Since the difference is >10, there is substantial support for the model that included the log number of L1 speakers. In other words, although the effect of the log number of L1 speakers on the degree of synthesis was small, it was nevertheless reasonable, as it clearly improved model fit[5].
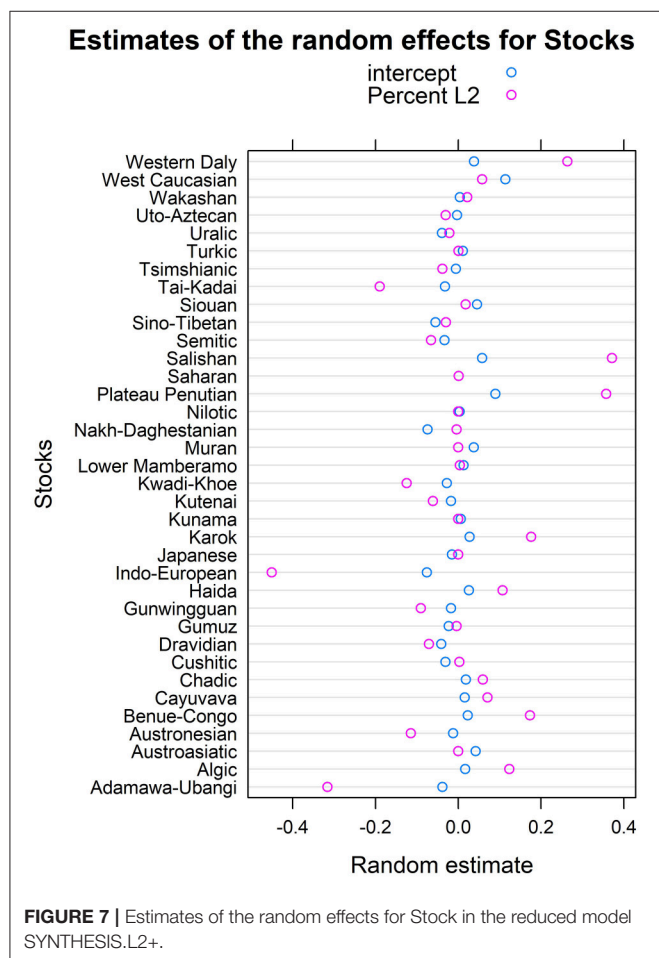
According to the mixed logistic regression of model SYNTHESIS.L2, the proportion of L2 speakers had an inverse effect on the degree of inflectional synthesis but this effect was not significant ($\log(\lambda) = -0.39 \pm 0.23$; $\chi^2(1) = 3.04$; $p = 0.081$). Again, while the maximal model converged the random effect variances for both Stocks and Area were very close to zero (of the magnitude of 1e-7), which suggests that some of the random structure may be superfluous. The maximum likelihood ratio tests confirm that both slopes may be removed from the model (random slope over Stocks: $\chi^2(1) = 0.24$; $p = 0.63$; random slope over Area: $\chi^2(1) = 0.002$; $p = 0.96$).

According to the reduced model the proportion of L2 speakers had an inverse and borderline significant effect on the degree of inflectional synthesis ($\log(\lambda) = -0.398 \pm 0.21$; $\chi^2(1) = 3.798$; $p = 0.051$). However, the borderline significant p-value makes the result somewhat uncertain. We further compared the

---

[5]We initially evaluated the predictive capacity of our models by using the index of concordance (C) between the predicted probability and the observed response, which is quite widely used in linguistics. For the reduced model the value of C was 0.74 (values of C between 0.7 and 0.8 are considered acceptable by Hosmer and Lemeshow, 2000, p. 162). However, a reviewer pointed out that this good index of concordance might be misleading, as its value might be only due to the random structure. To double-check this, we compared the C in models with and without the fixed effect. Since the difference of these models was only 0.006, this seemed to suggest that the log number of L1 speakers has virtually no predictive power. For two reasons, we think that this conclusion would be premature. First, Barth and Kapatsinski (2018) present results of a simulation, which shows that a real predictor may fail to contribute to a model's predictive capacity measured, for instance, by C. Their result suggests that perhaps the index of concordance should not be used to assess the predictive capacity of GLMMs. Second, the predictive capacity of models can also be evaluated using R^2, which measures the variance explained by the model. Although there is no consensus as to how or whether it

would be possible to reliably compute R^2 for GLMMs, many researchers currently use marginal R^2 to compute the variance explained by the fixed effects only and conditional R^2 to compute the variance explained by the whole model (both fixed and random effects) (following Nakagawa and Schielzeth, 2013; Johnson and O'Hara, 2014). Marginal and conditional R^2 can be computed using the R package MuMIn (Barton, 2018), but unfortunately this is not yet implemented for models produced with glmmADMB, which we used for modeling. For this purpose, we used Bayesian mixed effects modeling with R package blme (Chung et al., 2013) to build model SYNTHESIS.L1 (our reduced model that included only the random intercepts), as it produces objects that MuMIn understands but also because the models actually converged, unlike when using the package lme. The results produced by blme[$\log(\lambda) = -0.079 \pm 0.018$; $\chi^2(1) = 17.9$; $p = 0.000023$] were practically identical compared to those produced by glmmADMB [$\log(\lambda) = -0.078 \pm 0.018$; $\chi^2(1) = 17.8$; $p = 0.000024$]. Based on the model produced by blme the marginal $R^2 = 0.094$ and the conditional $R^2 = 0.279$. The marginal R^2 suggests that the log number of L1 speakers has reasonable predictive power, as it explains almost 10% of variance in the degree of synthesis. In addition, the conditional R^2 is similar to what we have often witnessed for typological data (e.g., Sinnemäki, 2010).

FIGURE 6 | Degree of inflectional synthesis as a function of log number of speakers on the left **(A)** and for the proportion of L2 speakers on the right **(B)**.
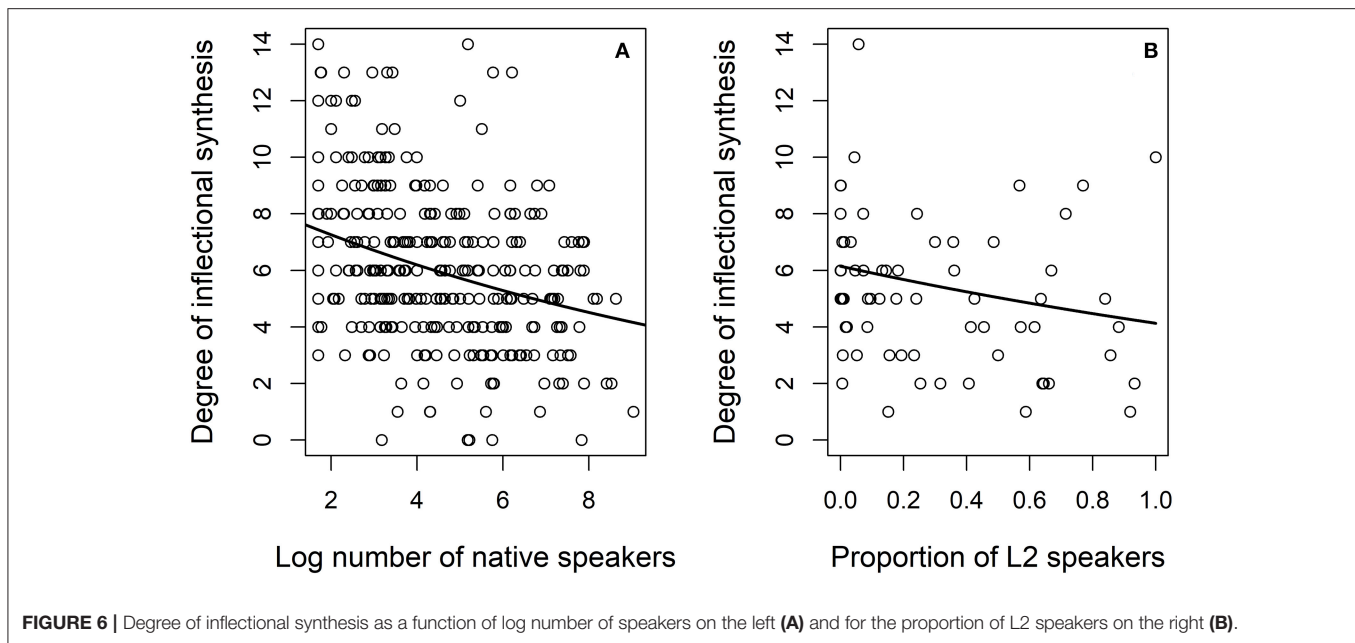
AIC values in the model which contained the proportion of L2 speakers (304.2) to a model that contained only the random intercepts (306.0). Adding the fixed effect decreased the AIC only by 1.8, which provides further evidence that the proportion of L2 speakers has a negligible effect on the degree of inflectional synthesis.

Figure 6 presents the degree of inflectional synthesis as a function of the demographic variables in models SYNTHESIS.L1 and SYNTHESIS.L2. The curve indicates the fit of the mixed regression model. The figure on the left (**Figure 6A**) presents the fit to log number of L1 speakers. In communities with about 1,000 speakers or less [$log_{(1,000)} = 3$] the predicted degree of synthesis is about 7 while it drops to about 5 in communities with a million or more L1 speakers [$log_{(1,000,000)} = 6$]. The downward slope is clear but not impressively large. The figure on the right (**Figure 6B**) presents the fit to the proportion of L2 speakers. There is a small downward trend so that in communities with few L2 speakers the predicted degree of synthesis is around 6, whereas in communities with close to 100% L2 speakers the predicted degree is about 4.

According to the mixed logistic regression of model SYNTHESIS.L2+, the proportion of L2 speakers (including semi-speakers) had an inverse effect on the degree of inflectional synthesis but this effect was not significant [$\log(\lambda) = -0.27 \pm 0.24$; $\chi^2(1) = 1.32$; $p = 0.25$]. We again tested the random effect structure with maximum likelihood ratio tests and removed the random slope for Area but not that for Stocks [random slope over Stocks: $\chi^2(1) = 3.96$; $p = 0.047$; random slope over Area: $\chi^2(1) = 1.83$; $p = 0.18$]. According to the reduced model the proportion of L2 speakers (including semi-speakers) had an inverse but non-significant effect on the degree of inflectional synthesis [$\log(\lambda) = -0.23 \pm 0.23$; $\chi^2(1) = 1.07$; $p = 0.30$]. The negative coefficient provides support for the hypothesis but the non-significant p-value goes against the hypothesis. According to

this model, the effect of L2 proportion on inflectional synthesis is largely lineage-specific. This is suggested by the significant random slope for Stock and by the large positive (e.g., in Salishan) and negative (e.g., in Indo-European) random variances for Stock (see **Figure 7**).

All in all when the effect of the demographic variables was researched in isolation only the number of L1 speakers had a clearly significant and negative effect on the degree of inflectional synthesis. The significant effect of the number of L1 speakers replicates the result by Lupyan and Dale (2010). However, compared to the proportion of L2 speakers the number of L1 speakers is a less direct measure of the kind of language contact effects that have been hypothesized to influence language structures (see section 2.3). For this reason it is somewhat surprising that it was the less direct measure of language contact effects that had a significant effect on language structures in the modeling. It is possible that this is mostly due to sample size. In the model SYNTHESIS.L1 the sample size was 309 languages but in the model SYNTHESIS.L2 the sample size was 65 languages. In order to test whether this result depended on sample size, we modeled the effect of the two demographic variables in the same model.

In model SYNTHESIS.ALL we model the effects of the log number of L1 speakers and the proportion of L2 speakers (excluding semi-speakers) in competition with one another. According to the mixed logistic regression of the maximal model, log number of L1 speakers had a significant inverse effect on the degree of inflectional synthesis [$\log(\lambda) = -0.12 \pm 0.026$; $\chi^2(1) = 15.2$; $p = 0.000095$]. The proportion of L2 speakers (excluding semi-speakers) had also an inverse effect and this time also a significant effect on the degree of inflectional synthesis [$\log(\lambda) = -0.47 \pm 0.20$; $\chi^2(1) = 5.8$; $p = 0.016$]. We again tested the random effect structure with maximum likelihood ratio tests because most of the random effect variances

for the slopes (both Stocks and Area) were very close to zero (of the magnitude of 1e-7) and ended up removing all the random slopes (all were non-significant).

Estimates of the random effects for Stocks

FIGURE 7 | Estimates of the random effects for Stock in the reduced model SYNTHESIS.L2+.

According to this reduced model the log number of L1 speakers had a significant inverse effect on the degree of inflectional synthesis [$\log(\lambda) = -0.10 \pm 0.026$; $\chi^2(1) = 12.3$; $p = 0.00046$] and so did the proportion of L2 speakers [$\log(\lambda) = -0.47 \pm 0.19$; $\chi^2(1) = 6.58$; $p = 0.010$]. For the purpose of model comparison, we modeled the log number of L1 speakers in isolation from the proportion of L2 speakers but just for this smaller data set ($n = 65$), keeping the random structure identical (that is, modeling just the random intercepts). In this model the log number of L1 speakers again had a significant but slightly smaller inverse effect on the degree of inflectional synthesis [$\log(\lambda) = -0.09 \pm 0.025$; $\chi^2(1) = 9.5$; $p = 0.0021$] than when modeling the log number of L1 speakers in the same model with the proportion of L2 speakers. The coefficient for log of L1 speakers was -0.10 and its inverse logarithm is 0.905. This means that (in this smaller sample) as the population size becomes 10 times larger the language will have on average 9.5% fewer inflectional categories per verb conditioned by the random effect structure. The coefficient for the proportion of L2 speakers was -0.47 and its inverse logarithm is 0.625. This means that languages spoken by communities with 100% L2 speakers have about 37.5% fewer inflectional categories per verb than those with no L2 speakers conditioned by the random effect structure.

Figure 8 presents the effect plots for the model predictors in model SYNTHESIS.ALL[6]. The plots present the predictors' values on the x-axis and the predicted values of the response on the y-axis. Based on the effect plot for log L1 speakers as the predictor, the predicted degree of synthesis drops from roughly eight categories in communities with about 10 speakers [$log_{(100)} = 2$] to about four in communities with 100 million or more L1 speakers [$log_{(100,000,000)} = 8$]. The downward slope is very clear. Based on the effect plot for the proportion for L2 speakers, the predicted degree of synthesis drops from roughly

---

[6]Created using R package `effects` (Fox, 2003).

FIGURE 8 | Effect plots of the model predictors in model SYNTHESIS.ALL (log number of L1 speakers on the left and the proportion of L2 speakers on the right).
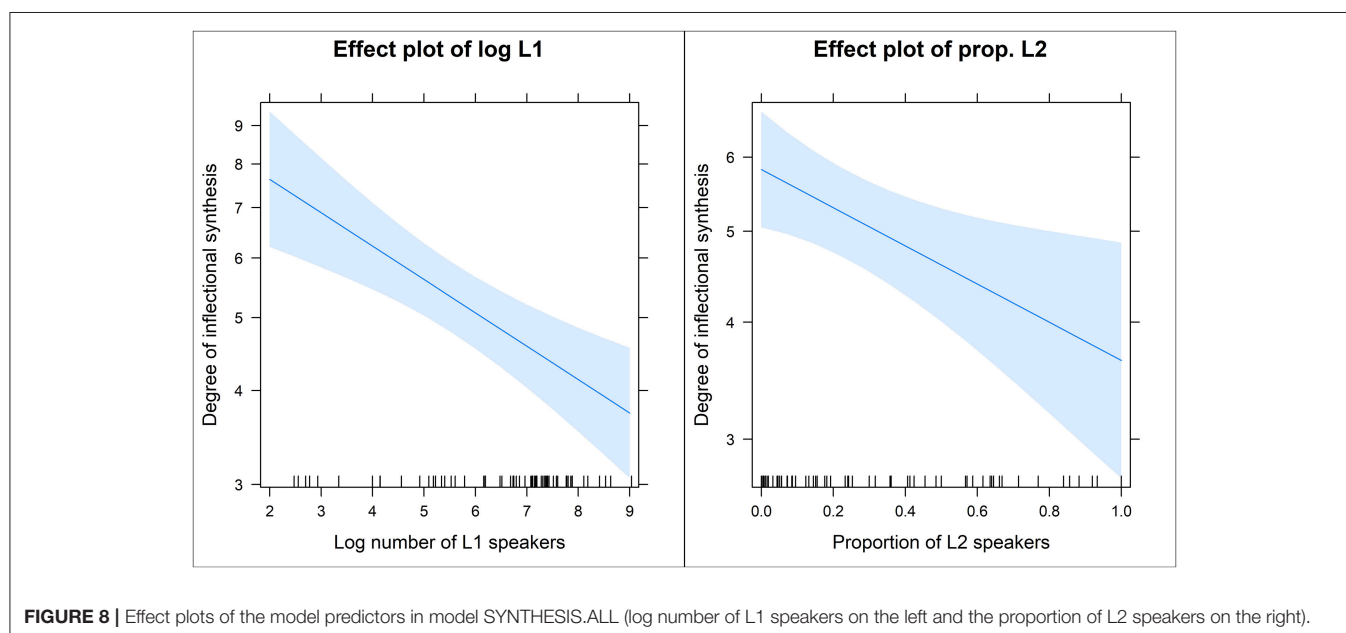
TABLE 4 | Results of model comparison for the reduced model SYNTHESIS.ALL. The full model includes both the log number of L1 speakers and the proportion of L2 speakers.

Model	Model structure	Df	AICc	Reduction in AICc	Akaike weights
(1)	Only random structure	3	306.4	0.0	0.003
(2)	Proportion of L2	4	304.9	1.5	0.006
(3)	Log number of L1 speakers	4	299.2	7.2	0.107
(4)	Full model	5	295.0	4.2	0.884

six categories in communities with no L2 speakers to about four in communities with about 80% or more L2 speakers. There is a downward slope but not as steep as for the log number of L1 speakers.

For model comparison we used AICc; the results are reported in **Table 4** in decreasing order of AICc. Based on the AICc values the model (1) which contained only the random intercepts but no fixed effects had the largest AICc value (306.4) and, therefore, it is the worst of the four models. In model (2) the proportion of L2 speakers was added as a fixed effect to the random intercepts-model and this decreased the AICc by 1.5 compared to model (1). This decrease is small and suggests that modeling the proportion of L2 speakers in isolation from the log number of native speakers produces a negligible effect. In model (3) the log number of L1 speakers was added to the random intercepts-model and this decreased the AICc by 7.2 compared to model (1). This large reduction suggests that the log number of L1 speakers has a reasonable effect on the degree of inflectional synthesis. In model (4) the proportion of L2 speakers was added as a fixed effect to model (3), which gives us the full model that contained the random intercepts and both of our fixed effects. In the full model the AICc value was the smallest, being 4.2 smaller than in model (3). We further used Akaike weights (the right-most column in **Table 4**) to compare these four models to one another (Burnham and Anderson, 2002). The Akaike weights scale the differences in the models' AIC values to a scale of 1 and thus provide an easy and effective way to interpret the models' AIC differences[7]. Based on the Akaike weights, the model (4) which includes both the log number of L1 speakers and the proportion of L2 speakers has 88.4% chance of being the best model among our four models. These results suggests that modeling both demographic factors in the same model significantly improves the model fit compared to modeling them in isolation from one another[8].

---

[7]The AICc values as well as the Akaike weights were computed using package MuMIn (Barton, 2018).
[8]Although our data is relatively small we also tested whether the interaction term between the log number of L1 speakers and the proportion of L2 speakers would have a significant effect on the degree of inflectional synthesis. It is possible that in very large languages the population size of L1 would act as a buffer against transfer effects from the L2 population (cf. our discussion in section 3.1). For this purpose we compared a model that included this interaction term to one that excluded it (using only random intercepts for both Stock and Area). Based on the result the interaction term had a negative but non-significant effect on the degree of

Thus, to summarize, the log number of L1 speakers has a significant effect on the degree of inflectional synthesis both in the larger sample (SYNTHESIS.L1; $n = 309$) and in the smaller sample (SYNTHESIS.ALL; $n = 65$). Conversely, the proportion of L2 speakers has a clearly significant effect on the degree of inflectional synthesis only when modeling it in competition with the number of L1 speakers ($p = 0.010$) but not when modeling it in isolation ($p = 0.051$). These results are confirmed by comparing the AIC values.

3.2.3. Discussion

Two of the four statistical tests that we carried out to investigate the effect of population data on the degree of inflectional synthesis yielded significant results. Altogether these findings replicate and expand on previous research (Lupyan and Dale, 2010; Bentz and Winter, 2013) and suggest that the hypothesis whereby verbal inflectional synthesis adapts to demographic variables is corroborated by the present data set.

Our first model (SYNTHESIS.L1) replicated the earlier findings by Lupyan and Dale (2010). However, our results were based on a data set (309 languages) that was more than two times larger than the data set (145 languages) in Lupyan and Dale (2010). We also used the original exact counts for the degree of inflectional synthesis from *AUTOTYP* rather than the conflated count categories from the *WALS*.

We then estimated the proportion of L2 speakers in the whole speech community in the spirit of Bentz and Winter (2013). In their study the proportion of L2 speakers had a significant inverse effect on the number of case distinctions but, importantly, the size of the speech community did not. The fact that in our models the proportion of L2 speakers (whether including or excluding semi-speakers) did not have a clearly significant effect on the degree of verbal inflectional synthesis suggests that the proportion of L2 speakers alone is not a sufficient predictor of adaptive effects for all kinds of different linguistic structures, although it may be sufficient for some, such as number of cases. This result is in line with the hypotheses of Trudgill (2011b), who argues that single sociolinguistic features may not be sufficient for showing correlations between language structure and sociolinguistic structure and that richer models of the sociolinguistic environment are necessary instead.

We also contrasted two measures for the proportion of L2 speakers, namely, one including semi-speakers and the other excluding them. While in the latter model (SYNTHESIS.L2) the proportion of L2 speakers was borderline significant, in the former (SYNTHESIS.L2+) it was not. In addition, in the former model the slope for Stocks was significant. This result may be related to the observation in section 3.2.2 that the median proportion of L2 speakers was much smaller than the median proportion of semi-speakers, all of which came from small languages of North America. In other words, the large median proportion of semi-speakers suggests a different sociolinguistic environment, and thus different conditioning factors for those

---

inflectional synthesis [$\log(\lambda) = -0.085 \pm 0.082$; $\chi^2(1) = 1.1$; $p = 0.30$]. This suggests that population size and the proportion of L2 speakers influence degree of inflectional synthesis independently of one another.

languages for which the number of semi-speakers was reported compared to those for which the number of L2 speakers was reported. For future research it may thus be necessary to treat L2 speakers separately from semi-speakers, to the extent that this is analytically possible.

Lastly in our model SYNTHESIS.ALL we included both the log number of L1 speakers and the proportion of L2 speakers (excluding semi-speakers) in the same model, which produced a set of interesting results. First, the number of L1 speakers had a significant effect even with the smaller sample (compared to model SYNTHESIS.L1). This result suggests that the number of L1 speakers is an important predictor of the degree of verbal inflectional synthesis and that the result in model SYNTHESIS.L1 was not just a consequence of larger sample size. Most interestingly, both our sociolinguistic factors had a significant inverse effect on the degree of inflectional synthesis when modeled as fixed effects in the same model and this model was also the best among competing models when using Akaike weights. In contrast, the proportion of L2 speakers did not have a clearly significant effect on the degree of inflectional synthesis when modeled in isolation (model SYNTHESIS.L2 and model SYNTHESIS.ALL). These results are in line with Trudgill (2011a,b)'s predictions. According to Trudgill, the sociolinguistic environment that attracts adaptation in the complexity of language structures cannot be systematically characterized by single sociolinguistic features, such as population size, but demands richer data. He further suggests that three sociolinguistic factors are decisive, namely, population size (here roughly the number of L1 speakers), degree of language contact (that we approximate by measuring the proportion of L2 speakers in the speech community), and the density of social networks. While our models did not include a factor for density of social networks, they still provided improved results compared to modeling the sociolinguistic factors in isolation. For future research our results suggest that the kind of sociolinguistic environment that may attract changes in the complexity of language structures cannot be easily captured by single demographic factors, but should preferably include information about population size, degree of contact vs. isolation, and possibly also other factors.

3.3. Study 2: Morphological Complexity and Grammatical Gender

3.3.1. Materials and Methods

We collected data on the number of genders in 345 languages. The material is provided in the **Supplementary Material**. The data is largely based on Sinnemäki (unpublish) and Corbett (2013a) and therefore we follow the definitions in these two studies.

As outlined in section 2.2, we define gender as a grammatical strategy that groups nouns into classes. These classificatory distinctions are not necessarily marked on nouns, but must be marked on clausal constituents that are in a syntactic relationship (also known as agreement) with nouns.

TABLE 5 | Model names and predictors in case study 2.

Model name	Predictor(s)
GENDER.L1	log number of L1 speakers
GENDER.L2	proportion of L2 speakers (excluding semi-speakers)
GENDER.L2+	proportion of L2 speakers (including semi-speakers)
GENDER.ALL	log number of L1 speakers and proportion of L2 speakers (including semi-speakers)

The number of genders in a language was counted based on number of distinguishable agreement classes. Usually a gender class is marked consistently across inflectional paradigms. However, often not all distinctions are present in all paradigms, as is the case in Mufian (**Table 1**). For instance, verb prefixes in Mufian are identical in classes 1, 2, and 3 in the plural, but in the singular the classes are distinguished from one another. For this reason each of these classes was counted as a separate gender in Mufian; all sample languages were analyzed with the same principles.

Our hypothesis is that an inverse relationship exists between the number of genders and the demographic factors used as independent variables. Similarly to case study 1, we constructed generalized linear mixed effects models using the package glmmADMB (Fournier et al., 2012; Skaug et al., 2016) in R (R Core Team, 2017) to assess the relationship between the number of genders and the demographic factors. The Poisson regression modeling is complicated by the large number of zeroes. The sample contains 345 languages but 200 (58%) of them have no genders. We accounted for this high number of zeroes by using zero inflation models offerred by glmmADMB. As in study 1, in this case study, too, we set the L1 population sizes to 50 when the actual number of L1 speakers was 50 or less (and for the same reasons; see section 3.2.1).

We constructed four models in this case study following the same principles as in case study 1 (see section 3.2.1). The model names and their predictors are listed in **Table 5**. In all of the models the number of genders was the response and the random structure was the same: *AUTOTYP* stocks were used as a grouping factor for genealogical affiliation and the 24 areas of *AUTOTYP* as the grouping factor for areas. However, models containing random slopes may lead to overfitting and the random effect variances being zero or approaching zero. To improve our models we tested whether some of the random slopes could be removed.

The number of genders is discrete count data, ranging from 0 to 17, and therefore we used Poisson regression to model the data. Poisson distribution assumes that the sample mean is identical with the sample variance. However, the dispersion ratios met the assumption about identical sample mean and variance (that is, the dispersion ratios were not significantly different from 1) only in model GENDER.L1. In models GENDER.L2, GENDER.L2+, and GENDER.ALL the dispersion ratio was significantly different from 1 which means that the assumption about identical sample mean and variance was not met for these models (see **Table 6**). Our solution was to use negative

TABLE 6 | Dispersion ratio and deviance from 1 for models in case study 2.

Model name	Dispersion ratio	Estimation of deviance
GENDER.L1	0.90	$\chi^2_{(337)} = 301.6; p = 0.92$
GENDER.L2	1.32	$\chi^2_{(57)} = 75.5; p = 0.051$
GENDER.L2+	1.35	$\chi^2_{(64)} = 86.3; p = 0.033$
GENDER.ALL	1.78	$\chi^2_{(57)} = 101.2; p = 0.00028$
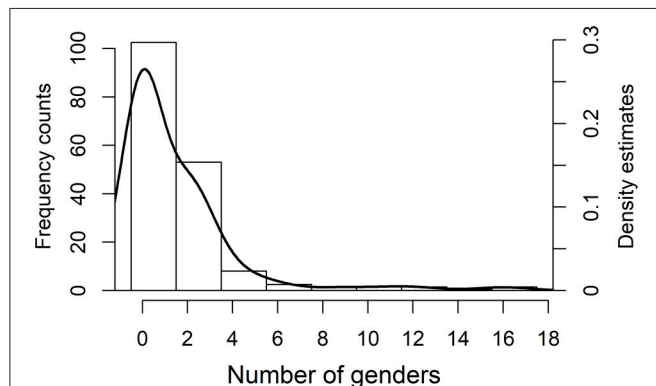
FIGURE 9 | Frequency histogram and the superimposed density estimates for the number of genders in the sample languages (because no language has exactly one gender we smoothed over this absence by using biased cross-validation for bandwidth in density estimation).

binomial models for these three models and Poisson regression for GENDER.L1.

3.3.2. Results

The sample contains data on log number of native speakers and the number of genders in 345 languages. It was possible to get data on the proportion of L2 speakers in 65 languages and for an additional 7 languages on the number of semi-speakers. The distribution of the number of genders is shown in **Figure 9**. The number of genders has a roughly negative exponential distribution, that is, it is strongly skewed to the right. This kind of distribution is typical for typological variables (Cysouw, 2010). The areal distribution of number of gender is provided in **Figure 10** on a world map.

The distribution of the demographic factors for the sample languages is shown in **Figure 11**. In this sample the median size of L1 populations was 10,000, which is somewhat smaller than in case study 1 but still larger than the total median of 7,000 for all spoken languages in the *Ethnologue*. The median proportion of L2 speakers was 19% and that of semi-speakers 58%. These figures are practically identical to those in case study 1 because roughly the same data was used.

According to the zero-inflated mixed logistic regression of the maximal model of GENDER.L1, log number of L1 speakers had a non-significant (positive) effect on the number of genders $[\log(\lambda) = 0.015 \pm 0.069; \chi^2(1) = 0.048; p = 0.83]$. However, while this maximal model converged the random effect variances for the slopes (both Stocks and Area) were very close

to zero (of the magnitude of 1e-7). The maximum likelihood ratio tests confirm that both slopes may be removed from the model [random slope over Stocks: $\chi^2(1) = 0.14; p = 0.71$; random slope over Area: $\chi^2(1) = 1.49; p = 0.22$]. According to the reduced model, the effect of log number of L1 speakers on the number of genders was non-significant $[\log(\lambda) = 0.024 \pm 0.059; \chi^2(1) = 0.17; p = 0.68]$. The non-significant p-value provides evidence that log number of native speakers has no effect on the number of genders.

According to the zero-inflated negative binomial mixed logistic regression of model GENDER.L2, the effect of the proportion of L2 speakers on the number of genders was negative but non-significant $[\log(\lambda) = -0.69 \pm 0.49; \chi^2(1) = 1.1; p = 0.30]$. Again, while the maximal model converged the random effect variances for both Stocks and Area were very close to zero (of the magnitude of 1e-7) and as a result the random slopes for both Area and Stocks were removed [random slope over Stocks: $\chi^2(1) = 1.1; p = 0.29$; random slope over Area: $\chi^2(1) = 0.006; p = 0.94$]. According to the reduced model the proportion of L2 speakers had an inverse and non-significant effect on the number of genders $[\log(\lambda) = -0.53 \pm 0.61; \chi^2(1) = 0.59; p = 0.44]$. Based on these results the proportion of L2 speakers has no effect on the number of genders.

Figure 12 presents the number of genders as a function of the demographic variables in models GENDER.L1 and GENDER.L2. The curve indicates the fit of the mixed regression model. The figure on the left (**Figure 12A**) presents the fit to log number of L1 speakers. As is evident from the plot, the fitted line is almost flat. The figure on the right (**Figure 12B**) presents the fit to the proportion of L2 speakers. There is a small downward trend so that in communities with few L2 speakers the predicted number of genders is about three and approaching two as the percentage of L2 speakers grows closer to 100%.

According to the zero-inflated negative binomial mixed logistic regression of model GENDER.L2+, the proportion of L2 speakers (including semi-speakers) had an inverse but non-significant effect on the number of genders $[\log(\lambda) = -0.67 \pm 0.46; \chi^2(1) = 1.29; p = 0.26]$. We again tested the random effect structure with maximum likelihood ratio tests and removed the random slope for both Stocks and Area [random slope over Stocks: $\chi^2(1) = 0.09; p = 0.76$; random slope over Area: $\chi^2(1) = 0.01; p = 0.93$]. According to the reduced model the proportion of L2 speakers (including semi-speakers) had an inverse but non-significant effect on the number of genders $[\log(\lambda) = -0.59 \pm 0.52; \chi^2(1) = 0.93; p = 0.34]$. Based on this result the proportion of L2 speakers had no effect on the number of genders.

In model GENDER.ALL we model the effects of the log number of L1 speakers and the proportion of L2 speakers in competition with one another. This time we include semi-speakers for reasons of improved convergence compared to when excluding semi-speakers. According to the zero-inflated negative binomial mixed logistic regression of the maximal model, log number of L1 speakers had a non-significant inverse effect on the number of genders $[\log(\lambda) = -0.13 \pm 0.21; \chi^2(1) = 0.88; p = 0.35]$. The proportion of L2 speakers had also an inverse but non-significant effect on the number of genders $[\log(\lambda) = -0.258 \pm$
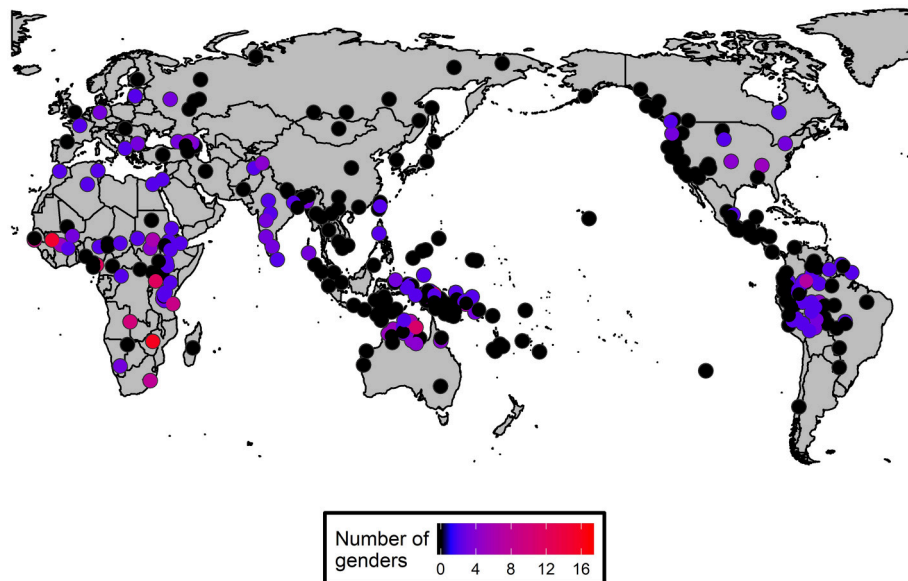
FIGURE 10 | The distribution of number of genders on a world map (In the figure black dots represent languages with no gender and blue dots represent those with two genders. The deeper the red color, the more genders the language has).
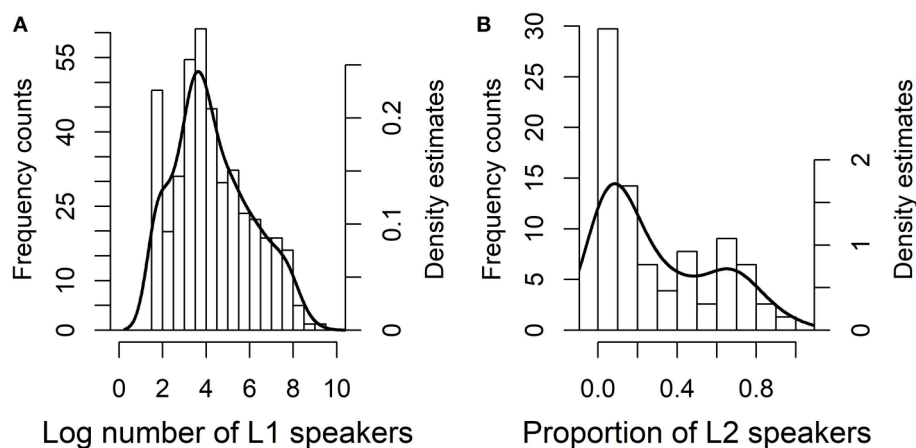

FIGURE 11 | Frequency histograms and superimposed density estimates for the independent demographic variables in case study 2, for log number of native speakers on the left **(A)** and for the proportion of L2 speakers (including semi-speakers) on the right **(B)**.

0.74; $\chi^2(1) = 0.51$; $p = 0.48$]. We again tested the random effect structure with maximum likelihood ratio tests because most of the random effect variances for the slopes (both Stocks and Area) were very close to zero (of the magnitude of 1e-7) and ended up removing all the random slopes (all were non-significant). According to the reduced model the log number of L1 speakers had a non-significant inverse effect on the number of genders [$\log(\lambda) = -0.05 \pm 0.10$; $\chi^2(1) = 0.21$; $p = 0.65$] and so did the proportion of L2 speakers [$\log(\lambda) = -0.62 \pm 0.50$; $\chi^2(1) = 1.08$; $p = 0.30$].

We further used AICc for model comparison; the results are reported in **Table 7** in decreasing order of AICc. The model (4)

which contained only the random intercepts but no fixed effects had the smallest AICc value (245.5). Based on the Akaike weights this model had more than 50% chance of being the best model among the four models. These results clearly suggests that neither of the demographic factors had any meaningful effect on the distribution of the number of genders.

As a summary, the results of study 2 suggest that the number of L1 speakers and the proportion of L2 speakers do not have a significant effect on the number of genders. The estimate was negative for both demographic factors (except in GENDER.L1), but since the effects were non-significant and the AICc values were small, the only reliable conclusion to draw from these results
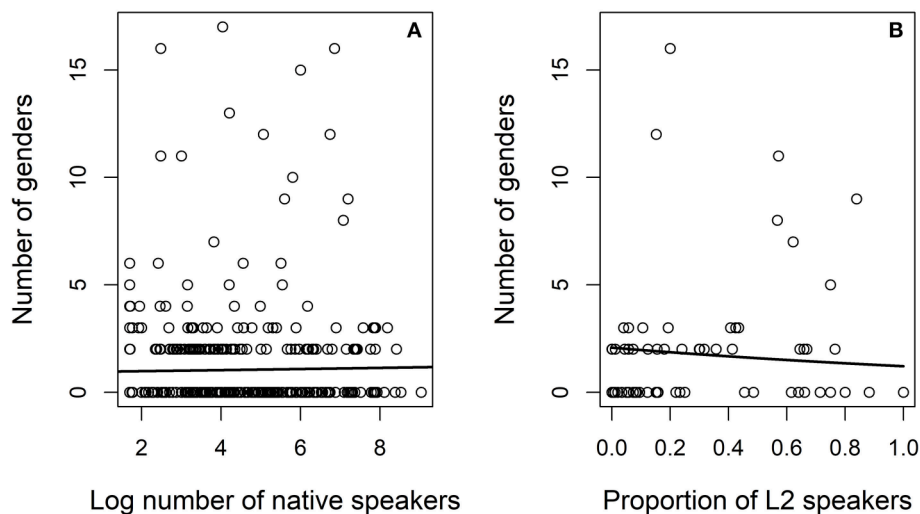
FIGURE 12 | Number of genders as a function of log number of speakers on the left **(A)** and for the proportion of L2 speakers on the right **(B)**.

TABLE 7 | Results of model comparison for the reduced model GENDER.ALL. The full model includes both the log number of L1 speakers and the proportion of L2 speakers (including semi-speakers).

Model	Model structure	Df	AICc	Reduction in AICc	Akaike weights
(1)	Full model	5	249.2	0.0	0.080
(2)	Log number of L1 speakers	4	247.9	1.3	0.160
(3)	Proportion of L2	4	247.0	2.2	0.247
(4)	Only random structure	3	245.5	1.5	0.512

is that the log number of L1 speakers and the proportion of L2 speakers have no effect on the number of genders.

3.3.3. Discussion
None of the four statistical tests that we carried out to investigate the relationship between number of gender distinctions and population data yielded significant results. These (negative) findings replicate and expand on previous research by Dahl (unpublish) and suggest that the hypothesis whereby gender systems adapt to demographic variables must be rejected, at least based on the present data set.

Even though all the tests failed to reach significance, one interesting pattern emerged from the data as a function of the feature values assigned to our dependent variable "Number of genders." We first tested whether the overall results could be affected by counting the exact number of genders for any of the sampled languages. Thereafter, we tested the relationship between the number of genders and population structure by using the classification of Corbett (2013a) in *WALS*. This classification uses five values for number of gender distinctions: "none," "two," "three," "four," "five or more." Conflating number

of genders greater than four into one bin, "five or more," means to assume that a language with, say, 12 genders would not behave differently from a language with five genders. However, we found that using the *WALS* classification had a big impact on the results.

In particular, when we modeled the effect of the proportion of L2 speakers on number of genders and used the exact count of gender distinctions for languages with more than five genders, we found a non-significant negative correlation between the proportion of L2 and the number of genders. When following the *WALS* coding, which collapses together all languages with five or more genders, the observed coefficient between the number of genders and L2 proportions was instead positive [maximal zero-inflated negative binomial model; $\log(\lambda) = 0.84 \pm 0.45$; $\chi^2(1) = 2.98$; $p = 0.11$], even though still non-significant. This same pattern was observed when the proportion of L2 speakers also included the number of semi-speakers. The correlation coefficient was negative (but non-significant) when the exact number of genders was factored in, but it became positive (and still non-significant) when we followed the *WALS* data coding structure [maximal negative binomial model; $\log(\lambda) = 0.66 \pm 0.44$; $\chi^2(1) = 1.61$; $p = 0.20$], that is, when we lumped together languages with five or more gender distinctions.

As for the number of L1 speakers, the choice of coding had a parallel outcome. When we modeled the effect of the number of L1 speakers on the number of genders and used the exact count of gender distinctions, we found a non-significant positive correlation between the variables. When, following the *WALS* coding, we collapsed together all languages with five or more genders the observed estimate was instead negative [maximal zero-inflated poisson model; $\log(\lambda) = -0.014 \pm 0.05$; $\chi^2(1) = 0.07$; $p = 0.79$], even though still non-significant.

While these results do not affect the overall outcome of the case study, the mismatching patterns demonstrate that data structure and data coding may act as crucial confounding factors when running statistical tests on already available databases.

In this particular case, the results suggest that a less abstract coding approach than the one adopted by *WALS* is preferable when investigating sociolinguistic correlates of number of gender distinctions and that the assumption we make about the behavior of languages with five or more genders matters crucially.

With regard to data coding, a parallel case reported in the literature is the correlation between phoneme inventory size and population size by Atkinson (2011). Using the *WALS* data, Atkinson (2011) arrived at a significant negative correlation between phoneme inventory and population size, which seemed to be connected to geographical spread, namely, to the spread of languages out of Africa. The *WALS* data for number of consonants divides data into five bins: "small," "moderately small," "average," "moderately large," "large." Maddieson et al. (2011) took the underlying data for the same *WALS* chapter and still found a significant correlation, but Donohue and Nichols (2011) and Moran et al. (2012) used completely different data sets and found no significant correlation between phoneme inventory and population size reflected there. Alongside with our own results from number of genders and population size, the controversy about phoneme inventory and population data thus suggests that data, and data coding, clearly matter.

In addition, our impression is that, particularly in the case of grammatical gender, the confounding effect of data and data coding may even be a reflection of the type of variable chosen as a proxy of complexity. As outlined in section 3, recent research (Audring, 2014, 2017; Di Garbo, 2016) posits that number of gender distinctions is one of the three main dimensions of complexity variation in gender systems, along with gender assignment rules (whether gender assignment is semantic/formal, rigid/flexible), and formal marking (which word classes inflect for gender in a given language). These studies show that complexity at the level of gender distinctions predicts complexity in other domains of the gender system. For instance, Di Garbo (2016) observes that out of a sample of 84 African languages, particular instances of flexible gender assignment are only attested in languages with a high number of gender distinctions or a high degree of formal marking. Similarly, Audring (2014) observes that in languages with a high number of gender distinctions, complexity in the domain of formal marking (i.e., presence of gender marking on different types of targets in the clause) may facilitate the learning and use of gender distinctions (the more occurrences of gender marking within the utterance the easier to remember the gender of a noun). Thus, while it is no doubt that complexity in the domain of number of gender distinctions bears relevant interactions with complexities in other areas of the gender system of a language, it may well be that this type of complexity is not sensitive (or not in straightforward ways) to the effect of sociolinguistic variables. This would suggest that, in order to investigate the sociolinguistic typology of gender systems from a quantitative point of view, other typological variables than number of genders must be used. This consideration, which is also embraced by Dahl (unpublish), is the point of departure of recent research by Di Garbo and Verkerk (2017). They observe that neither the number of genders nor any of the other *WALS* variables for gender systems directly tackle the morphosyntactic encoding of gender distinctions, that is, the structural properties of gender marking systems. Under the assumption that it is morphosyntax which is directly sensitive to the effect of sociolinguistic variables, they thus look at synchronic variation in gender marking patterns in a sample of 253 Bantu languages, which are well known in the literature for their rather elaborated systems of gender marking. The study finds a significant positive correlation between incidence of restructuring in gender marking and population size whereby languages with larger populations show a preference for restructured gender marking systems[9]. This result partially contradicts the findings on creole languages by Blasi et al. (2017), who find no evidence for adaptive patterns in gender marking on adjectival modifiers and personal pronouns, the two gender-related variables included in the *APICS* database (Michaelis et al., 2013), which the study is based upon. However, while Blasi et al. (2017) only look at these two domains of gender marking, Di Garbo and Verkerk take into account a wider range of syntactic domains (adnominal modification, predication, relative constructions and pronouns) and, within each of these domains they consider different kind of gender marking hosts (for instance, within the domain of adnominal modification, they look not only at adjectival modifiers but also at numerals, demonstratives, quantifiers and question words). These results thus suggest that support to the linguistic adaptation hypothesis in the domain of grammatical gender comes from typological variables that are not (entirely) part of those typological databases that have so far been used to run exploratory studies on the relationship between language structures and social structures.

4. GENERAL DISCUSSION AND CONCLUDING REMARKS

Starting from the assumption that languages are complex adaptive systems (Beckner et al., 2009), in this paper we investigated the hypothesis that morphological complexity is sensitive to sociolinguistic variables concerning population structure. This was done by means of two case studies, one in the verbal domain (degree of inflectional synthesis) and one in the nominal domain (grammatical gender). In both case studies, the same type of sociolinguistic data were operationalized as independent variables: population size (measured as log number of L1 speakers) and proportion of L2 speakers (including/excluding semi-speakers in different models). The raw data for the typological variables came from the *AUTOTYP* database for inflectional synthesis on the verb and from Sinnemäki (unpublish) and *WALS* (Corbett, 2013a) for grammatical gender. The raw demographic data were taken mostly from the *Ethnologue* (see the Supporting Material). While the results of case study 1 confirm that morphological complexity in the verbal domain is sensitive to population dynamics thus bringing support to the main hypothesis, the same could not be observed in the case of grammatical gender (case study 2).

However, irrespectively of how well the individual case studies support the main hypothesis, we think that both make a relevant

---

[9]In this study, restructured gender systems are systems in which gender marking is partially or heavily based on animacy distinctions

contribution to the understanding of non-linguistic correlates of linguistic diversity. First, the results of the two case studies suggest that not all domains of grammar adapt to sociolinguistic variables to the same extent. More specifically, our data show that while the degree of inflectional synthesis is sensitive to population data, the number of gender distinctions is not. Whether this discrepancy is related to the different functions that the two grammatical domains display in discourse is an open question whose answer we leave to further studies. Our results ultimately suggest that no general prediction can be made about the relationship between morphological complexity and population data because the outcomes of this relationship are very much specific to the grammatical domain under study.

Second, the results from study 1 suggest that competitive models, where the effect of multiple sociolinguistic variables on language structures are tested simultaneously, are somewhat better than non-competitive models, where each factor is tested in isolation. These findings bring quantitative evidence in support of Trudgill's (2011a,b) suggestion that the effect of social structures on language structures must be studied by factoring in a multifaceted array of interacting variables, ranging from population size to degree of contact and social network density. While our study covers two of the three suggested dimensions—population size and degree of contact (of which the proportion of L2 speakers is taken as a proxy)—nothing could be said about social network density. Operationalizing social network density as one of the critical variables in quantitative sociolinguistic typology would, in fact, require accessing a type of data that is at present not featured in existing databases.

Third, in line with previous studies addressing similar research questions, case study 2 fails to show any significant relationship between the complexity of grammatical gender systems (measured in terms of number of gender distinctions) and sociolinguistic variables. These results contradict the well-known observation (supported by evidence from different linguistic families and areas) that while gender systems are generally very stable, their transmission tends to be disrupted under the pressure of language contact. In line with a recent suggestion by Dahl (unpublish) and ongoing research on the topic (Di Garbo and Verkerk, 2017) we think that a reasonable explanation behind this mismatch may be that the number of gender distinctions is not a suitable measure to test hypotheses on linguistic adaptation in the domain of grammatical gender, and that typological variables pertaining to patterns of gender marking should instead be considered. In addition, we found that using the number of gender distinctions as coded in *WALS*, with five cut-off points between no gender, two, three, four, five

or more gender distinctions, leads to less accurate results than following a less abstract coding procedure where languages with richer gender systems are coded based on the exact number of distinctions that they display. For these reasons, we conclude that existing typological databases are not fully equipped to support quantitative sociolinguistic typologies of grammatical gender systems.

To sum up, while at least for one of the grammatical domains used as test cases this paper confirms the validity of the linguistic adaptation hypothesis, the paper also shows that a precondition to the advancement of research on nonlinguistic correlates of linguistic diversity lies in the refinement of the statistical methodologies used to test this hypothesis as well as in the types of data and data coding principles that are fed into the analyses. In order to test hypotheses about sociolinguistic typology, comparative data on sociolinguistic variables other than demographic variables, such as relative prestige, literacy, and multilingualism, need to be collected. Furthermore, given that approaching linguistic structures (and their complexity) from different perspectives may produce radically different results about adaptation, more exploratory studies need to be run in order to test which domains of grammar and what types of language structures within a given domain are most sensitive to the effect of social structures.

AUTHOR CONTRIBUTIONS

The research was designed together by KS and FD, data collection was done primarily by KS (data on the number of genders for languages with 5+ genders was collected together by FD and KS), statistical analyses for the typological case studies were done by KS, and the write-up was done together by KS and FD.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01141/full#supplementary-material

REFERENCES

Alungum, J., Conrad, R. J., and Lukas, J. (1978). "Some Muhiang grammatical notes," in *Miscellaneous Papers on Dobu and Arapesh, Volume 25 of Workpapers in Papua New Guinea Languages*, ed R. Loving (Ukarumpa: Summer Institute of Linguistics), 89–130.

Arefyev, A. L. (2012). *The Russian Language at the Turn of the 21st Century (in Russian)*. Moscow: Center of Social Forecasting and Marketing.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–349. doi: 10.1126/science.1199295

Audring, J. (2014). Gender as a complex feature. *Lang. Sci.* 43, 5–17. doi: 10.1016/j.langsci.2013.10.003

Audring, J. (2017). Calibrating complexity: how complex is a gender system? *Lang. Sci.* 60, 53–68. doi: 10.1016/j.langsci.2016.09.003

Baayen, H. R. (2013). "Multivariate statistics," in *Research Methods in Linguistics*, eds R. J. Podesva and D. Sharma (Cambridge: Cambridge University Press), 337–372.

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

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Barth, D., and Kapatsinski, V. (2018). "Evaluating logistic mixed-effects models of corpus-linguistic data in light of lexical diffusion," in *Mixed-Effects Regression Models in Linguistics, Quantitative Methods in the Humanities and Social Sciences*, eds D. Speelman, K. Heylen, and D. Geeraerts (Cham: Springer), 99–116.

Barton, K. (2018). *MuMIn: Multi-Model Inference*. R package version 1.40.4.

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59, 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Bentz, C., and Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Lang. Dyn. Change* 3, 1–27. doi: 10.1163/22105832-13030105

Bickel, B., and Nichols, J. (2007). "Inflectional morphology," in *Language Typology and Syntactic Description, Vol. 3*, ed T. Shopen (Cambridge: Cambridge University Press), 169–240.

Bickel, B., and Nichols, J. (2013). "Chapter 22: Inflectional synthesis of the verb," in *The World Atlas of Language Structures Online*, eds M. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., et al. (2017). *The AUTOTYP Typological Databases*. Version 0.1.0. Available online at: https://github.com/autotyp/autotyp-data/tree/0.1.0

Blasi, D. E., Michaelis, S. M., and Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nat. Hum. Behav.* 1, 723–729. doi: 10.1038/s41562-017-0192-4

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edn.* New York, NY: Springer.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., and Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* 78, 685–709. doi: 10.1007/s11336-013-9328-2

Clahsen, H., Fester, C., Neubauer, K., Sato, M., and Silva, R. (2010). Morphological structure in native and non-native language processing. *Lang. Learn.* 60, 21–43. doi: 10.1111/j.1467-9922.2009.00550.x

Corbett, G. (2013a). "Chapter 30: Number of genders," in *The World Atlas of Language Structures Online*, eds M. Dryer and M. Haspelmath (Max Planck Digital Library). Available online at: http://wals.info/chapter/30 (Accessed February 2, 2014).

Corbett, G. (2013b). "Chapter 31a: Sex-based and non-sex-based gender systems," in *The World Atlas of Language Structures Online*, eds M. Dryer and M. Haspelmath (Max Planck Digital Library). Available online at: http://wals.info/chapter/31 (Accessed February 2, 2014).

Corbett, G. (2013c). "Systems of gender assignment," in *The World Atlas of Language Structures Online*, eds M. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: http://wals.info/chapter/32 (Accessed February 2, 2014).

Cysouw, M. (2010). Dealing with diversity: towards an explanation of NP-internal word order frequencies. *Linguist. Typol.* 14, 253–286. doi: 10.1515/lity.2010.010

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.

de Groot, C. (2008). "Morphological complexity as a parameter of linguistic typology: Hungarian as a contact language," in *Language Complexity: Typology, Contact, Change, Vol. 94 of Studies in Language Companion Series*, eds M. Miestamo, K. Sinnemäki, and F. Karlsson (Amsterdam: John Benjamins), 191–215.

Dediu, D., and Cysouw, M. (2013). Some structural aspects of language are more stable than others: a comparison of seven methods. *PLoS ONE* 8:e55009. doi: 10.1371/journal.pone.0055009

DeLancey, S. (2014). Sociolinguistic typology in North East India: a tale of two branches. *J. South Asian Lang. Linguist.* 1, 59–82. doi: 10.1515/jsall-2014-0004

Desmuth, K. (2000). "Bantu noun class systems: loanword and acquisition evidence of semantic productivity," in *Systems of Noun Classification*, ed G. Senft (Cambridge: Cambridge University Press), 270–292.

Di Garbo, F. (2016). Exploring grammatical complexity crosslinguistically: the case of gender. *Linguist. Discov.* 14, 46–85. doi: 10.1349/PS1.1537-0852.A.468

Di Garbo, F., and Verkerk, A. (2017). *On the Evolution of Bantu Gender Marking Systems: Typology, Phylogeny, and Social History*. Invited lecture at the Department of African Languages and Linguistics, Humboldt University, Berlin, 14 November 2017.

Dixon, R. M. W., and Aikhenvald, A. Y. (eds.). (2002). *Word: A Cross-Linguistic Typology*. Cambridge: Cambridge University Press.

Donohue, M., and Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguist. Typol.* 15, 161–170. doi: 10.1515/LITY.2011.011

Eichler, N., Jansen, V., and Müller, N. (2013). Gender acquisition in bilingual children: French-German, Italian-German, Spanish-German and Italian-French. *Int. J. Bilingual.* 17, 550–572. doi: 10.1177/1367006911435719

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1322–1327. doi: 10.1073/pnas.1417413112

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., et al. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* 27, 233–249. doi: 10.1080/10556788.2011.597854

Fox, J. (2003). Effect displays in R for generalised linear models. *J. Stat. Softw.* 8, 1–27. doi: 10.18637/jss.v008.i15

Gagliardi, A., and Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90, 58–89. doi: 10.1353/lan.2014.0013

Givón, T. (2009). *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam: John Benjamins.

Göksel, A., and Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. London: Routledge.

Grinevald, C. (2003). "Speakers and documentation of endangered languages," in *Language Documentation and Description, Vol. 1*, ed P. K. Austin (London: SOAS), 52–72.

Gunter, T. C., and Friederici, A. D. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 14, 556–568. doi: 10.1162/089892900562336

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguist.* 45, 31–80. doi: 10.1515/flin.2011.002

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression, 2nd Edn*. New York, NY: John Wiley & Sons.

Jaeger, T. F., Graff, P., Croft, W., and Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguist. Typol.* 15, 281–319. doi: 10.1515/lity.2011.021

Johnson, P. C. D., and O'Hara, R. B. (2014). Extension of nakagawa & schielzeth's R2GLMM to random slopes models. *Methods Ecol. Evol.* 5, 944–946. doi: 10.1111/2041-210X.12225

Kettunen, K. (2014). Can type-token ratio be used to show morphological complexity of languages? *J. Quant. Linguist.* 21, 223–245. doi: 10.1080/09296174.2014.911506

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

Ladd, D. R., Roberts, S. G., and Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annu. Rev. Linguist.* 1, 221–241. doi: 10.1146/annurev-linguist-030514-124819

Lewis, M. P., Simons, G. F., and Fennig, C. D. (eds.). (2016). *Ethnologue: Languages of the World, 19th Edn*. Dallas, TX: SIL International).

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Maddieson, I., Bhattacharya, T., Smith, D. E., and Croft, W. (2011). Geographical distribution of phonological complexity. *Linguist. Typol.* 15, 267–279. doi: 10.1515/lity.2011.020

Maitz, P., and Németh, A. (2014). Language contact and morphosyntactic complexity: evidence from German. *J. German. Linguist.* 26, 1–29. doi: 10.1017/S1470542713000184

McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguist. Typol.* 5, 125–156. doi: 10.1515/lity.2001.001

McWhorter, J. (2007). *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars.* New York, NY: Oxford University Press.

Michaelis, S. M., Maurer, P., Haspelmath, M., and Huber, M., (eds.). (2013). *Atlas of Pidgin and Creole Language Structures Online.* Max Planck Institute for Evolutionary Anthropology. Available online at: http:/apics-online.info

Mills, A. E. (1986). *The Acquisition of Gender: A Study of English and German.* Berlin: Springer.

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs phoneme inventory size. *Language* 88, 877–893. doi: 10.1353/lan.2012.0087

Moravcsik, E. A., and Wirth, J. R. (1986). "Markedness: an overview," in *Markedness*, eds F. R. Eckman, E. A. Moravcsik, and J. R. Wirth (New York, NY: Plenum), 1–11.

Moscoso del Prado Martín, F., Kostic, A., and Baayen, H. R. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94, 1–18. doi: 10.1016/j.cognition.2003.10.015

Mulford, R. (1985). Comprehension of Icelandic pronoun gender: semantic versus formal factors. *J. Child Lang.* 12, 443–453.

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Nichols, J. (2009). "Linguistic complexity: a comprehensive definition and survey," in *Language Complexity as an Evolving Variable*, eds G. Sampson, D. Gil, and P. Trudgill (Oxford: Oxford University Press), 110–125.

Parodi, T., Schwartz, B. D., and Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics* 42, 669–705. doi: 10.1515/ling.2004.022

Plank, F. (1998). The co-variation of phonology with morphology and syntax. *Linguist. Typol.* 2, 195–230.

Plank, F. (1999). Split morphology: how agglutination and flexion mix. *Linguist. Typol.* 3, 279–340.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Sapir, E. (1921). *Language.* New York, NY: Harcourt.

Schielzeth, H., and Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behav. Ecol.* 20, 416–420. doi: 10.1093/beheco/arn145

Shosted, R. K. (2006). Correlating complexity: a typological approach. *Linguist. Typol.* 10, 1–40. doi: 10.1515/LINGTY.2006.001

Sinnemäki, K. (2009). "Complexity in core argument marking and population size," in *Language Complexity as an Evolving Variable*, eds G. Sampson, D. Gil, and P. Trudgill (Oxford: Oxford University Press), 125–140.

Sinnemäki, K. (2014). Cognitive processing, language typology, and variation. *WIREs Cogn. Sci.* 5, 477–487. doi: 10.1002/wcs.1294

Sinnemäki, K. (2010). Word order in zero-marking languages. *Stud. Lang.* 34, 869–912. doi: 10.1075/sl.34.4.04sin

Skaug, H., Fournier, D., Bolker, B., Magnusson, A., and Nielsen, A. (2016). *Generalized Linear Mixed Models using 'AD Model Builder'.* R package version 0.8.3.3.

Skaug, H. J., and Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Comput. Stat. Data Anal.* 51, 699–709. doi: 10.1016/j.csda.2006.03.005

Skelton, C. (2017). Greek-Anatolian language contact and the settlement of Pamphylia. *Classic. Antiq.* 36, 144–129. doi: 10.1525/ca.2017.36.1.104

Slobin, D. I. (2005). "From ontogenesis to phylogenesis: what can child language tell us about language evolution?," in *Biology and Knowledge Revisited: From Neurogenesis to Psychogenesis*, eds S. T. Parker, J. Langer, and C. Milbrath (Mahwah, NJ: Lawrence Erlbaum Associates), 255–286.

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.

Suzman, S. (1980). Acquisition of the noun class system in Zulu. *Pap. Rep. Child Lang. Dev.* 19, 45–52.

Thomason, S. G. (2015). *Endangered Languages: An Introduction.* Cambridge: Cambridge University Press.

Thompson, L. C. (1987). *A Vietnamese Reference Grammar.* Honolulu, HI: University of Hawaii Press.

Trudgill, P. (1999). Language contact and the function of linguistic gender. *Poznań Stud. Contemp. Linguist.* 35, 133–152.

Trudgill, P. (2001). Contact and simplification: historical baggage and directionality in linguistic change. *Linguist. Typol.* 5, 372–375.

Trudgill, P. (2011a). Social structure and phoneme inventories. *Linguist. Typol.* 15, 155–160. doi: 10.1515/lity.2011.010

Trudgill, P. (2011b). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity.* Oxford: Oxford University Press.

Trudgill, P. (2017). "The anthropological setting of polysynthesis," in *The Oxford Handbook of Polysynthesis*, eds M. Fortescue, M. Mithun, and N. Evans (Oxford: Oxford University Press), 186–202.

Ullman, M. T. (2005). "A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model," *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*, ed C. Sanz (Washington, DC: Georgetown University Press), 141–178.

von Mengden, F., and Coussé, E. (2014). "Introduction: The role of change in usage-based conceptions of language," in *Usage-Based Approaches to Language Change*, eds E. Coussé and F. von Mengden (Amsterdam: John Benjamins), 1–19.

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Conflict of Interest Statement: 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.

Sociolinguistic Typology and Sign Languages

Adam Schembri[1]*, Jordan Fenlon[2], Kearsy Cormier[3] and Trevor Johnston[4]

[1] Department of English Language and Applied Linguistics, University of Birmingham, Birmingham, United Kingdom,
[2] Languages and Intercultural Studies, Heriot-Watt University, Edinburgh, United Kingdom, [3] Deafness Cognition and Language Research Centre, University College London, London, United Kingdom, [4] Department of Linguistics, Macquarie University, Sydney, NSW, Australia

This paper examines the possible relationship between proposed social determinants of morphological 'complexity' and how this contributes to linguistic diversity, specifically via the typological nature of the sign languages of deaf communities. We sketch how the notion of morphological complexity, as defined by Trudgill (2011), applies to sign languages. Using these criteria, sign languages appear to be languages with low to moderate levels of morphological complexity. This may partly reflect the influence of key social characteristics of communities on the typological nature of languages. Although many deaf communities are relatively small and may involve dense social networks (both social characteristics that Trudgill claimed may lend themselves to morphological 'complexification'), the picture is complicated by the highly variable nature of the sign language acquisition for most deaf people, and the ongoing contact between native signers, hearing non-native signers, and those deaf individuals who only acquire sign languages in later childhood and early adulthood. These are all factors that may work against the emergence of morphological complexification. The relationship between linguistic typology and these key social factors may lead to a better understanding of the nature of sign language grammar. This perspective stands in contrast to other work where sign languages are sometimes presented as having complex morphology despite being young languages (e.g., Aronoff et al., 2005); in some descriptions, the social determinants of morphological complexity have not received much attention, nor has the notion of complexity itself been specifically explored.

Keywords: sign languages, sociolinguistics, typology, language complexity, morphology, linguistic diversity

INTRODUCTION

In this paper, we examine the possible relationship between proposed social determinants of morphological complexity (Trudgill, 2011), the typological nature of the sign languages of deaf communities, and how this contributes to an understanding of linguistic diversity. We review the notion of morphological complexity as defined by Trudgill and how it applies to the grammar of sign languages, with a focus on British Sign Language (BSL), Australian Sign Language (Auslan) and American Sign Language (ASL). We then discuss the sociolinguistic situation of sign languages.

SOCIOLINGUISTIC TYPOLOGY

Interest in social structures and linguistic diversity dates back at least a century, as pointed out by Perkins (1992). Based on cross-linguistic evidence, a number of scholars have proposed that spoken languages which undergo extensive second language acquisition by adults appear to have relatively less inflectional complexity (Kusters, 2003; Dahl, 2004; McWhorter, 2007; Wray and Grace, 2007; Miestamo et al., 2008; Sampson et al., 2009). This would suggest that the default state for human languages (i.e., those which lack a history of extensive second language acquisition by adults) is a high degree of morphological complexification, as appears to be true of languages such as the Athabaskan language Navajo (with its highly irregular verbal system) or Yimas (with its rich tense system) spoken in Papua New Guinea. As a result, the moderate degree of morphological complexity of languages like English and French might thus be viewed as a 'sociohistorical anomaly' (McWhorter, 2012), resulting from the particular sociolinguistic histories of these two major languages.

Trudgill (2011) introduced the term *sociolinguistic typology*: a 'sociolinguistically informed' approach to linguistic typology. This approach assumes that, despite a common set of communicative pressures and cognitive abilities in all humans, different types of languages develop in different places and at different points in time partly as a result of the influence of varying sociolinguistic situations. In particular, this theory proposes that there are specific distinctive social characteristics of speech communities that mold the grammatical organization of their languages. Trudgill (2011) proposed the following factors: (1) population size, (2) social network density, (3) degree of communally shared information, (4) social stability, and (5) degree of language and dialect contact. Morphological complexification, Trudgill suggests, tends to be found in languages used by small communities, composed of dense social networks, with high degrees of communally shared information and social stability, and stable situations of language contact. Stable language contact situations refer here to multilingual communities in which one or more languages are learned as children, as opposed to language contact situations in which large numbers of adults learn a second or additional language, perhaps as the result of some significant social change (e.g., displacement caused by war).

MORPHOLOGICAL COMPLEXITY

What does Trudgill (2011) mean by morphological 'complexification'? He proposes that it consists of the following factors: high degrees of (1) irregularity, (2) morphological opacity, (3) syntagmatic redundancy and (4) morphological marking of categories such as tense, gender, voice etc. Trudgill (2011) illustrates (1) by discussing the irregular system of noun declension in Faroese, with the paradigm for the noun *dagur* 'day' showing, for example, completely unrelated forms for accusative [dɛa], genitive [daås] and dative case [de:ji] (compare this to the more regular system for *batur* 'boat,' with accusative

bat, genitive *bats* and dative *bati*). By (2), Trudgill (2011) is referring to the notion that the relationship of form and meaning should be as transparent as possible (Kusters, 2003). In a dialect of North Frisian, however, Trudgill reports that, depending on the syntactic context, the infinitive form of 'do' has several variant morphological forms (i.e., allomorphs), with it appearing either as *douen*, *doue* or *dou*. Trudgill (2011) illustrates (3) with data from East Flemish dialects in which subject arguments involves triple-marking as in *we zulle-me wij dat doen* 'we shall do that' (literally 'we shall-we we that do'). Lastly, with (4) he explores how the morphological marking in the demonstrative system in some dialects of Norwegian has evolved a three-way distinction between proximal demonstratives *denne/dette/desse* which are equivalent to 'this' in English, distal demonstratives *danna/data/dassa* which are similar to English 'that' but are used for something that the speaker can point to in contrast to a third type of demonstrative – i.e., the forms *den/dae/dei* which refer to something that is not visible but has been recently mentioned in the conversation.

These aspects of morphological complexity, Trudgill (2011) claims, predominate in smaller, dense, stable communities without large-scale adult second language contact. In fact, many of the examples he describes in Faroese, Frisian, Flemish and Norwegian have emerged in small dialect speaking communities, and represent complexifications in comparison to more standard varieties of each language. He suggests that, as all of these features appear to be difficult for post-critical-period adult learners to master, this reflects that fact that one expects to see morphological simplification – i.e., the reduction in features (1) to (4) – in languages spoken by larger communities with looser social networks that have greater numbers of adult second language learners. Evidence supporting this hypothesis includes a recent study, for example, showing that spoken languages with large numbers of adult second language learners tend to lose nominal case systems (Bentz and Winter, 2013).

MORPHOLOGICAL COMPLEXITY AND SIGN LANGUAGES

We would like to focus here on how Trudgill's (2011) notion of sociolinguistic typology can inform, and can be informed by, the study of sign languages of deaf communities. To our knowledge, this notion has only been partly explored in relation to sign languages (Meir et al., 2012), and the specific predictions of Trudgill's proposal have not yet been applied to the languages of deaf communities. Sign languages can be divided into two very broad subclasses: (1) 'macro-community' sign languages which may be used across an entire national deaf community, such as BSL, Auslan, ASL, German Sign Language (DGS) and Taiwan Sign Language (TSL), and (2) 'micro-community' sign languages which are used by smaller communities within a nation state, such as the so-called 'village sign languages' Kata Kolok in Bali and Al Sayyid Bedouin Sign Language in Israel (see Schembri, 2010 for a description of these two community types). These two types of sign language have developed in quite different social situations, so below we explore how they may provide an

interesting test case for the proposal by Trudgill (2011), albeit with some important qualifications.

First, we consider how the notion of morphological complexity might apply to sign languages. Applying Trudgill's (2011) theory to sign languages is controversial because there is little consensus on how some aspects of their structural organization are best analyzed. Sign languages are often described as morphologically complex languages (e.g., Supalla, 1982, unpublished). Indeed, some researchers have characterized the fact that sign languages appear to have complex morphology despite being young languages a 'paradox' (e.g., Aronoff et al., 2005). In contrast, a small number of linguists (e.g., Bergman and Dahl, 1994; Liddell, 2003a) have described sign languages as inflectionless languages, but this view is not widely accepted. After a brief overview of morphology in sign languages, we will work through each of the main features of morphological complexity that Trudgill (2011) discusses, with a focus on BSL, Auslan and ASL (the sign language varieties with which the authors of this paper are most familiar). As we will see, it appears that Trudgill's notion of morphological complexity and the social determinants associated with it offer some fresh insights into this debate about the structure of sign languages: drawing on this work, we might argue that there is, in fact, no 'paradox' to solve.

First, we provide a little background about sign language structure. Formationally, signs in BSL, Auslan and ASL are composed of contrastive hand configurations, locations on the body or in the space around the signer, movements of the hands, and non-manual features, such as mouth gestures and facial expressions. Morphologically, these formational features may be modified to convey a range of meanings, some of which we explain in more detail below (Sutton-Spence and Woll, 1999; Liddell, 2003a; Johnston and Schembri, 2007). Many of these morphological patterns are widely found in unrelated sign languages, perhaps because they are clearly iconically motivated. For example, time-related signs may incorporate numeral handshapes to show number (e.g., TOMORROW versus IN-TWO-DAY'S-TIME in Auslan in **Figure 1A**). A subset of verb signs, which we will refer to here as *indicating verbs*, may be directed toward locations associated with the referents of the verb's arguments, as we see in **Figure 3**. Another category of verb signs, known as *classifier constructions* or *depicting signs*, include handshape morphemes that represent classification of a referent into a number of semantic or shape categories. These handshapes combine with movement and spatial components to build complex iconic representations of the specific referent in motion, its relative location and/or its distribution, as we can see in **Figure 1B**. This example shows three possible combinations of a Auslan classifier handshape for person in relation to another classifier handshape for vehicle. These forms represent perhaps the most complex constructions in signs languages, but researchers do not agree on the most appropriate morphological analysis (e.g., Liddell, 2003b). For example, do the changes in relative location in the sign in **Figure 1B** act as discrete morphemes, or are they some kind of gradient gestural representation? In addition to alternations of distinctive formational features of a sign, reduplication of a subset of nouns is used to signal plurality (e.g., Auslan HOUSE versus

HOUSE[PLURAL], see **Figure 1C**). Fast or slow reduplication of some verb signs may be used to signal habitual versus continuative aspect (as in Auslan JOKE versus JOKE[continuative] in **Figure 1D**). The rich system for modification of signs is what contributes to the claim by many sign language linguists (e.g., Aronoff et al., 2005) that sign languages are morphologically complex languages.

In terms of Trudgill's (2011) criteria for morphological complexity, however, the picture seems more mixed, as few of the phenomena identified as morphologically complex by sign linguists (e.g., classifier constructions) fit into his definition. First, none of these three sign languages (BSL, Auslan, or ASL) exhibit high levels of irregularity in any of the morphological phenomena described above. There are a very small number of irregular negative verb and modal forms in each sign language, including CAN and CANNOT in Auslan and in ASL; SHOULD and SHOULD-NOT in BSL, and HAVE and HAVE-NOT in BSL. Some of the negative forms in BSL/Auslan, however, appear to involve a now unproductive negative suffix, as in DISAGREE (cf. AGREE). This suffix appears to be related to the negative lexical item in BSL/Auslan which can mean 'not have,' 'did not,' 'without' etc. There are also irregular forms meaning 'people' in Auslan and BSL (unrelated to signs meaning 'person'). Apart from these small number of examples, however, there are few other examples of irregularity attested (see BSL SignBank and Auslan SignBank for examples of these signs[1,2]).

There is only limited allomorphy in ASL, BSL and Auslan that cannot be predicted on the basis of morphophonemic processes. For example, in all three sign languages, there is a high degree of variation in the handshape in first person singular pronouns, with the pointing sign directed to the chest appearing as an extended index finger in isolation, but often as some other handshape in connected signing (as we see in **Figure 2** BSL PRO1SG BREATHE 'I breathe' where the handshape in the first person pronoun has all fingers extended, matching the handshape of the following sign BREATHE). Empirical studies indicate that this variation may be conditioned in part by the handshape of the following sign (i.e., it is due to co-articulation, see Bayley et al., 2002; Fenlon et al., 2013). Some isolated examples of unpredictable allomorphy do occur in verbs. In one regional variety of Auslan, there are two forms of the non-first person to first person form of the sign GIVE. The form with the Y handshape (i.e., a little finger and thumb extended from the fist), anecdotal reports suggest, cannot be modified for first to non-first person marking[3]. In ASL, there is a non-first person to first person marked form for CONVINCE that is directed toward a location on the neck, unlike other forms of the verb produced in the signing space in front of the signer's chest. The first person object form has been argued to be an idiosyncratic form (Lillo-Martin and Meier, 2011). However, it could be argued that this form is actually similar to other first person object forms for other indicating verbs which are directed toward particular parts of the body but otherwise are predictable in form (e.g., REMIND, LOOK-AT, etc.).

---

[1] http:bslsignbank.ucl.ac.uk

[2] http://www.auslan.org.au

[3] http://www.auslan.org.au/dictionary/words/give%20back-1.html
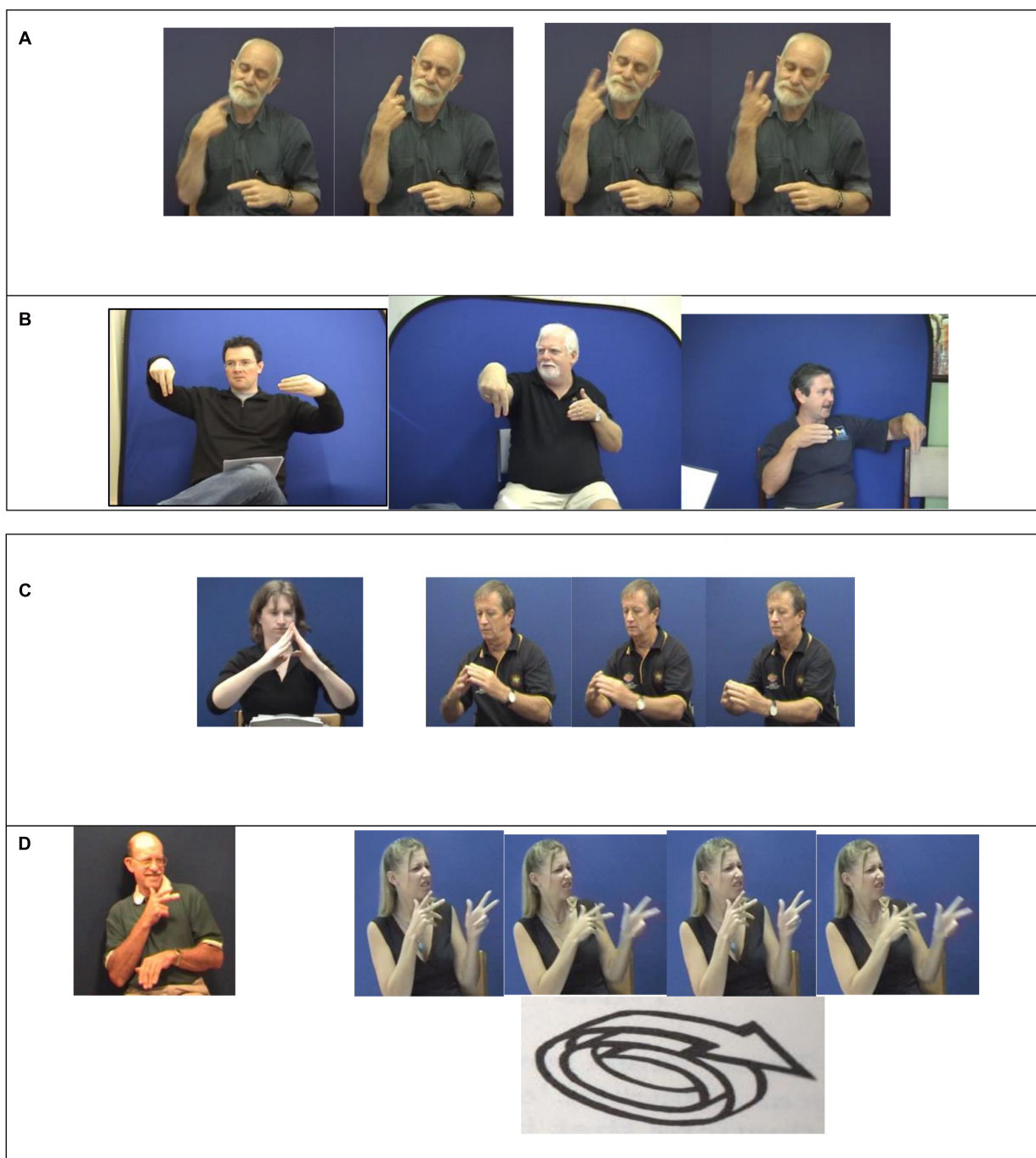
FIGURE 1 | (A) TOMORROW and IN-TWO-DAY'S-TIME in Auslan. **(B)** Various Auslan depicting sign constructions. **(C)** Auslan HOUSE versus Auslan HOUSE [PLURAL].
(D) Auslan JOKE versus Auslan JOKE [CONTINUATIVE].

There is limited syntagmatic redundancy in ASL, BSL, and Auslan, with plural marking of most nouns being optional, for example, even when the nominal occurs with a lexical quantifier or verb modified for number.

ASL, BSL, and Auslan do not employ any morphological markers for gender, tense, or voice. Although some scholars claim that ASL does mark for tense and passive voice (Neidle

et al., 1999; Janzen et al., 2001), the claims are based on syntactic, rather than morphological, phenomena. The marking of aspect mentioned above is clearly iconically motivated and does not appear highly grammaticalized in Auslan (Gray, 2013). Furthermore, the aspect marking system is predictable: it involves the reduplication of punctual verbs marking habitual aspect, for example, whereas a similar modification for durative verbs
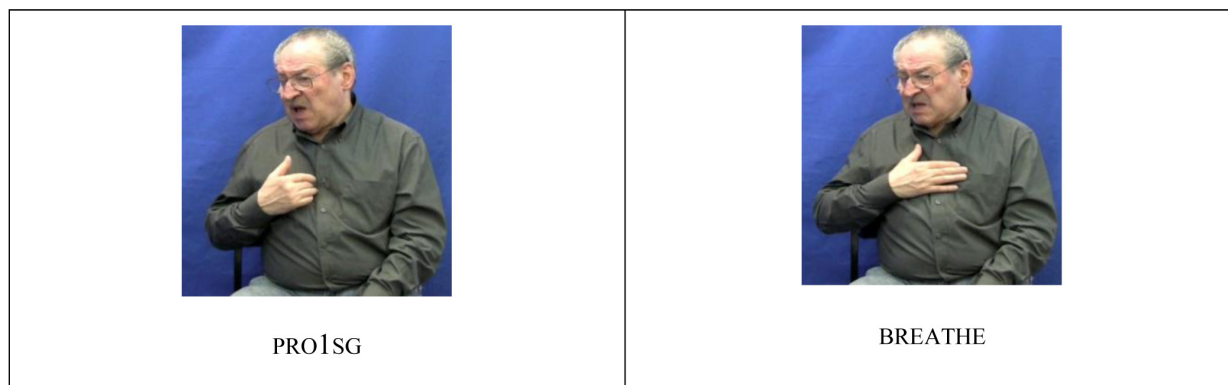
FIGURE 2 | Handshape assimilation in PRO1SG.
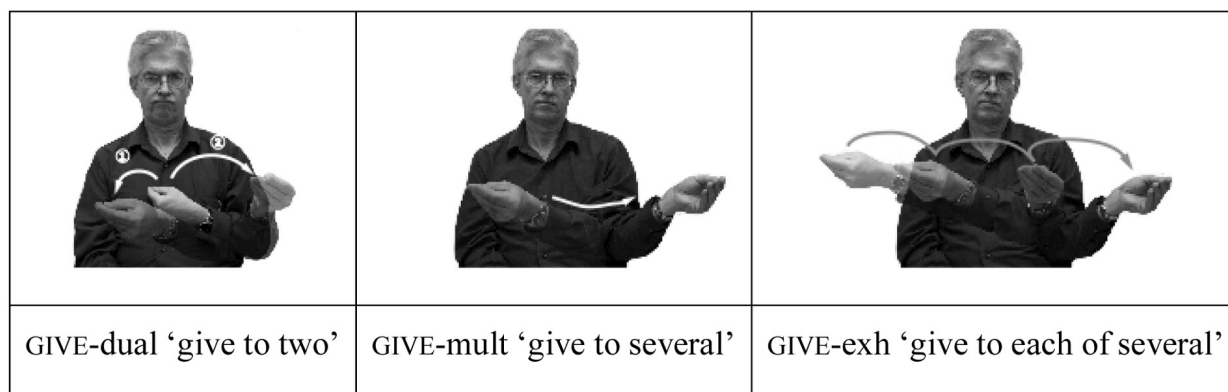
FIGURE 3 | ASL plural forms of indicating verb GIVE.

represents durational aspect. In some sign languages, in fact, aspect marking has been considered ideophonic (Bergman and Dahl, 1994).

Genitive case is optionally marked on nouns in Auslan and some varieties of BSL (Johnston and Schembri, 2007; Cormier and Fenlon, 2009): a possessive marker that is based on fingerspelled '-s' (borrowed from English) is sometimes used, as in (1). ASL also has a possessive marker based on a modified form of fingerspelled '-s' which is also optional (Pichler et al., 2008). This appears to be an example of morphological complexification as a result of language contact.

(1) MOTHER POSSESSIVE-S SISTER 'mother's sister'

Indicating verbs appear to share some characteristics with person and number agreement in spoken languages (Sandler and Lillo-Martin, 2006; Johnston and Schembri, 2007). This modification has been called 'agreement' because it was originally assumed that the form of the verb reflects aspects of the form or semantics of the subject or object noun phrase. In fact, these modifications, like pointing used by non-signers, actually most often reflect the location of a present referent, or the association between an absent referent and a location in the space around the signer's body (Liddell, 2003a; Fenlon et al., in press). This is arguably quite different from what we see in spoken language

agreement systems (Corbett, 2006), and there is considerable debate in the literature about whether it should be called an agreement system at all (e.g., Liddell, 2011; Lillo-Martin and Meier, 2011). Regardless of this debate, it is clear from studies of BSL and Auslan data that this modification is not obligatory (e.g., de Beuzeville et al., 2009; Fenlon et al., in press), as one would expect from a canonical agreement system (Corbett, 2006).

Indicating verb signs may also be modified for number. An optional alternation of location features and reduplication is used to represent number and distribution of object arguments, as shown in **Figure 3**. With two object arguments, the sign may reduplicate to different locations, or may use a two-handed construction ('dual inflection'). With more than two, a sweeping movement may be added across the signing space ('multiple inflection'). Multiple reduplications may signal marking for distribution (the 'exhaustive inflection'). Again, these modifications are clearly iconically motivated, and do not appear to be obligatory for any sign language.

Overall, it might be argued that BSL, Auslan, and ASL are languages with relatively little obligatory inflection and, based on Trudgill's (2011) criteria, low to moderate levels of morphological complexity (in contradistinction to Aronoff et al., 2005). Indeed, previous analyses have compared ASL, BSL, and Auslan grammar

FIGURE 4 | Mouth gesture 'pah' in Kata Kolok (from De Vos, 2012) [reproduced with permission].

to spoken language creoles (Fischer, 1978; Ladd and Edwards, 1982; Johnston, 1989). Aronoff et al. (2005) pose this similarity to creoles as a "young language puzzle": i.e., why is it that sign languages are similar in some ways to spoken language creoles and yet they have complex morphology? Our response is that sign languages, by Trudgill (2011)'s definition, are not as morphologically complex as previously assumed.

SOCIAL STRUCTURE AND SIGN LANGUAGE COMMUNITIES

So, what about the social factors at play in deaf communities? Sign language communities tend to be small, but not as small as many spoken languages. For example, Lupyan and Dale (2010) show that the median number of speakers of the 6,192 languages cataloged by Ethnologue is only 7000, although the mean is over 828,000. The total number of signers in North America, the United Kingdom and Australia numbers in the thousands (although this is likely to be in the hundreds of thousands in the North American case), so all of these sign languages would have a lower number than the mean for all languages given in Ethnologue, with only Auslan possibly approaching the much lower median. In terms of the density of social networks, there has been relatively little research into the network densities of macro-community sign languages (the work of Morris, 2016, being the only example). A small number of deaf individuals are from deaf families, work with deaf people and have deaf partners, and this core of the deaf community might have dense social ties with other signers. Over 95% of deaf people, however, are from hearing families (Mitchell and Karchmer, 2004). It is also likely that most deaf adults work with hearing people, and thus they have considerable contact with social networks that do not include people who can sign. It is not clear how to operationalize the variable related to the degree of communally shared information. This is likely to be high

in terms of deaf community specific information, but access to information about the wider community is often limited and inconsistent, as the provision of sign language interpreting and captioning on broadcast video is patchy in deaf communities. With regards to social stability, deaf communities are undergoing a period of social change, with traditional centralized schools for deaf children closing, and deaf clubs having increasingly less importance. Both these factors are leading to changing patterns of language transmission. Given only a minority of signers who have ASL, BSL, or Auslan as a first language from signing deaf parents (e.g., Fischer, 1978; Mitchell and Karchmer, 2004), many deaf adults thus acquire these sign languages from other deaf children in primary or secondary school, or in early adulthood in deaf clubs. Some of these deaf adults may not have fully acquired English, and thus may have learnt these sign language varieties as delayed first languages (e.g., Emmorey, 2002). In fact, together with hearing adult second language learners of ASL, Auslan, and BSL, non-native deaf signers constitute the overwhelming majority of the signing community. Together with extensive exposure to spoken and written English, native signers are in constant contact with delayed first language and second language learners. This leads to a sociolinguistic situation that is quite unique, although with some similarities to pidgin language contact situations in which nobody is a native speaker of the variety being used to communicate across language barriers (cf. Fischer, 1978).

MORPHOLOGICAL COMPLEXITY IN VILLAGE SIGN LANGUAGES

One might predict that the relatively more dense, stable environments of some micro-community sign languages, such as Kata Kolok, might provide an environment in which complexification is more likely to emerge. We need more research

to explore this claim (see Zeshan and De Vos, 2012), but there are some possible hints in the literature. For example, we see some possible complexification in the pronoun and verb systems in Kata Kolok, where the grammar exhibits distinctions in person and aspect marking (Trudgill's criterion 4, see above). While pointing signs are used for present referents, list buoys (where signers point to fingers on their non-dominant hand, often used to refer to a list of items, cf. Liddell, 2003a) are reportedly used for absent referents (De Vos, 2012). Both pointing signs and list buoys exist in other sign languages, but studies appear to suggest the use of these systems is allocated different grammatical functions categorically in Kata Kolok. Another example might be the emergence of a mouth gesture in Kata Kolok (closed mouth opening, resembling the syllable 'pah', see **Figure 4**) which co-occurs with manual verbs to indicate perfective aspect (De Vos, 2012). This is a type of aspect marking which represents an increase in morphological complexity (a similar mouth gesture has been identified in other sign languages, although it does not appear to have the same grammatical role). Perfective aspect marking in ASL, BSL, and Auslan, however, involves the grammaticalization of a manual lexical verb sign meaning 'finish' (e.g., Johnston et al., 2015). Therefore, it may be the case that micro-community sign languages provide more dense, stable environments compared to macro-community sign languages, and it is here that we might see some emergent complexification, but more detailed investigation needs to be undertaken.

CONCLUSION AND FUTURE DIRECTIONS

In this article, we have briefly explored the idea that socio-cultural and other non-linguistic factors can contribute to linguistic diversity using Trudgill's (2011) framework of sociolinguistic typology, and we have discussed this proposal with regards to sign languages used by deaf communities for the first time. We have argued that the unique sociolinguistic situation

and language transmission patterns of sign languages may contribute as a factor (in addition to the relative youth of sign languages) to explain their relative lack of morphological complexification. This conclusion is controversial since sign languages are sometimes presented as morphologically complex languages that present a puzzle for linguistic theory when their youth is taken into consideration. However, when we apply Trudgill's notion of linguistic complexity, as we have done here, a clearer picture of the nature of sign languages and their relationship to their sociolinguistic situation emerges. If Trudgill is correct, even considerably longer histories may not lead to morphological complexification in macro-community sign languages. In future, more research needs to be carried out on the specific sociolinguistic situation of sign languages, particularly with regards to the relative impact of social network density on these languages, as well as their youth and propensity for highly iconic structures (e.g., Cuxac and Sallandre, 2007).

AUTHOR CONTRIBUTIONS

AS, JF, KC, and TJ all made substantial contributions to the conception of the work and the interpretation of data. AS led on the writing of the paper, with JF, KC, and TJ all contributing to revising it critically for intellectual content and style. AS, JF, KC, and TJ all gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

REFERENCES

Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language* 81, 301–344. doi: 10.1353/lan.2005.0043

Bayley, R., Lucas, C., and Rose, M. (2002). Phonological variation in American Sign Language: the case of 1 handshape. *Lang. Var. Change* 14, 19–53. doi: 10.1017/S0954394502141020

Bentz, C., and Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Lang. Dyn. Change* 3, 1–27.

Bergman, B., and Dahl, Ö (1994). "Ideophones in Sign Language? The place of reduplication in the tense-aspect system of Swedish sign language," in *Tense, Aspect and Action - Empirical and Theoretical Contributions to Language Typology*, eds C. Bache, H. Basboll, and C. E. Lindberg (Berlin: Mouton de Gruyter), 397–422.

Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.

Cormier, K., and Fenlon, J. (2009). "Possession in the visual-gestural modality: how possession is expressed in British sign language," in *The Expression of Possession*, ed. W. McGregor (Berlin: Mouton de Gruyter), 389–422.

Cuxac, C., and Sallandre, M. (2007). "Iconicity and arbitrariness in French sign language: highly iconic structures, degenerated iconicity and diagrammatic iconicity," in *Verbal and Signed Languages: Comparative Structures, Concepts,*

and Methodologies, eds E. Pizzuto, P. Pietrandrea, and R. Simone (Berlin: Mouton de Gruyter), 13–33.

Dahl, O. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins. doi: 10.1075/slcs.71

de Beuzeville, L., Johnston, T., and Schembri, A. (2009). The use of space with indicating verbs in Auslan: a corpus-based investigation. *Sign Lang. Linguist.* 12, 53–82. doi: 10.1075/sll.12.1.03deb

De Vos, C. (2012). *Sign-Spatiality in Kata Kolok: How a Village Sign Language in Bali Inscribes its Signing Space*. Doctoral dissertation, Radboud University, Nijmegen.

Emmorey, K. (2002). *Language, Cognition and the Brain: Insights from Sign Language Research*. Mahwah, NJ: Erlbaum.

Fenlon, J., Schembri, A., Rentelis, R., and Cormier, K. (2013). Variation in handshape and orientation in British sign language: the case of the '1' hand configuration. *Lang. Commun.* 22, 69–91. doi: 10.1016/j.langcom.2012.09.001

Fischer, S. D. (1978). "Sign language and creoles," in *Understanding Language Through Sign Language Research*, ed. P. Siple (New York, NY: Academic Press), 309–331.

Gray, M. (2013). *Aspect Marking in Australian Sign Language: A Process of Gestural Verb Modification*. Doctoral dissertation, Macquarie University, Sydney, NSW.

Janzen, T., O'Dea, B., and Shaffer, B. (2001). The construal of events: passives in American sign language. *Sign Lang. Stud.* 1, 281–310. doi: 10.1353/sls.2001.0009

Johnston, T. (1989). *Auslan: The Sign Language of the Australian Deaf Community*. Doctoral dissertation, University of Sydney, Sydney, NSW.

Johnston, T., Cresdee, D., Schembri, A., and Woll, B. (2015). Finish variation and grammaticalization in a signed language: How far down this well-trodden pathway in Auslan (Australian Sign Language)? *Lang. Var. Change* 27, 117–155. doi: 10.1017/S0954394514000209

Johnston, T., and Schembri, A. (2007). *Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511607479

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

Ladd, P., and Edwards, V. (1982). British sign language and West Indian Creole. *Sign Lang. Stud.* 35, 101–126. doi: 10.1353/sls.1982.0028

Liddell, S. K. (2003a). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615054

Liddell, S. K. (2003b). "Sources of meaning in ASL classifier predicates," in *Perspectives on Classifier Constructions in Sign Language*, ed. K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates).

Liddell, S. K. (2011). Agreement disagreements. *Theor. Linguist.* 37, 161–172. doi: 10.1515/thli.2011.012

Lillo-Martin, D., and Meier, R. P. (2011). On the linguistic status of 'agreement' in sign languages. *Theor. Linguist.* 37, 95–141. doi: 10.1080/02699200801916808

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559

McWhorter, J. H. (2007). *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195309805.001.0001

McWhorter, J. H. (2012). A review of Trudgill, P. (2011). *Rev. Crioulos Base Lex. Port. Espanhol*. 3, 44–45.

Meir, I., Assaf, I., Sandler, W., Padden, C., and Aronoff, M. (2012). The influence of community on language structure: evidence from two young sign languages. *Linguist. Var.* 12, 247–291. doi: 10.1075/lv.12.2.04mei

Miestamo, M., Sinnemäki, K., and Karlsson, F. (eds) (2008). *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins.

Mitchell, R. E., and Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in The United States. *Sign Lang. Stud.* 4, 138–163. doi: 10.1353/sls.2004.0005

Morris, C. D. (2016). *Examining the Associations Between Social Network ties and Linguistic Production*. Doctoral dissertation, Gallaudet University, Washington, DC.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. G. (1999). *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, MA: MIT Press.

Perkins, R. D. (1992). *Deixis, Grammar, and Culture*. Amsterdam: John Benjamins. doi: 10.1075/tsl.24

Pichler, D. C., Schalber, K., Hochgesang, J., Milkovic, M., Wilbur, R. B., Vulje, M., et al. (2008). "Possession and existence in three sign languages," in *Sign Languages: Spinning and Unraveling the Past, Present and Future. TISLR 9, Forty-five Papers and Three Posters from the 9th Theoretical Issues in Sign Language Research Conference*, ed. R. M. D. Quadros (Petrópolis: Editora Arara Azul), 440–458.

Sampson, G., Gil, D., and Trudgill, P. (eds). (2009). *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139163910

Schembri, A. (2010). "Documenting sign languages," in *Language Documentation and Description 7: Lectures in Language Documentation and Description*, ed. P. K. Austin (London: School of Oriental and African Studies), 105–143.

Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Zeshan, U., and De Vos, C. (2012). *Sign Languages in Village Communities: Anthropological and Linguistic Insights*. Berlin: Mouton De Gruyter. doi: 10.1515/9781614511496

Blues in Two Different Spanish-Speaking Populations

Fernando González-Perilli[1,2]*, Ignacio Rebollo[1], Alejandro Maiche[1] and Analía Arévalo[3]

[1]*Centro de Investigación Básica en Psicología, Universidad de la República, Montevideo, Uruguay,* [2]*Facultad de Información y Comunicación, Universidad de la República, Montevideo, Uruguay,* [3]*Departamento de Neurologia, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil*

Several studies investigating color discrimination across languages have shown a facilitation effect in groups that employ more than one term to refer to a given color. While Uruguayans use "*azul*" to refer to dark blue and "*celeste*" for light blue, Spaniards use "*azul*" for dark blue and the compound terms "*azul celeste*" or "*azul claro*" for light blue. In this study, Uruguayan and Spanish participants discriminated between pairs of color stimuli that lie at different distances from each other on the blue color spectrum in three different sessions: a session with no interference (basic task), one with verbal and one with visual interference. Only the Uruguayans were more accurate at distinguishing between stimuli associated with different color terms. Furthermore, while both Uruguayans and Spaniards showed a category effect in response times, the effect was strongest for Uruguayans when items were closer to each other on the color spectrum (i.e., more difficult). This study is unique in that we observed different Whorfian effects in two groups that speak the same language but differ in their use of color-specific terms. Our results contribute to the discussion of whether and to what extent language or other cultural variables affect the perception of different color categories.

Keywords: color perception, categorical perception, linguistic relativity, Sapir–Whorf hypothesis, cross-cultural cognition

INTRODUCTION

To what extent do language and/or culture affect the way we process and organize the information and experiences that make up our world? The work of Sapir, Whorf, and others sparked this famous debate at least a century ago, and these questions continue to interest academics across fields to this day (Whorf, 1956; Lucy and Shweder, 1979; Kay and Kempton, 1984; Vygotsky, 1987; Lupyan, 2012; Levelt, 2014).

Most investigations addressing this topic have been characterized as either descriptive, simply reporting interesting differences between two or more languages, or aiming to explain how observed disparities are associated with different cognitive processes (Zlatev and Blomberg, 2015). These two perspectives are also associated with weak and strong versions of the Sapir–Whorf hypothesis (language and thought are interrelated vs. language determines thought, Brown, 1976). Both hypotheses have been criticized for being trivial and non-informative (weak version) or theoretically and/or methodologically wrong (strong version) (Bloom and Keil, 2001).

Zlatev and Blomberg (2015) propose approaching each investigation according to whether the focus is on the structure of language or on its implementation (discourse). Traditional cognitive approaches focus on abstract structural aspects of language and search for innate universal features. On the other hand, linguistic relativism concentrates on how the phenomenon of categorical

perception (CP, Harnad, 2005) is affected by different contextual factors, such as language and culture.

According to Lucy (1997), there are three "logical components" that are typically taken into account when studying linguistic relativity: (1) the distinction between language and thought, (2) the mechanisms explaining the instantiation of a possible influence, and (3) the identification of other factors involved in the phenomenon.

Regarding the first point, relativists often agree with a broad definition of *thought*, not just as a conscious reflective process (as understood in folk psychology) but also involving less aware, automatic processes, such as perception and categorization. Moreover, language and perception are not understood as isolated modules—as in classic cognitivism (Pylyshyn, 1999)—but are thought to interact with a myriad of processes. Thus, the role of verbal labels affecting perception and categorization is a key issue in contemporary approaches (Thierry, 2016). How basic cognitive processes are influenced by implicit recovery of linguistic (but also contextual and sociocultural information) is another key question, which involves points 2 and 3.

Therefore, the key notion leading the research on linguistic relativity is not whether minds are dependent on a given language but how verbal labels and categories interact with cognition across different contexts (Thierry, 2016; Zhong et al., 2017). Topics currently being studied include: cross-cultural comparisons (i.e., Boroditsky, 2001; Casasanto, 2008), the exploration of categorical effects under different interference conditions (i.e., Roberson and Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007), and the time course of the effect, which informs whether perception or higher cognitive processes are involved (Mo et al., 2011; Clifford et al., 2012; He et al., 2014; Forder et al., 2017).

One line of research within this debate concerns the way in which different languages divide color space. The key question within this work is whether these varying linguistic representations affect performance on tasks that are seemingly non-linguistic. In other words, does the way in which a particular language categorizes colors affect the way its speakers think about and organize color in their minds, even in the absence of an explicitly linguistic task? One special case—that of the color *blue*—has been studied by researchers across a number of languages, including Greek (Androulaki et al., 2006; Athanasopoulos, 2009; Thierry et al., 2009), Italian (Bimler and Uusküla, 2014), Japanese (Athanasopoulos et al., 2010), Korean (Roberson et al., 2009), and Russian (Witthoft et al., 2003; Winawer et al., 2007). These languages share a common feature that distinguishes them from English: they divide the color blue into two distinct linguistic categories, one depicting lighter blues, and the other depicting darker blues. In the above studies, speakers of those languages were relatively better than English speakers at distinguishing between color samples along the blue color spectrum when the samples' names came from different linguistic categories, even though the task did not require linguistic output.
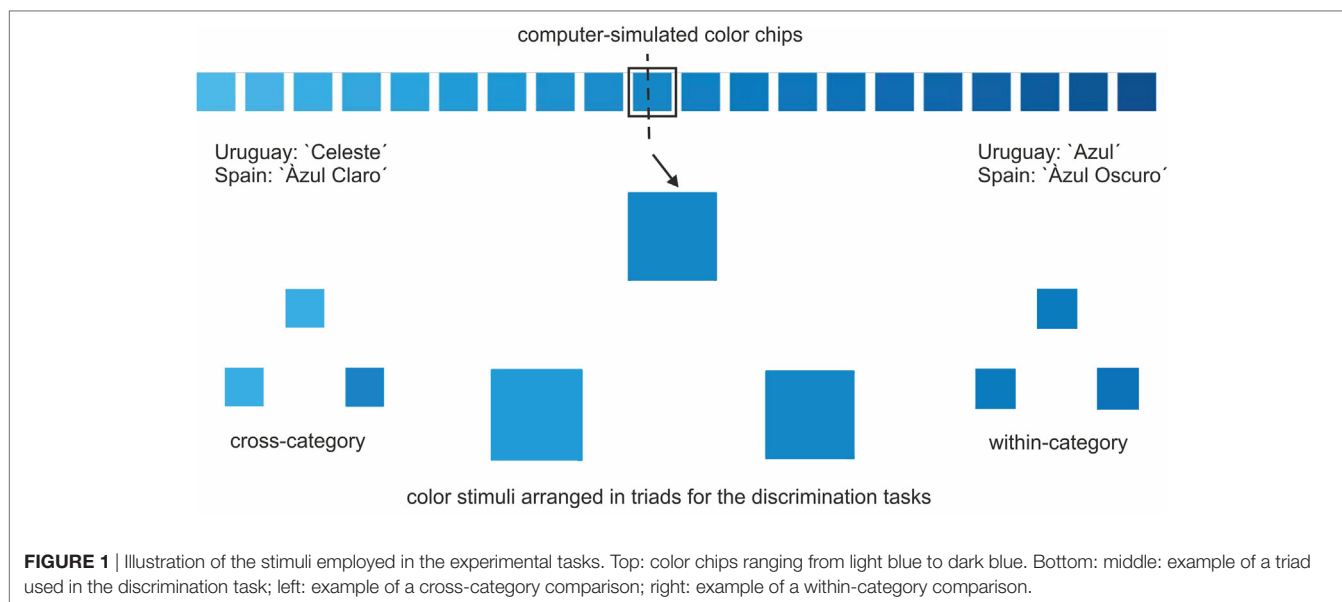
This kind of implicit linguistic effect is explained by theories arguing that linguistic labels can aid in the discrimination of stimuli that are hard to categorize (Lupyan, 2012) thanks to a predicting coding process in which "every level of the hierarchically organized system that constitutes the brain works to predict the activity in the level below" (Lupyan and Clark, 2015, p. 279). In such a predictive framework, the brain's function is to produce a percept that fits the best hypothesis regarding the state of the world that is being conceived (Lupyan and Clark, 2015). That is acquired through an interplay of top-down knowledge about the world and incoming bottom-up sensory information (Bar, 2003). In Lupyan's view, labels work as hubs of perceptual, semantic and contextual information related to specific categories. Their function is to reduce prediction error by enhancing the perception of typical categorical features. Therefore, verbal labels can be elicited to foster predictability and support cognition.

Aiming to clarify this issue, several studies include a verbal interference condition. That is, they introduce a concurrent task demanding linguistic resources (e.g., remembering a string of digits). This interference is expected to disrupt categorical effects (advantage for the discrimination of stimuli pertaining to different categories) if linguistic processes are necessary for CP to occur. For instance, Winawer et al. (2007) showed that when an additional task requiring verbal memory was included, the categorical effects found for the Russian participants vanished, suggesting linguistic resources are used by Russian speakers in this seemingly non-linguistic color perception task. The authors also presented a spatial interference condition that did not alter categorical effects, further supporting the view that the a disruption of the CP advantages was in fact due to a disruption in linguistic processing and not to the heavier cognitive load imposed by any interference task.

In the current study, we compared two groups of speakers of the same language that employ different verbal labels for the same color. This comparison is interesting because, unlike previous studies where groups of speakers spoke different languages, differences between the current groups should be much subtler, and may reflect cultural variations that affect the frequency of use of such labels.

Similarly to the languages investigated in previous studies (Androulaki et al., 2006; Winawer et al., 2007), in some variants of Spanish, the color blue is associated with two different linguistic terms: dark blues are *azul* and light blues are *celeste*. However, the Spanish language presents an interesting case, in that different populations of Spanish speakers differ in the way they implement this distinction. Namely, in some South American countries such as Uruguay, the term *celeste* (light blue) is used on its own. By contrast, in Spain, the term "*celeste*" is used as part of a compound word, i.e., *azul celeste*, making *celeste* a subcategory within the larger category of *azul*, or (regular or dark) blue. The word (and color) *celeste* also carries significant cultural weight in Uruguay, given that it is found on national emblems and by extension, national sports team uniforms. A recent study conducted by our group confirmed the use of *celeste* as a separate basic color term (BCT) for light blues in Uruguay. Thirty healthy participants were given 2 min to write down as many color names as they could remember while keeping their eyes closed (Elicited List task: Corbett and Davies, 1997). Following Berlin and Kay's (Berlin and Kay, 1969) work, one would predict that only 11 different color names would be elicited in more than 50% of the lists produced by participants. In this study, however,

FIGURE 1 | Illustration of the stimuli employed in the experimental tasks. Top: color chips ranging from light blue to dark blue. Bottom: middle: example of a triad used in the discrimination task; left: example of a cross-category comparison; right: example of a within-category comparison.

Uruguayan participants consistently produced 12 names, as they included *celeste* as its own color category. In fact, both *azul* and *celeste* were consistently found among the first BCTs reported by Uruguayans (Lillo et al., 2016).

For the current experiment, we tested Uruguayan as well as Spanish participants on a color discrimination task we designed using stimuli along the *azul-celeste* boundary. Since cultural as well as linguistic differences have been used to explain Whorfian effects across different populations, the Uruguay-Spain comparison is interesting because the two populations come from different cultures but use the same language and very similar color space partitions. That is, when asked to assign segments of the color spectrum to different color terms, Uruguayans and Spaniards coincide perfectly on all terms except for *celeste*: the space Uruguayans call "*celeste*" falls into the greater category of "*azul*" for Spaniards (Lillo et al., 2016). Given the presence of the 12th BCT for the Uruguayans, we hypothesized that this group would display a relatively stronger categorical advantage than Spaniards.

MATERIALS AND METHODS

Participants

A total of 73 individuals participated in this study: 35 were recruited from the Universitat Autònoma de Barcelona, Spain, and 38 were recruited from the Universidad de la República in Montevideo, Uruguay. All of them were native speakers of the Spanish spoken in their country, and 22 of the Spanish participants were also Catalan speakers. Nine participants (2 from Spain and 7 from Uruguay) who produced more than 25% errors and RTs < 200 and >3,000 ms were excluded from the analysis, for a final group of 33 Spaniards (mean age = 25.1, SD = 3; 18 female) and 31 Uruguayans (mean age = 22.5, SD = 3.2; 17 female). Groups did not differ significantly from each other in terms of gender or age [$F(1,62) = 0.802$, $p = 0.374$].

FIGURE 2 | Examples of far (left) and near (right) comparison triads.
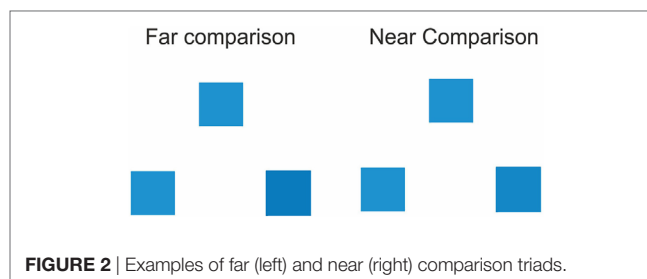
Stimuli

We created 20 computer-simulated color chips that ranged from light blue (*azul celeste* in Spain and *celeste* in Uruguay) to dark blue (*azul oscuro* in Spain and *azul* in Uruguay) (**Figure 1**). Stimuli coordinates (Commission Internationale de l'Eclairage, *Yxy*) ranged from $Y = 29.26$, $x = 0.217$, $y = 0.274$ for stimulus 1 to $Y = 4.18$, $x = 0.182$, $y = 0.167$ for stimulus 20. Stimuli varied primarily in the luminance axis (Y) and the y chromaticity axis, and were selected taking into account previous research on color categories in Spanish (Lillo et al., 2007) as well as cross-linguistic comparisons (Winawer et al., 2007; Roberson et al., 2009). The color squares measured 2.5 cm per side, and subjects viewed the screen from a distance of 60 cm. In addition, there were two categories of deviant stimuli: near and far. "Near" stimuli were colors that were two chips away from the target stimulus while "far" stimuli were four chips away (**Figure 2**). Discrimination between "near" stimuli was expected to be more difficult than between "far" stimuli.

Procedure

Prior to participation, an investigator explained the study to participants, who then signed an informed consent form. All study procedures were conducted with the approval of the Research Ethics Committee of the Department of

Psychology at University of the Republic (Uruguay) and the Department of Basic Psychology at the Autonomous University of Barcelona (a separate ethics approval was not required as per the Autonomous University of Barcelona guidelines and as per Spanish regulations) and were in accordance with the Declaration of Helsinki. Participants viewed three color squares arranged in triads (1 above and 2 below) (**Figure 1**) and were asked to decide which of the two lower squares matched the one on top. The side (right or left) on which the distractor was presented was counterbalanced across trials. Each participant completed three blocks of 136 color discrimination trials: one regular block (Basic Task), one block that also included a secondary spatial interference task, and a third block that included a verbal interference task. Half of the comparisons included "near" stimuli and half included "far" stimuli. The two interference tasks (one verbal and one spatial) were included, following Winawer et al. (2007), to test whether either type of interference affected any observed categorical effects, thus shedding light on the type of processing employed by participants during the basic task.

Interference Tasks

(a) *Spatial interference:* participants viewed a 4 × 4 square grid in which four randomly chosen squares were shaded black (**Figure 3**) and were instructed to maintain a picture of it in mind until tested. A two-choice test was presented every eight color discrimination trials.

(b) *Verbal interference:* participants were shown an eight-digit number series (**Figure 3**) for 3 s every eight color discrimination trials and were asked to rehearse it while completing the color discrimination task. Their recall was then tested by having them choose between the original series and a foil that differed by one digit.

Participants' Boundaries

Following the categorization tasks, participants also completed a *Border detection task* designed to test each individual's color boundary between dark and light blues. Participants viewed the 20 stimuli (which appeared 10 times and in random order) and pressed a key to indicate whether each color was *celeste* or *azul* (for Uruguayans) and *azul celeste* or *azul oscuro* (for Spaniards). They were asked to make all judgments as quickly and accurately as possible.

Overall, 36% of participants identified Stimulus 10 as the categorical boundary, 24% chose Stimulus 9, 20% chose Stimulus 8, 14% chose Stimulus 11, and 6% chose Stimulus 7. All Uruguayans categorized Stimulus 1 as *celeste* (light blue) and stimulus 20 as *azul* (dark blue), while all Spanish participants categorized Stimulus 1 as *azul celeste* (sky blue) or *azul claro* (light blue) and Stimulus 20 as *azul oscuro* (dark blue). Each participant's score was determined individually by using his/her color boundary to classify the color discrimination trials as either cross-category or within-category. This classification was made individually (i.e., not based on the group average).

Errors and Outliers

In order that we only analyzed data from trials in which participants were actively following the interference tasks, we systematically discarded all eight color trials preceding each incorrectly answered interference trial (5.74% of trials).

We also eliminated all trials with reaction times below 200 or above 3,000 ms (2.41% of trials across participants). RT analyses were conducted only on accurate responses (87.5%).

RESULTS

We conducted a mixed ANOVA with three within-subject factors (Distance × Interference × Category) and one between-subjects factor (country: Uruguay vs. Spain).

Accuracy

Groups did not differ in terms of overall accuracy: Uruguay (M = 86.1, SD = 0.61) vs. Spain (M = 88.3, SD = 0.63), $F(1,62) = 1.942$, $p = 0.168$, $\eta^2 = 0.030$.

There were two significant main effects: Distance, $F(1,62) = 303.109$, $p < 0.0001$, $\eta^2 = 0.830$, and Category, $F(1,62) = 5.845$, $p = 0.01$, $\eta^2 = 0.086$. When analyzed together, participants were more accurate at distinguishing between far trials (M = 0.94, SD = 0.04) than between near trials (M = 0.80, SD = 0.09), and between cross-category trials (M = 0.87, SD = 0.07) than between within-category trials (M = 0.86, SD = 0.06). There were also three significant interactions: Interference × Country, Distance × Country and, most interestingly, Category × Country.

Interference × Country, $F(1, 62) = 3.219$, $p = 0.043$, $\eta^2 = 0.049$. *Post hoc* analyses showed that the interference factor was not significant when analyzed separately for each group,

Stimuli employed in the spatial interference blocks

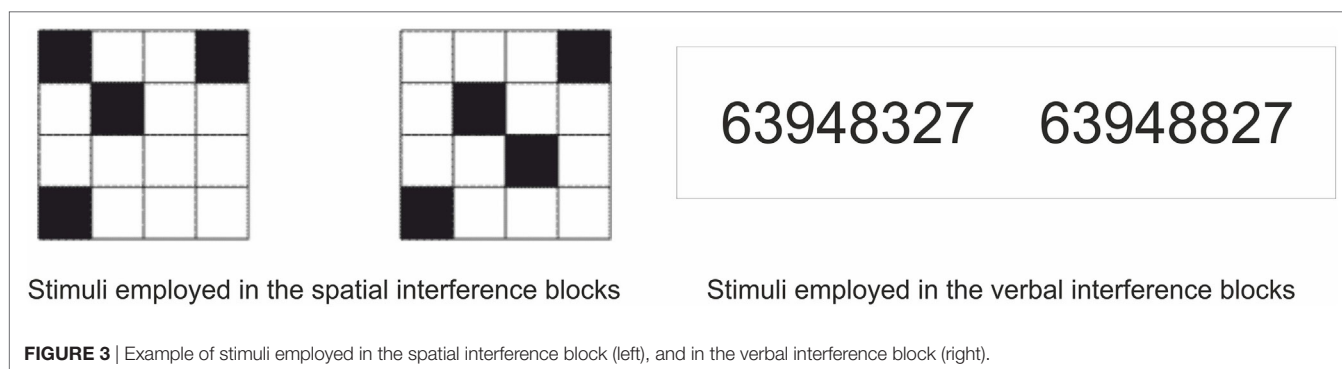Stimuli employed in the verbal interference blocks

FIGURE 3 | Example of stimuli employed in the spatial interference block (left), and in the verbal interference block (right).

and that the difference between groups was significant only in the verbal interference condition, $F_{(1,62)} = 2.304$, $p = 0.025$, $d = 0.4$.

Distance × Country, $F_{(1, 62)} = 4.252$, $p = 0.043$, $\eta^2 = 0.064$. Uruguayans had relatively greater difficulty discriminating between near stimuli (near: M = 0.78, SD = 0.13; far: M = 0.94, SD = 0.06) than did Spaniards (near: M = 0.82, SD = 0.13; far: M = 0.95, SD = 0.06).

Post hoc analyses (separate one-way ANOVAs for each group) showed that distance effects were significant for both countries, Uruguay.

$F_{(1, 30)} = 157.375$, $p < 0.0001$, $\eta^2 = 0.840$., Spain: $F_{(1, 32)} = 145.353$, $p < 0.0001$, $\eta^2 = 0.820$. Moreover, pairwise comparisons showed that neither near nor far cases showed differences between countries ($p > 0.05$).

Category × Country, $F_{(1,62)} = 2.123$, $p = 0.19$, $\eta^2 = 0.086$. Uruguayans showed an advantage for cross-category trials compared to within category trials (M = 0.87, SD = 0.07 vs. M = 0.85, SD = 0.07); *post hoc* analyses: $t_{(1,30)} = 3.268$, $p = 0.003$, $d = 0.29$. Spaniards, on the other hand, did not show this advantage (within: M = 0.88, SD = 0.06, cross: M = 0.88, SD = 0.07), $p > 0.05$ (see **Figure 4**). All other effects and interactions were not significant (all $p > 0.05$).

RT

Overall, Uruguayans were significantly slower than Spaniards, $F_{(1,62)} = 8.196$, $p = 0.006$, $\eta^2 = 0.117$ (M = 1043 ms, SD = 278 ms vs. M = 900 ms, SD = 287 ms). There were also significant main effects of Distance, $F_{(1,62)} = 267.638$, $p < 0.0001$, $\eta^2 = 0.812$, and Category, $F_{(1,62)} = 27.331$, $p < 0.0001$, $\eta^2 = 0.306$.

In line with the accuracy results, participants were faster at discriminating between far trials (M = 862 ms, SD = 175 ms) than near ones (M = 1,081 ms, SD = 235 ms), and on cross-category (M = 952 ms, SD = 206) compared to within-category trials (M = 991 ms, SD = 198 ms) (see **Figure 5**).

The first-order interaction of Interference × Country was significant, $F_{(1,61)} = 3.517$, $p = 0.033$, $\eta^2 = 0.054$. Sessions with spatial interference, in which Uruguayans performed best, resulted in the Spanish group's slowest responses (Spain: Basic: M = 889, SD = 336; Spatial: M = 941, SD = 328; Verbal: M = 869, SD = 343; Uruguay: Basic: M = 1087, SD = 347: Spatial: M = 994, SD = 342; Verbal: M = 1048, SD = 354).

Post hoc analyses showed that differences across sessions were not significant within countries, but results comparing Spain and Uruguay were different for two of the three interference conditions. Differences between groups were significant in the Basic (no interference) session, $t_{(62)} = 3.271$, $p = 0.002$, $d = 0.58$, and in the Verbal interference session, $t_{(1,62)} = 2.895$, $p = 0.005$, $d = 0.51$.

Distance × Category, $F_{(1,62)} = 3.769$, $p = 0.019$, $\eta^2 = 0.085$. A category advantage (difference between cross- and within-category trials) was stronger for far ($M_{difference} = 54$ ms) than for near color comparisons ($M_{difference} = 25$ ms).

Nevertheless, *post hoc* analyses reflected that both differences were significant: Far, $F_{(1,63)} = 3.769$, $p = 0.003$, $\eta^2 = 0.129$; Near, $F_{(1,63)} = 3.769$, $p = 0.000$, $\eta^2 = 0.257$. Additionally, categorical effects were significant at both distance conditions.
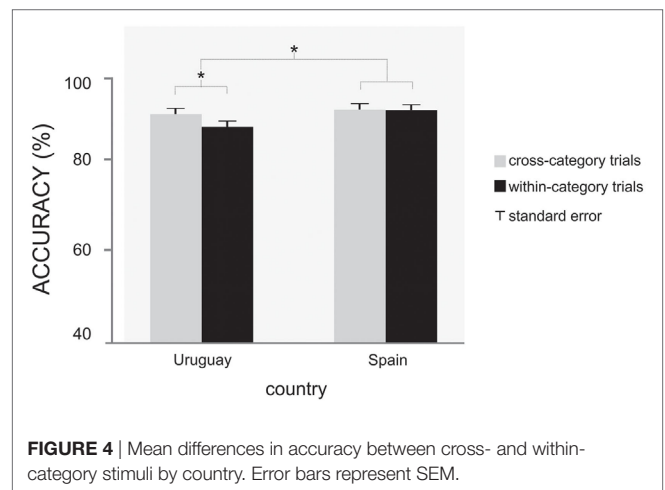

FIGURE 4 | Mean differences in accuracy between cross- and within-category stimuli by country. Error bars represent SEM.
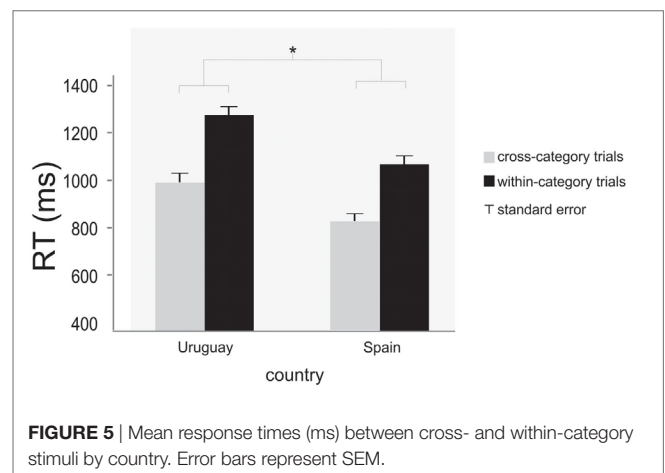

FIGURE 5 | Mean response times (ms) between cross- and within-category stimuli by country. Error bars represent SEM.

Cross-category: $F_{(1,62)} = 27.811$, $p = 0.000$, $\eta^2 = 0.310$; within-category: $F_{(1,62)} = 62.927$, $p = 0.000$, $\eta^2 = 0.504$.

While the Category × Country interaction was not significant ($p = 0.090$), the three-way Country × Distance × Category interaction was, $F_{(1,62)} = 6.596$, $p = 0.013$, $\eta^2 = 0.096$. Uruguayans showed a stronger categorical effect on near trials than on far trials.

Separate two-way ANOVAs conducted for each group showed that the interaction between distance and category was significant for Uruguayans, $F_{(1, 30)} = 11.041$, $p = 002$, $\eta^2 = 0.269$, but not for Spaniards, $F_{(1, 32)} = 0.635$, $p = 0.902$. $\eta^2 = 0.00$. For the Uruguayan group, RTs were faster for near cross-category trials than near within-category trials (M = 1112 ms, SD = 238 ms vs. M = 1193 ms, SD = 231 ms); *post hoc* analyses were significant: $t_{(1, 30)} = 5.312$, $p < 0.0001$, $d = 0.34$, while far cross-category trials did not differ significantly from far within-category trials (M = 922 ms, SD = 194 ms: vs. M = 944 ms, SD = 198 ms; *post hoc* analyses: $p > 0.05$) (see **Figures 5** and **6**).

Post hoc analyses also showed that categorical differences between countries were significant for near trials, $F_{(1,62)} = 6.852$, $p = 0.011$, $\eta^2 = 0.100$, but not for far ones, $p > 0.05$.
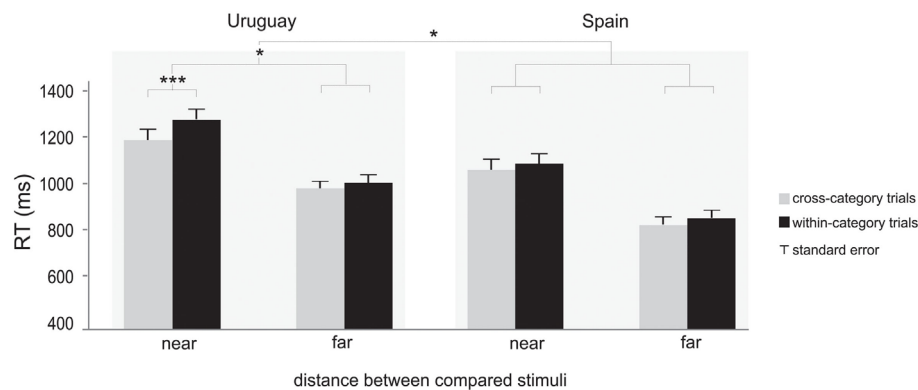
FIGURE 6 | Mean response times (ms) between cross- and within-category stimuli by country and distance. Error bars represent SEM.
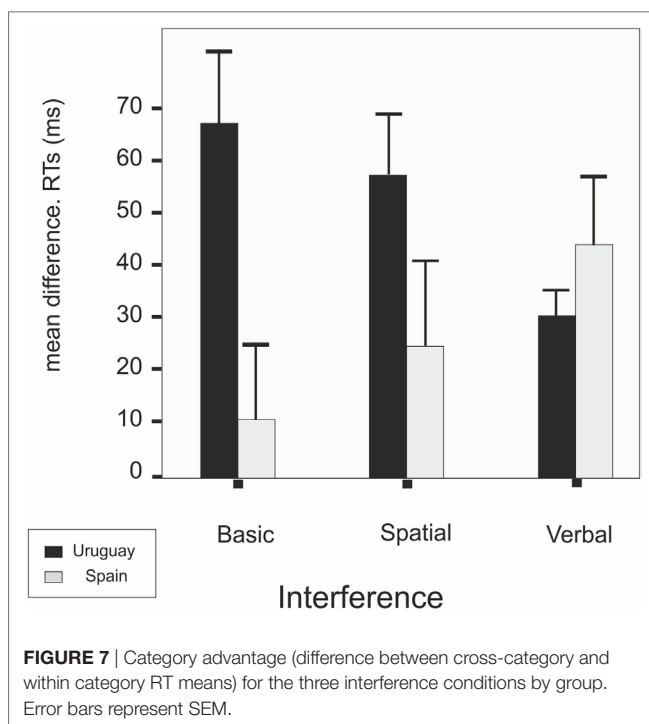

FIGURE 7 | Category advantage (difference between cross-category and within category RT means) for the three interference conditions by group. Error bars represent SEM.

Interestingly, a non-significant difference was observed for categorical effects between countries in the different interference conditions (country by category by interference, $p = 0.059$). We calculated the differences between cross- and within-category trials to obtain a categorical effect score. Categorical effect size was greater for Uruguayans (68 ms) than Spaniards (10 ms) in the basic condition [pos hoc: $F(1,62) = 6.089$, $p = 0.016$, $e = 0.089$], more similar between groups in the spatial condition [56 vs. 24; $F(1,62) = 1.513$, $p = 0.223$, $e = 0.024$] and almost equal between groups in the verbal interference condition [30 vs. 43; $F(1,62) = 0.407$, $p = 0.526$, $e = 0.007$] (see **Figure 7**).

In sum, participants were faster and more accurate when discriminating between far stimuli than near stimuli and when stimuli pertained to different categories. Uruguayans were slower than Spaniards overall, less accurate and slower in the verbal interference condition, and slower in the no interference condition. Also, Uruguayans were less accurate than Spaniards at discriminating between near stimuli. The Uruguayan group showed more categorical effects in terms of accuracy, while both groups showed stronger categorical effects for near cases in terms of RT (with Uruguayans displaying significantly stronger effects). Finally, there was a non-significant trend for differences in the effects of verbal interference on categorical effects between groups for RT.

DISCUSSION

The current study supports the Whorfian notion that language can influence color perception and is unique in that we were able to show differences in categorical effects in two groups of participants who speak the same language. Specifically, we found that Uruguayans, who have distinct color terms for light and dark blue, were more sensitive to color boundaries than Spaniards, who use a single color term for dark blue and two different compound terms for light blue. We also observed that a less frequent non-BCT—*azul celeste*—yielded some categorical facilitation. In this study, while both groups presented categorical effects in RT, the effect was strongest for Uruguayans on the more difficult "near" trials. Furthermore, only the Uruguayans were significantly more accurate at cross-category comparisons.

In contrast to previous studies where the color categories employed by the two populations clearly distinguished between dark and light blues (e.g., Russian and American participants in Winawer et al., 2007), one of the compound terms for light blue used by Spaniards (*azul celeste*) contains the monolexemic term (*celeste*) used by Uruguayans. From Lillo et al. (2016), we know that Spaniards do not consider "*celeste*" or "*azul celeste*" as a 12th BCT, as Uruguayans do, which may explain the weaker categorical effects observed among Spaniards relative to Uruguayans. Furthermore, as mentioned above, "*celeste*" is particularly salient in Uruguayan Spanish for cultural reasons, and may therefore appear more frequently for this population. According to several authors, the degree of exposure to color categories correlates with the strength of categorical effects in color discrimination tasks (Witthoft et al., 2003; Thierry et al.,

2009; Athanasopoulos et al., 2011). An interesting future study would be to test category effects with a monolexemic color term whose frequency of use differs between two populations that speak the same language.

Importantly, several studies have shown that categorical effects on perception can be elicited by newly learned categories (Zhou et al., 2010; Clifford et al., 2012). In Zhou et al. (2010), participants who learned two new categories depicting light and dark shades of blue showed a categorical advantage compared with a control group, suggesting that the introduction of a novel verbal label can affect CP.

In Winawer et al. (2007), verbal interference disrupted CP for Russian but not for English speakers, suggesting a key role of language in CP (Roberson and Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007). The results of the present study suggest that category saliency may also be affected by cultural factors.

Although the effect did not reach significance, we also observed that verbal interference diminished the categorical effect in Uruguayans and increased it in Spaniards (see **Figure 7**), which suggests CP effects are affected by linguistic input. Interestingly, Spaniards showed greater CP during the verbal interference block, suggesting the recruitment of the verbal label "*azul*" was inhibited. As shown by the *Stroop* effect (Stroop, 1935), automatic elicitation of a verbal label can interfere with color discrimination. Arguably, the discrimination between stimuli representing dark and light blues would benefit from the inhibition of the verbal label "*azul*" linked to the Spaniards' main blue category. Thus, further work is needed to clarify this issue. If replicated, it would be an unusual finding that has not been reported for English speakers in previous cross-cultural studies.

It should be noted that because part of our study was conducted in Barcelona, some of our Spanish participants also spoke Catalan, which uses "*blau cel*" as a term for light blue. We have not studied "*blau cel*" or Catalan speakers specifically, so we cannot say whether this term is more similar to any of the terms used by Spaniards in Spanish or by Uruguayans. In order to exclude this variable as a possible confound, we conducted an additional ANOVA comparing the subset of Catalan-speaking Spaniards ($n = 18$) to the non-Catalan-speaking Spaniards ($n = 15$) and found that groups did not differ on any of the variables or interactions of interest.

A recent interpretation of Whorfian effects (proposed more than 100 years ago by William James; James, 1890) is called the *Label feedback hypothesis* (Lupyan, 2008, 2012), which proposes that labels (i.e., words) are automatically recovered to solve difficult discrimination cases and are recruited unconsciously when an object is perceived in order to highlight characteristic features and thus assist in the categorization process.

Furthermore, recent studies have revealed that neural networks of color perception show strong connections between basic visual areas V1 and V4 and inferotemporal and nearby regions associated with categorization (Walsh, 1999; Roe et al., 2012; Gilbert and Li, 2013; Simanova et al., 2015; Winawer and Witthoft, 2015). Moreover, an fMRI study showed activation of language regions during color perception, supporting the notion of an interaction between higher level cognition and perceptual processes (Siok et al., 2009; Brouwer and Heeger, 2013).

In the present study, perceptual processes seemed to benefit from the words' referential attributes, but the effect differed between Spanish-speaking groups. This suggests that the interplay between categorization and perception only partially depends on a particular language's structure (Ozgen and Davies, 2002; Harnad, 2005; Lupyan et al., 2007; Collins and Olson, 2014).

An alternative interpretation is that perception could be driven by cultural—and not just linguistic—influences. In fact, cultural differences in speakers of the same language may even be the driving force behind the creation of different linguistic terms. The Emergence Hypothesis for BCTs (Kay and Maffi, 1999) proposes an explanation for how BCTs have evolved in different cultures. Kay and McDaniel (1978) suggest that derived categories are a fuzzy set of intersections among primary terms. According to this view, the emergence of a new category denoting a light shade of blue would be the result of the intersection between the blue and white categories, as Androulaki et al. (2006) proposed for Greek. Exactly why a language would add a new BCT is not clear. Casson (1997) proposed that a society's technological development will increase the importance of color as a distinguishing property of objects. Paramei (2005) and Steels and Belpaeme (2005) agree that cultural and social factors are key in the development of color lexicons. Such constraints imply that color names map onto color appearances in a culturally modal pattern (Frumkina, 1999; Jameson, 2005) and, in certain languages, could emerge as culturally basic.

Probably the main debate in linguistic relativity is whether CP occurs early on (during stimulus perception; Notman et al., 2005; Lupyan, 2012) or at the time a response is given (affecting post-perceptual processes; e.g., Pinker, 1995; Li and Gleitman, 2002). This question has been investigated using ERP, with studies showing early (Fonteneau and Davidoff, 2007; Thierry et al., 2009; Clifford et al., 2010; Mo et al., 2011; Forder et al., 2017), post perceptual (Clifford et al., 2012; He et al., 2014; Witzel and Gegenfurtner, 2016) and both effects (Holmes et al., 2009). This suggests that a strictly linguistic theory of CP is at best incomplete.

One unexpected result in the current study was that Uruguayans were both most accurate and fastest at the spatial interference block, relative to the other two blocks. One possible interpretation for this is that unlike verbal interference, spatial interference had a minimal effect on performance on a task where verbal aspects were critical, and that the added challenge resulted in higher accuracy. This would not, however, explain why that interference block would result in better accuracy than the block with no interference. We do not have enough data to answer this question at the moment but will investigate it in future studies.

Another interesting but not totally unexpected finding was that overall, Uruguayans gave slower responses than Spaniards. As observed by previous investigators, this may reflect differences in groups' experience as study participants (Witthoft et al., 2003; Winawer et al., 2007; Witzel and Gegenfurtner, 2015). In the present study, while both groups were recruited within university psychology departments, the Spanish group was generally more familiar with psychophysical experiments than the Uruguayan group. In order to ensure that categorical effects across groups were not related to overall RT, additional analyses were performed

on the subset (50%) of Uruguayans with the fastest responses. Results confirmed the trends observed for the whole group.

To conclude, color terms (both monolexemic and compound) carry different degrees of enhanced frequency and saliency within a linguistic community, which in turn depend on social, cultural, and historical factors (see Berlin and Kay, 1969; Casson, 1997; Kay and Maffi, 1999; Paramei, 2005, but also see Saunders, 2000). The present work shows that these differences can lead to different CP effects across groups that speak the same language.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of University of the Republic and Autonomous University of Barcelona ethics committees with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of the Republic ethics committee.

AUTHOR CONTRIBUTIONS

AA and FG-P conceived the study which was designed with the collaboration of IR and AM. IR and FG-P carried out the experiments and the analyses were conducted by AA, IR, and FG-P. All the authors contributed to the writing of the article.

REFERENCES

Androulaki, A., Gômez-Pestaña, N., Mitsakis, C., Jover, J. L., Coventry, K., and Davies, I. (2006). Basic colour terms in modern Greek: twelve terms including two blues. *J. Greek Linguist.* 7, 3–47. doi:10.1075/jgl.7.03and

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: the case of Greek blues*. *Bilingualism* 12, 83–95. doi:10.1017/S136672890800388X

Athanasopoulos, P., Damjanovic, L., Krajciova, A., and Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: the case of Japanese blues. *Bilingualism* 14, 9–17. doi:10.1017/S1366728909990046

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J. R., and Thierry, G. (2010). Perceptual shift in bilingualism: brain potentials reveal plasticity in pre-attentive colour perception. *Cognition* 116, 437–443. doi:10.1016/j.cognition.2010.05.016

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi:10.1162/089892903321662976

Berlin, B., and Kay, P. (1969). Basic color terms: their universality and evolution. *David Human Series Philos. Cogn. Sci. Reissues* 19, 178.

Bimler, D., and Uusküla, M. (2014). "Clothed in triple blues": sorting out the Italian blues. *J. Opt. Soc. Am. A* 31, A332–A340. doi:10.1364/JOSAA.31.00A332

Bloom, P., and Keil, F. C. (2001). Thinking through language. *Mind Lang.* 16, 351–367. doi:10.1111/1468-0017.00175

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 43, 1–22. doi:10.1006/cogp.2001.0748

Brouwer, G. J., and Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *J. Neurosci.* 33, 15454–15465. doi:10.1523/JNEUROSCI.2472-13.2013

Brown, R. (1976). Reference in memorial tribute to Eric Lenneberg. *Cognition* 4, 125–153. doi:10.1016/0010-0277(76)90001-9

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Lang. Learn.* 58, 63–79. doi:10.1111/j.1467-9922.2008.00462.x

Casson, R. W. (1997). "Color shift: evolution of English color terms from brightness to hue," in *Color Categories in Thought and Language*, eds C. L. Hardin and L. Maffi (Cambridge: Cambridge University Press), 225–239.

Clifford, A., Franklin, A., Holmes, A., Drivonikou, V. G., Özgen, E., and Davies, I. R. (2012). Neural correlates of acquired color category effects. *Brain Cogn.* 80, 126–143. doi:10.1016/j.bandc.2012.04.011

Clifford, A., Holmes, A., Davies, I. R., and Franklin, A. (2010). Color categories affect pre-attentive color perception. *Biol. Psychol.* 85, 275–282. doi:10.1016/j.biopsycho.2010.07.014

Collins, J. A., and Olson, I. R. (2014). Knowledge is power: how conceptual knowledge transforms visual cognition. *Psychon. Bull. Rev.* 29, 843–860. doi:10.3758/s13423-013-0564-3

Corbett, G. G., and Davies, I. R. L. (1997). "Establishing basic color terms: measures and techniques," in *Color Categories in Thought and Language*, eds C. L. Hardin and L. Maffi (Cambridge: Cambridge University Press), 197–223.

Fonteneau, E., and Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport* 18, 1323–1327. doi:10.1097/WNR.0b013e3282c48c33

Forder, L., He, X., and Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS ONE* 12:e0178097. doi:10.1371/journal.pone.0178097

Frumkina, R. M. (1999). What does my eye tell your mind? *Behav. Brain Sci.* 22, 951–952. doi:10.1017/S0140525X99302211

Gilbert, A. L., Regier, T., Kay, P., and Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U.S.A.* 103, 489–494. doi:10.1073/pnas.0509868103

Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi:10.1038/nrn3476

Harnad, S. (2005). "To cognize is to categorize: cognition is categorization," in *Handbook of Categorization in Cognitive Science*, eds H. Cohen and C. Lefebvre (Oxford: Elsevier Science), 19–43.

He, X., Witzel, C., Forder, L., Clifford, A., and Franklin, A. (2014). Color categories only affect post-perceptual processes when same-and different-category colors are equally discriminable. *JOSA A* 31, A322–A331. doi:10.1364/JOSAA.31.00A322

Holmes, A., Franklin, A., Clifford, A., and Davies, I. (2009). Neurophysiological evidence for categorical perception of color. *Brain Cogn.* 69, 426–434. doi:10.1016/j.bandc.2008.09.003

James, W. (1890). *The Principles of Psychology*, Vol. I. New York: Henry Holt.

Jameson, K. A. (2005). Culture and cognition: what is universal about the representation of color experience? *J. Cogn. Cult.* 5, 293–348. doi:10.1163/156853705774648527

Kay, P., and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *Am. Anthropol.* 86, 65–79. doi:10.1525/aa.1984.86.1.02a00050

Kay, P., and Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *Am. Anthropol.* 101, 1–32. doi:10.1525/aa.1999.101.4.743

Kay, P., and McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language* 54, 610–646. doi:10.2307/412789

Levelt, W. (2014). *A History of Psycholinguistics: The Pre-Chomskyan Era*. Oxford: Oxford Univeristy Press.

Li, P., and Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition* 83, 265–294. doi:10.1016/S0010-0277(02)00009-4

Lillo, J., Moreira, H., Vitini, I., and Martin, J. (2007). Locating basic Spanish colour categories in CIE L*u*v* space: identification, lightness segregation and correspondence with English equivalents. *Psicologica* 28, 21–24.

Lillo, J., Prado-León, L., Gonzalez, F., Alvaro, L., Moreira, H., and Melnikova, A. (2016). *Spanish Basic Color Categories Are 11 or 12, Depends on the Dialect, Presented at Progress in Colour Studies Conference, London, 2016.* London: University College of London.

Lucy, J., and Shweder, R. (1979). Whorf and his critics: linguistic and nonlinguistic influences on color memory. *Am. Anthropol.* 81, 581–615. doi:10.2307/675777

Lucy, J. A. (1997). Linguistic relativity. *Annu. Rev. Anthropol.* 26, 291–312. doi:10.1146/annurev.anthro.26.1.291

Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108, 566–577. doi:10.1016/j.cognition.2008.03.009

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3:54. doi:10.3389/fpsyg.2012.00054

Lupyan, G., and Clark, A. (2015). Words and the world: predictive coding and the language-perception-cognition interface. *Curr. Direct. Psychol. Sci.* 24, 279–284. doi:10.1177/0963721415570732

Lupyan, G., Rakison, D. H., and McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychol. Sci.* 18, 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x

Mo, L., Xu, G., Kay, P., and Tan, L. H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14026–14030. doi:10.1073/pnas.1111860108

Notman, L. A., Sowden, P. T., and Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition* 95, B1–B14. doi:10.1016/j.cognition.2004.07.002

Ozgen, E., and Davies, I. R. L. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol.* 131, 477–493. doi:10.1037/0096-3445.131.4.477

Paramei, G. V. (2005). Singing the Russian blues: an argument for culturally basic color terms. *Cross-Cult. Res.* 39, 10–38. doi:10.1177/1069397104267888

Pinker, S. (1995). Language acquisition. *Language* 1, 135–182.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365. doi:10.1017/S0140525X99002022

Roberson, D., and Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cognit.* 28, 977–986. doi:10.3758/BF03209345

Roberson, D., Hanley, J. R., and Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition* 112, 482–487. doi:10.1016/j.cognition.2009.06.008

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., et al. (2012). Toward a unified theory of visual area V4. *Neuron.* 74, 12–29. doi:10.1016/j.neuron.2012.03.011

Saunders, B. (2000). Revisiting basic color terms. *J. R. Anthropol. Inst.* 6, 81–99. doi:10.1002/col.5080170514

Steels, L., and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *Behav. Brain Sci.* 28, 469–488.

Simanova, I., Francken, J. C., de Lange, F. P., and Bekkering, H. (2015). Linguistic priors shape categorical perception. *Lang. Cogn. Neurosci.* 3798, 1–7. doi:10.1080/23273798.2015.1072638

Siok, W. T., Kay, P., Wang, W. S., Chan, A. H., Chen, L., Luke, K. K., et al. (2009). Language regions of brain are operative in color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8140–8145. doi:10.1073/pnas.0903627106

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643. doi:10.1037/h0054651

Thierry, G. (2016). Neurolinguistic relativity: how language flexes human perception and cognition. *Lang. Learn.* 66, 690–713. doi:10.1111/lang.12186

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., and Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4567–4570. doi:10.1073/pnas.0811155106

Vygotsky, L. S. (1987). Thinking and speech. *The Collected Works of LS Vygotsky*, Vol. 1, New York: Plenum Press, 113–114.

Walsh, V. (1999). How does the cortex construct color? *Proc. Natl. Acad. Sci. U.S.A.* 96, 13594–13596. doi:10.1073/pnas.96.24.13594

Whorf, B. L. (1956). *Language, Thought, and Reality*. Cambridge, MA: MIT Press.

Winawer, J., and Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Vis. Neurosci.* 32, E020. doi:10.1017/S0952523815000176

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785. doi:10.1073/pnas.0701644104

Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A., and Boroditsky, L. (2003). "Effects of language on color discriminability," in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 1247–1252.

Witzel, C., and Gegenfurtner, K. R. (2015). Categorical facilitation with equally discriminable colors. *J. Vis.* 15, 22. doi:10.1167/15.8.22

Witzel, C., and Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *J. Exp. Psychol.* 42, 540–570. doi:10.1037/xhp0000154

Zhong, W., Li, Y., Huang, Y., Li, H., and Mo, L. (2017). Is the lateralized categorical perception of color a situational effect of language on color perception? *Cogn. Sci.* doi:10.1111/cogs.12493

Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., and Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9974–9978. doi:10.1073/pnas.1005669107

Zlatev, J., and Blomberg, J. (2015). Language may indeed influence thought. *Front. Psychol.* 6:1631. doi:10.3389/fpsyg.2015.01631

Check for updates

Linking Adult Second Language Learning and Diachronic Change: A Cautionary Note

Vera Kempe[1]* and Patricia J. Brooks[2]

[1] *School of Social and Health Sciences, Abertay University, Dundee, United Kingdom,* [2] *College of Staten Island and The Graduate Center, City University of New York, Brooklyn, NY, United States*

Traditionally, diachronic language change has been attributed to intra-linguistic factors, which, in analogy to genetic drift, result in diversification of languages as a consequence of the social and geographical separation of linguistic communities (Lupyan and Dale, 2016). More recently, extra-linguistic factors have been implicated in language change as languages adapt to ecological niches formed by geographic, demographic, and cultural characteristics of social environments (Dale and Lupyan, 2012; Reali et al., 2018). One way of conceptualizing these extra-linguistic factors is to distinguish linguistic communities along a continuum of variation in population size, geographical spread, and amount of contact with other languages: Inward-facing, esoteric communities have small populations with shared knowledge and little language contact whereas outward-facing, exoteric communities have large populations, assembled into diverse social networks with substantial amounts of non-shared knowledge and contact with other languages (Thurston, 1987; Wray and Grace, 2007).

According to the Linguistic Niche Hypothesis (Lupyan and Dale, 2010; LNH: Dale and Lupyan, 2012), larger proportions of non-native speakers in exoteric communities promote morphological simplification of the majority language. This is thought to occur because simplifying adjustments to non-native interlocutors produced by native speakers (Little, 2011) or linguistic forms better adapted to learning constraints of adult second-language (L2) learners are adopted and transmitted to subsequent generations. Support for this hypothesis comes from qualitative (McWhorter, 2007; Trudgill, 2011) and quantitative (Szmrecsanyi and Kortmann, 2009; Lupyan and Dale, 2010; Bentz and Winter, 2013) analyses suggesting a negative correlation between the proportion of L2-learners in a linguistic community and the morphological complexity of the majority language (but see Nichols, 1992; Atkinson et al., 2016, for failures to observe this link). Below we evaluate evidence for this proposal, consider an alternative, and suggest directions for future research.

Adult language-learners differ from children in terms of prior real-world knowledge and literacy levels. Such differences allow adults to map L2s onto fully developed conceptual and linguistic representations, and may render them oblivious to aspects of morpho-syntactic structure that are not present in their L1, especially if not underpinned by awareness gained through literacy (Tarone et al., 2007). Adults and children also differ in learning mechanisms: Children rely on procedural memory whereas adults utilize declarative memory, at least in the initial stages of L2 grammar learning (Hamrick et al., 2018). Finally, relative to adults, children's cognitive limitations restrict their ability to consider contextual and referential information (Trueswell et al., 1999; Snedeker and Trueswell, 2004; Weighall, 2008). Nettle (2012) has conjectured that, as a result, children might benefit more than adults from over-specification afforded by redundant cues in complex morphological systems. Indeed, for at least one

esoteric language, Choguita Rarámuri, processing benefits have been observed from redundant morphological marking in situations where meanings of constructions are difficult to recognize (Caballero and Kapatsinski, 2015); however, evidence that benefits from over-specification are amplified in children is lacking. To test how learning and processing differences between adults and children shape morphology, the LNH has operationalized morphological complexity through estimates of the amount of morphologically-marked grammatical features and bound morphemes marking those features (Lupyan and Dale, 2010; Bentz and Winter, 2013).

Yet how strong is the evidence that children's cognitive limitations support learning of complex morphology? According to the "Less-Is-More" hypothesis (Newport, 1990), limited processing capacity focuses children's attention on smaller chunks of the input, facilitating its decomposition into sublexical units, such as inflectional affixes, and the mapping of these units onto grammatical features. Adults, in contrast, tend to process larger chunks of input, which may prevent them from noticing fine-grained variation crucial for learning inflectional morphology. Evidence for adults' limited decomposition ability has mainly been obtained from studies comparing the processing of regularly inflected vs. irregular forms (e.g., English past-tense verbs, German past participles). Evidence from priming and ERP studies suggests that native speakers rapidly decompose inflected regular forms into constituent stems and affixes, whereas adult L2-learners treat both regular and irregular forms as unanalyzed wholes (Clahsen et al., 2010), presumably because their initial reliance on declarative memory taxes cognitive resources and thus constrains the complexity of what can be learned (McDonald, 2006; Hamrick et al., 2018). Morphological complexity is also thought to impose a burden on production because it commits speakers to engage in additional "thinking for speaking," i.e., the obligatory encoding of information that may go beyond their immediate communicative intentions (Slobin, 1996, 2003). However, direct empirical support for the idea that cognitive limitations confer advantages for learning complex morphology is lacking: First, we know of no study that has directly compared children vs. adults in their tendency to decompose unfamiliar pseudo-linguistic stimuli. Second, neither connectionist models that varied memory capacity (Elman, 1993) nor experimental studies that imposed concurrent cognitive load on adult language learners (Cochran et al., 1999) yielded unequivocal and replicable evidence for superior decomposition or faster learning of morpho-syntax as a consequence of processing capacity limitations (Rohde and Plaut, 1999, 2003). There is to date no convincing evidence that cognitive limitations benefit input decomposition as an aid to morphology learning.

A related proposal attributes children's language-learning advantage to limitations in cognitive control (Thompson-Schill et al., 2009; Chrysikou et al., 2011). When exposed to artificial languages with competing variants of free morphemes distributed in unpredictable ways, children typically regularize the input by dropping less frequent variants, whereas adults tend to probability-match, i.e., to reproduce the statistical distribution of morpheme variants in the input (Hudson Kam and Newport, 2005). Such results suggest that children's inability to inhibit pre-potent responses may lead to regularization of unpredictable variation of the type encountered in pidgins. However, direct attempts to induce regularization in adults by imposing concurrent cognitive load have been unsuccessful (Perfors, 2012), suggesting that regularization is not a consequence of limitations in processing capacity and executive control, but rather a strategic response (Perfors, 2016). Additionally, while children's propensity to regularize may play an important role in creolization, it is unclear how it could facilitate morphology acquisition in natural languages, given that morphological structure has evolved to be quasi-regular and largely predictable (Kirby et al., 2015). If children were to regularize complex morphological systems, this would lead to neutralization of features and erosion of morphological contrasts—a prediction that contradicts the idea that children drive morphological complexity. Adults, on the other hand, regularize only at much higher levels of complexity, and only when variation is truly unpredictable, but not when it resembles the lexically-conditioned morphological variation of natural languages (Hudson Kam and Newport, 2009). This and other evidence that adults are quite capable of learning complex morphological systems, adopt similar learning strategies as children, and may even often outperform children in controlled experimental studies (Braine et al., 1990; Brooks et al., 1993; Wonnacott et al., 2008; Wonnacott, 2011) is difficult to reconcile with the idea that non-native speakers of a language are responsible for the erosion of its morphological complexity. Moreover, for simpler morphological patterns to become established in a language, the changes must be adopted by the next generation of L1-speakers. Although the children of non-native speakers may regularize their parents' unpredictable input, this process may not yield a less complex system, as documented in case studies of children acquiring sign language (Singleton and Newport, 2004).

Other accounts have emphasized that the morphological features of languages used by esoteric communities are idiosyncratic, low in compositionality, and replete with irregularities and formulaic expressions (Wray and Grace, 2007). Such systems arise because members of esoteric communities share a great extent of knowledge, which enables them to use contextual cues to discern utterance meanings and leave the linguistic expressions themselves more ambiguous. While this view would be compatible with the general idea of language adapting to a sociocultural niche, it is at odds with the idea that redundant marking of grammatical features by bound morphemes is the relevant characteristic of esoteric communication (Lupyan and Dale, 2010). Instead, it leads to an alternative prediction: that morphological systems acquired predominantly by children should be more idiosyncratic and less transparent than the regular, transparent, and compositional morphological systems preferred by adult L2-learners. This alternative aligns with evidence of children's propensity to learn from larger, unanalyzed chunks (Peters, 1983; Pine and Lieven, 1993)—a proposal contradicting Newport's (1990) version of the "Less-Is-More" hypothesis. Indeed, recent evidence (Arnon and Christiansen, 2017; Arnon et al., 2017) suggests that due to limited processing capacity and lack of conceptual knowledge, children may under-segment the input and form

representations of multi-word utterances along with their constituent components. Such concurrent representations enable children to harness predictive information inherent in the constituent components, which benefits learning of adjacent dependencies such as Spanish determiner-noun gender agreement (Arnon and Ramscar, 2012) or Chinese classifier-noun associations (Paul and Grüter, 2016). Even if adults form representations of multi-word utterances through chunking, their already existing conceptual knowledge may lead them to miss out on the predictive information from free morphemes contained in these utterances, focussing instead on the mapping of novel L2 content words onto existing concepts. However, while children's learning from larger, only partially decomposed multi-word utterances can explain acquisition of grammatical features marked by predictive free morphemes (e.g., determiners, prepositions), it does not explain the acquisition of bound morphemes, which are at the heart of the LNH.

A possible way to reconcile the different conceptualizations of how esoteric vs. exoteric communication affects language change is to acknowledge that transparency and complexity of morphological systems are orthogonal dimensions that jointly affect learnability, irrespective of whether instantiated by bound or free morphemes. Consider the following example: German nominal morphology comprises free (determiners) and bound (suffixes) morphemes marking number (singular, plural), gender (masculine, feminine, neuter) and case (nominative, genitive, dative, accusative), yet a considerable degree of neutralization and inflectional syncretism in its declension paradigm renders case markers fairly non-transparent and uninformative. In contrast, Russian nominal inflections are considerably more complex, with suffixes varying according to number (singular, plural), gender (masculine, feminine, neuter; with further inflectional variation for several nominal subclasses), and case (nominative, genitive, dative, accusative, instrumental, locative), yet the degree of neutralization and inflectional syncretism is substantially lower, rendering case markers more transparent and informative. If complexity is the main obstacle for adults learning nominal morphology, then L2-learners should exhibit greater difficulty with Russian than with German. If, however, lack of transparency poses the challenge, then German should be more difficult for L2-learners. Comprehension tasks comparing adult learners of Russian and German with comparable levels of L2-proficiency revealed that L2-learners of Russian processed case markers much more efficiently than L2-learners of German (Kempe and MacWhinney, 1998). When potential confounds between different L2s were controlled by manipulating features of morphological systems within

languages, native English speakers who learned Russian case inflections for transparently gender-marked nouns progressed much faster than those who learned inflections for non-transparently gender-marked nouns, even though the two subsystems were of comparable complexity (Kempe and Brooks, 2008). These findings align with evidence that learners are biased toward morphological systems that maximize communicative efficiency (Fedzechkina et al., 2012) and suggest that conceptualizations of morphological complexity need to consider the informativeness of morphemes as cues to underlying syntactic and semantic structure (Bates and MacWhinney, 1989)—an approach compatible with connectionist (Kempe and MacWhinney, 1998, 1999; Mirković et al., 2011) and information-theoretical approaches to learning and processing of inflectional morphology (Milin et al., 2009).

To provide more stringent tests of the role of child vs. adult learners as drivers of morphological change, a cognitively-grounded typology of informativeness—obtained through quantitative approaches—is needed, for example, using connectionist or deep-learning algorithms that estimate strength of association between morphological markers and thematic roles from morphologically-tagged language corpora, or inferential algorithms that operate on probability distributions of markers over thematic roles in analogy to what has been suggested for semantic typology (Kemp et al., 2018). Such estimates should be integrated with findings from cross-linguistic studies of how adults and children learn and process different morphological systems to complement existing models of learner biases in terms of exposure (Bentz and Berdicevskis, 2016) or preference for regularization (Cuskley et al., 2017), while taking into account more subtle differences in children's cognitive and pragmatic capacities. We expect especially strong insights to be gained from amplification of adult vs. child biases during transmission of language in iterated learning studies. Initial forays into this line of inquiry indicate that compositional morpho-syntax emerges more readily when systems are learned and transmitted by adults than by children (Flaherty and Kirby, 2008; Raviv and Arnon, 2016), suggesting that at present questions about how languages adapt to different learnability constraints imposed by children and adults are far from settled.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication. The authors developed the ideas jointly and collaborated in writing the article.

REFERENCES

Arnon, I., and Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Top. Cogn. Sci.* 9, 621–636. doi: 10.1111/tops.12271

Arnon, I., McCauley, S. M., and Christiansen, M. H. (2017). Digging up the building blocks of language: age-of-acquisition effects for multiword phrases. *J. Mem. Lang.* 92, 265–280. doi: 10.1016/j.jml.2016.07.004

Arnon, I., and Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition* 122, 292–305. doi: 10.1016/j.cognition.2011.10.009

Atkinson, M., Smith, K., and Kirby, S. (2016). "Adult language learning and the evolution of linguistic complexity," in *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, eds S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, and T. Verhoef (New Orleans, LA). doi: 10.17617/2.2248195

Bates, E., and MacWhinney, B. (1989). "Functionalism and the competition model," in *The Crosslinguistic Study of Sentence Processing*, eds B. MacWhinney and E. Bates (New York, NY: Cambridge University Press), 3–76.

Bentz, C., and Berdicevskis, A. (2016). "Learning pressures reduce morphological complexity: linking corpus, computational and experimental evidence," in *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)* (Osaka: 26th International Conference on Computational Linguistics), 222–232.

Bentz, C., and Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Lang. Dyn. Change* 3, 1–27. doi:10.1163/22105832-13030105

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al. (1990). Exploring language acquisition in children with a miniature artificial language: effects of item and pattern frequency, arbitrary subclasses, and correction. *J. Mem. Lang.* 29, 591–610. doi: 10.1016/0749-596X(90)90054-4

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., and Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: the contribution of phonological markers to learning. *J. Mem. Lang.* 32, 76–95. doi: 10.1006/jmla.1993.1005

Caballero, G., and Kapatsinski, V. (2015). Perceptual functionality of morphological redundancy in Choguita Rarámuri (Tarahumara). *Lang. Cogn. Neurosci.* 30, 1134–1143. doi: 10.1080/23273798.2014.940983

Chrysikou, E. G., Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2011). The other side of cognitive control: can a lack of cognitive control benefit language and cognition? *Top. Cogn. Sci.* 3, 253–256. doi: 10.1111/j.1756-8765.2011.01137.x

Clahsen, H., Felser, C., Neubauer, K., Sato, M., and Silva, R. (2010). Morphological structure in native and nonnative language processing. *Lang. Learn.* 60, 21–43. doi: 10.1111/j.1467-9922.2009.00550.x

Cochran, B. P., McDonald, J. L., and Parault, S. J. (1999). Too smart for their own good: the disadvantage of a superior processing capacity for adult language learners. *J. Mem. Lang.* 41, 30–58. doi: 10.1006/jmla.1999.2633

Cuskley, C., Castellano, C., Colaiori, F., Loreto, V., Pugliese, M., and Tria, F. (2017). The regularity game: investigating linguistic rule dynamics in a population of interacting agents. *Cognition* 159, 25–32. doi: 10.1016/j.cognition.2016.11.001

Dale, R., and Lupyan, G. (2012). Understanding the origins of morphological diversity: the linguistic niche hypothesis. *Adv. Complex Syst.* 15:1150017. doi: 10.1142/S0219525911500172

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99. doi: 10.1016/0010-0277(93)90058-4

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17897–17902. doi: 10.1073/pnas.1215776109

Flaherty, M., and Kirby, S. (2008). "Iterated language learning in children," in *Proceedings of the 7th International Conference (EVOLANG7), Vol. 12*, eds A. D. M. Smith, K. Smith, and R. Ferrer i Cancho (Singapore: World Scientific Publishing Co.). doi: 10.1142/9789812776129_0064

Hamrick, P., Lum, J. A. G., and Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1487–1492. doi: 10.1073/pnas.1713975115

Hudson Kam, C. L., and Newport, E. L. (2005). Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Lang. Learn. Dev.* 1, 151–195. doi: 10.1207/s15473341lld0102_3

Hudson Kam, C. L., and Newport, E. L. (2009). Getting it right by getting it wrong: when learners change languages. *Cogn. Psychol.* 59, 30–66. doi: 10.1016/j.cogpsych.2009.01.001

Kemp, C., Xu, Y., and Regier, T. (2018). Semantic typology and efficient communication. *Ann. Rev. Linguist.* 4, 109–128. doi: 10.1146/annurev-linguistics-011817-045406

Kempe, V., and Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Lang. Learn.* 58, 703–746. doi: 10.1111/j.1467-9922.2008.00477.x

Kempe, V., and MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Stud. Second Lang. Acquis.* 20, 543–587. doi: 10.1017/S0272263198004045

Kempe, V., and MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Lang. Cogn. Process.* 14, 129–171. doi: 10.1080/016909699386329

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Little, H. (2011). *The Role of Foreigner Directed Speech in the Cultural Transmission of Language and the Resulting Effects on Language Typology*. MSc. thesis, University of Edinburgh.

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5:e8559. doi: 10.1371/journal.pone.0008559

Lupyan, G., and Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cogn. Sci.* 20, 649–660. doi: 10.1016/j.tics.2016.07.005

McDonald, J. L. (2006). Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *J. Mem. Lang.* 55, 381–401. doi: 10.1016/j.jml.2006.06.006

McWhorter, J. (2007). *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars.* Oxford, UK: Oxford University Press.

Milin, P., Đurđević, D. F., and del Prado Martín, F.M. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: evidence from Serbian. *J. Mem. Lang.* 60, 50–64. doi: 10.1016/j.jml.2008.08.007

Mirković, J., Seidenberg, M. S., and Joanisse, M. F. (2011). Rules versus statistics: insights from a highly inflected language. *Cogn. Sci.* 35, 638–681. doi: 10.1111/j.1551-6709.2011.01174.x

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philos. Trans. R. Soc. B* 367, 1829–1836. doi: 10.1098/rstb.2011.0216

Newport, E. L. (1990). Maturational constraints on language learning. *Cogn. Sci.* 14, 11–28. doi: 10.1207/s15516709cog1401_2

Nichols, J. (1992). *Linguistic Diversity in Space and Time.* Chicago, IL: University of Chicago Press.

Paul, J. Z., and Grüter, T. (2016). Blocking effects in the learning of Chinese classifiers. *Lang. Learn.* 66, 972–999. doi: 10.1111/lang.12197

Perfors, A. (2012). When do memory limitations lead to regularization? An experimental and computational investigation. *J. Mem. Lang.* 67, 486–506. doi: 10.1016/j.jml.2012.07.009

Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Lang. Learn. Dev.* 12, 138–155. doi: 10.1080/15475441.2015.1052449

Peters, A. M. (1983). *Units of Language Acquisition.* Cambridge, UK: Cambridge University Press.

Pine, J., and Lieven, E. V. (1993). Reanalysing rote-learned phrases: individual differences in the transition to multi-word speech. *J. Child Lang.* 20, 551–571.

Raviv, L., and Arnon, I. (2016). "Language evolution in the lab: the case of child learners," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, eds A. Papafragou, D. Grodner, D. Mirman, and J. C. Trueswell (Austin, TX: Cognitive Science Society), 1643–1648.

Reali, F., Chater, N., and Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: how population size affects language. *Proc. R. Soc. B* 285:20172586. doi: 10.1098/rspb.2017.2586

Rohde, D. L., and Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition* 72, 67–109. doi: 10.1016/S0010-0277(99)00031-1

Rohde, D. L., and Plaut, D. C. (2003). "Less is less in language acquisition," in *Connectionist Modelling of Cognitive Development*, ed P. Quinlin (Hove: Psychology Press), 189–231.

Singleton, J. L., and Newport, E. L. (2004). When learners surpass their models: the acquisition of American Sign Language from inconsistent input. *Cogn. Psychol.* 49, 370–407. doi: 10.1016/j.cogpsych.2004.05.001

Slobin, D. I. (1996). "From "thought and language" to "thinking for speaking"", in *Studies in the Social and Cultural Foundations of Language, No. 17. Rethinking Linguistic Relativity*, eds J. J. Gumperz and S. C. Levinson (New York, NY: Cambridge University Press), 70–96.

Slobin, D. I. (2003). "Language and thought online: cognitive consequences of linguistic relativity," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 157–191.

Snedeker, J., and Trueswell, J. C. (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cogn. Psychol.* 49, 238–299. doi: 10.1016/j.cogpsych.2004.03.001

Szmrecsanyi, B., and Kortmann, B. (2009). The morphosyntax of varieties of English worldwide: a quantitative perspective. *Lingua* 119, 1643–1663. doi: 10.1016/j.lingua.2007.09.016

Tarone, E., Bigelow, M., and Hansen, K. (2007). "The impact of alphabetic print literacy level on oral second language acquisition," in *Low-Educated Second Language and Literacy Acquisition: Research, Policy, and Practice, Proceedings of the Second Annual Forum*, ed N. Faux (Richmond, VA: The Literacy Institute at Virginia Commonwealth University), 99–122. Available online at: http://www.leslla.org/files/resources/LESLLAProceedingsFinal.pdf

Thompson-Schill, S. L., Ramscar, M., and Chrysikou, E. G. (2009). Cognition without control: when a little frontal lobe goes a long way. *Curr. Dir. Psychol. Sci.* 18, 259–263. doi: 10.1111/j.1467-8721.2009.01648.x

Thurston, W. R. (1987). *Processes of Change in the Languages of North-Western New Britain (No. 99)*. Department of Linguistics, Research School of Pacific Studies. The Australian National University.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition* 73, 89–134. doi: 10.1016/S0010-0277(99)00032-3

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford, UK: Oxford University Press.

Weighall, A. R. (2008). The kindergarten path effect revisited: children's use of context in processing structural ambiguities. *J. Exp. Child Psychol.* 99, 75–95. doi: 10.1016/j.jecp.2007.10.004

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: an artificial language study with child learners. *J. Mem. Lang.* 65, 1–14. doi: 10.1016/j.jml.2011.03.001

Wonnacott, E., Newport, E. L., and Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cogn. Psychol.* 56, 165–209. doi: 10.1016/j.cogpsych.2007.04.002

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Recursive Combination Has Adaptability in Diversifiability of Production and Material Culture

Genta Toya* and Takashi Hashimoto

School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

It has been suggested that hierarchically structured symbols, a remarkable feature of human language, are produced via the operation of recursive combination. Recursive combination is frequently observed in human behavior, not only in language but also in action sequences, mind-reading, technology, etc. in contrast, it is rarely observed in animals. Why is it that only humans use this operation? What is the adaptability of recursive combination? We aim (1) to identify the environmental feature(s) in which recursive combination is effective for survival and reproduction, and that has facilitated the evolution of this ability, and (2) to demonstrate the possible evolutionary processes of recursive combination. To achieve this, we constructed an evolutionary simulation of agents that generated products using recursive combination and used the results to explore the types of fitness functions (that reflect the kinds of adaptive environments) that give rise to this ability. We identified two types of adaptability of the recursive combination: (1) diversifiability of production and (2) diversifiability of products. Through the former, recursive combination promotes robustness against failure of production caused by inaccurate manipulations or irreversible changes. In an environment in which diversified products are preferable, sharing a portion of the production process for these products entails producing multiple products in which recursive combination plays a key role. We suppose that recursive combination works as a driving force of material culture. Finally, we discuss the possible evolutionary scenarios of recursive combination that is later generalized to encompass many aspects of human cognition, including human language.

Keywords: recursive combination, hierarchical structure, evolutionary simulation, action grammar, evolutionary linguistics, tool manufacturing

INTRODUCTION

One of the most remarkable features of human language is its hierarchically embedded structure (Chomsky, 1957). Although both animal calls and human languages use one-dimensional sound signals in communication, words are organized hierarchically into sentences in the latter unlike in the former (Hauser et al., 2002). This feature recognizes the fact that the meaning of a sentence depends on its hierarchical structure and not on word order alone (**Figure 1**). This structural dependency may cause misunderstandings in communication, since the structure determining the meaning is not expressed unambiguously in a linear word sequence but only via interpretation (involving selections from multiple possibilities inside the speaker's and the listener's minds). If the adaptability of language contributes to information transmission and mutual understanding, for

example, to promote cooperation in a group, structural dependency will cause a disadvantage. We need to consider the adaptive value of language equipped with hierarchically embedded structures and structural dependencies in the period of the language's origin.

An important perspective was proposed by Kirby (2017) that cultural effects have a stronger impact than biological effects on the origin of linguistic structure. Kirby claims that human behaviors developed rich systematic structure such as recursion, compositionality and hierarchical structure to be expressive. Although we agree with this claim, we need to clarify the origin of linguistic ability, that is, operation, to construct rich structures.

Studies on hierarchical structure as the fundamental aspect of language (from the perspective of generative grammar) assume that recursive combination capacity, defined as the capacity to combine two items into a set, is the most important ability required for constructing hierarchical structures. This capacity is applied to enable a recursive syntactic operation. The different hierarchical structures are created by two types of combination, recursive combination and non-recursive combination. On one hand, theoretical linguists suggest that *Merge* (**Figure 2**) is a set-formation operation that can be used to create an unbounded number of sentences through its recursive application (recursive Merge) (Chomsky, 1993, 2013; Everaert et al., 2015). On the other hand, *Unification* is also a set-formation operation that has been proposed by other researchers (Jackendoff, 2002, 2011). Jackendoff claims that *Recursion* is found everywhere in higher cognition; therefore, operations such as unification that can be applied to language expression and to other mental structures are needed. The important point is that both Merge and Unification share recursive combination as the core of the operations.

Recursive combination of language has been hypothesized as a human-unique trait (Hauser et al., 2002; Fitch and Hauser,

FIGURE 1 | Structure dependency, which may cause misunderstanding in communication.

FIGURE 2 | Syntactic operation for making sentences referred to as *Merge*.

2004; Fujita, 2009, 2016). How did recursive combination and structure-dependency originate? Structure-dependency greatly increases the capacity for ambiguity in language communication. Thus, it is unreasonable to assume that recursive combination evolved to meet the needs of simple communication with one-to-one mapping between meanings and forms.

Hierarchical structure and recursive combination are described in other domains as follows:

- In action, the complex action sequences are represented as a hierarchical structure including the final goal and subordinate goals (Jackendoff, 2011). This structure is made up of three parts: the HEAD, the PREPARATION, and the CODA. For instance, when we use an automatic drip coffee maker, the action sequence "put coffee in machine" is made up of smaller steps, such as preparing the filter (preparation), putting the coffee in (head), and closing the filter (coda). "Putting the coffee in" also includes steps such as getting the coffee can out of the freezer (preparation), measuring the coffee (head), and putting the coffee away (coda), and each of these steps can be broken down further.

- In technology, new technology is constantly derived from components that already exist; in turn, these new technologies offer themselves as possible components for the construction of further new technologies (Arthur and Polak, 2006; Arthur, 2009). Technology, which is the collection of mechanical devices and methods available to a culture, becomes assimilated and combined. Therefore, this self-production is the combination of combined objects.

- In music, "discrete structural elements such as beats are hierarchically combined into larger groups according to rules" (Asano and Boeckx, 2015, p. 2). Musical metrical structures have a hierarchical structure that include the recursive embedding of beats into beats. The existence of hierarchical structure in meter has received support from neurological evidence (Bouwer et al., 2014).

- In the Theory of Mind, intentionality can be correlated to hierarchical structure with recursive embedding. Intentional states provide a natural platform for communication through mentalizing capacity. In this way, intentionality forms a naturally reflexive hierarchy (i.e., I suppose that you intend that I believe that you want me to understand that …). Representing another person's mental state is thus inherently recursive (Oesch and Dunbar, 2017).

These frameworks indicate that human behaviors and mental or physical structures can be treated as combinatorial objects. Boeckx (2017) claims that the neural basis of recursion is realized from the pairing of the fronto-parietal and fronto-temporal networks. He takes it that although both networks may be of the finite-state variety, pairing two finite-state devices could have the effect of boosting computational possibilities. Instead of operating on one-dimensional sequences, one now operates on two-dimensional tree representations. The fronto-parietal network may have the role of the global workspace as proposed by Dehaene et al. (1998). The global workspace is inherently hierarchical: It sits on top of modular networks of other cognitive domains and acts as a chunking device in a sequence producer.
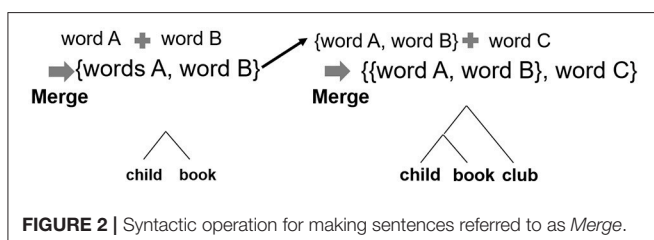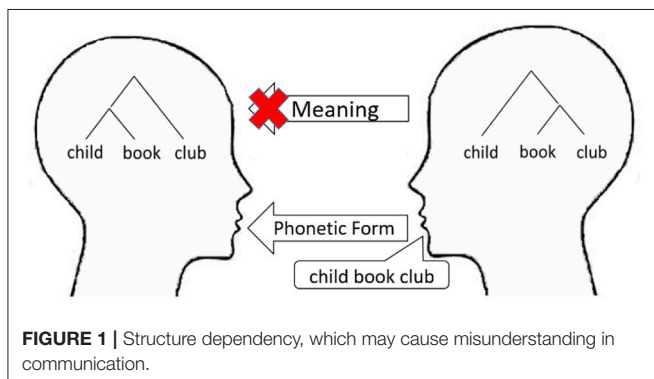
If this device is to be integrated with another sequencing machine, sequences of sequences would naturally emerge. Once this network is established, a variety of cognitive domains co-opt and account for other aspects of human-specific cognition (Boeckx, 2017). In this paper, the domain-general characteristics of discrete object combination and recursion are focused upon these.

When do object recognition and its recursive manipulation advance? Recursive combination has been observed in the object manipulation of animals and has also been researched in a cup-combining experiment with human infant participants (Greenfield et al., 1972). Greenfield posited the notion of a *grammar of action,* or in other words, a set of syntactic rules for behaviors such as object manipulation. Sequential behaviors are classified into two strategies in the framework of action grammar: the *pot strategy* and the *sub-assembly strategy*, visualized below (**Figure 3**) with the manipulation of cups used to illustrate object manipulation.

- Pot strategy: Repeated combination. Multiple active objects act on a single static object.
- Sub-assembly strategy: Recursive combination. Two objects are combined into a pair, which is then manipulated as a single unit in the combination.

It has been noted that sub-assembly strategy or other equally similar behavior such as tool-making is rarely observed in animal behavior (Greenfield, 1991; Conway and Christiansen, 2001). Therefore, it is assumed to be a precursor of recursive combination in syntax (Maynard Smith and Szathmáry, 1995; Fujita, 2009, 2016). Although Chomsky and Berwick (2015) insist that the recursive combination, Merge, abruptly appeared at some time in human evolution, we assume that it was a gradual evolutionary scenario. We further presume that object manipulation of a physical entity was a pre-adaptation to recursive syntactic operation, and the target of manipulation was qualitatively generalized. This is a reasonable hypothesis derived from the following evidence in addition to the results of comparative cognitive experiments and analysis in archeology:

- Children develop the ability to perform sub-assembly strategy and hierarchical language structure almost simultaneously (Greenfield et al., 1972). Further, the order of development in object manipulation is similar to syntactic operation. When developing object manipulation, children acquire pot strategy first and then acquire sub-assembly strategy at a later stage. During the development of syntactic operation, they learn to combine two words to form a higher order grammatical relation as an early operation. Later, children discover the grammatical complexity of adjectives and nouns that combine to form a superordinate noun phrase that enters a still higher order combination with a verb (Brown, 1973; Greenfield, 1991).
- Archeological evidence of the use of recursive combination to make stone tools from 0.28 million years ago (mya) encourages this hypothesis (Moore, 2010, 2011). Moore (2011) compares the production methods of stone tools of the Oldowan and Acheulian types and points out that the difference lies in the

FIGURE 3 | Two strategies of action grammar (adapted from Greenfield et al., 1972).
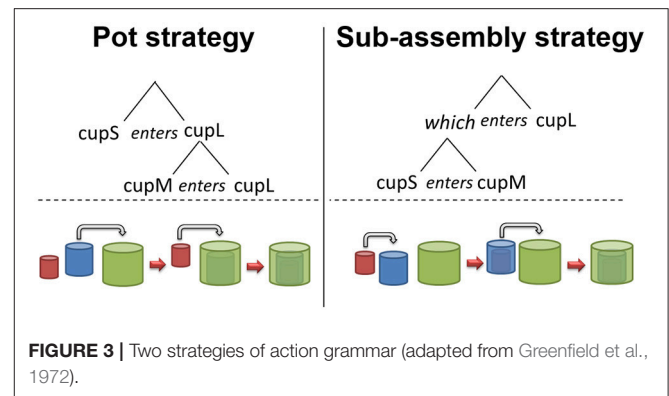
hierarchical structure of the action sequences. The Oldowan tool is generally produced by making stone flakes from a stone core. Making stone flakes from a stone core is called the flaking process. The Acheulian tool is produced by shaping a large stone flake in combination with this flaking process. This production method reflects hierarchically organized higher order intention and suggests that recursive combination of action sequences is followed. In addition, Stout (2011) illustrates stone tool-making using tree diagram. Stout shows that hominins used recursive combination in a production sequence with sub-goals when making stone tools. These are dated earlier than the appearance of symbolic behavior in human evolution (Mithen, 1996). It suggests that the recursive combination of objects pre-dated the recursive combination of lexical items.

These findings suggest that humans might have acquired recursive combination (that is a different evolutionary effect of language on communication such as sharing information) through an action sequencing process such as tool-making. Henceforth, we term the pot and sub-assembly strategies *non-recursive combination* and *recursive combination*, respectively.

The hypothesis that social recognition and population size cause recursive mental structure is reasonable because it assumes an evolutionary continuity carried over from non-human animals (Dunbar, 2009; Oesch and Dunbar, 2017). According to this hypothesis, recursive thinking became the necessary cognitive scaffolding. Dunbar claims that recursion in the language structure is boot-strapped by a primitive mentalizing ability as evidenced by an experiment that investigated correlation between recursive syntax and intentionality. However, it must be noted that *recursion*, which is often assumed to be the subordinate clause in a sentence is not equal to "recursive combination" in this paper. Recursive combination means "combination of combined objects," thus this interpretation of recursion can also be applied to mental object manipulation like mind-reading. We will elaborate on this point later in the Discussion section.

It is most important that we answer the following questions. What is the evolutionary process of recursive combination? What does the adaptability of the recursive combination consist of, if the process is adaptive evolution?

According to Tinbergen (1963), adaptability (which is effectiveness in survival and reproduction) is an important aspect used to explain the characteristics of animals. Although the adaptability of the human ability of recursive combination has been investigated in comparative cognitive science; similar traits have not yet been discovered. Furthermore, the phenomenon of evolution can only be observed in living things that have rapid generation alternation. The evolution of higher cognitive ability is not that easily studied. This problem can be solved partially by using simulations (Hashimoto, 2001). The advantage of simulation is that it allows the elaboration of hypotheses and the consideration of evolutionary processes. This is enabled by repeating the experiments in a constructive environment on a phenomenon that is difficult to observe empirically. It is not possible to prove a hypothesis solely by using this method. However, we can explain the process of the generation of a system (in this research, capacity of agent and ecological environment) causing a specific phenomenon (the evolution of recursive combination) by reproducing the phenomenon by implementing and operating a model derived from the hypothesis.

In this paper, we study the evolutionary process and adaptability of recursive combination using evolutionary simulations. The objectives are (1) to demonstrate the conditions in which recursive combination could have evolved, and (2) the possible evolutionary processes by which recursive combination could have evolved. We will claim that recursive combination has two adaptabilities; the diversifiability of production methods that promotes the secure manufacturing of the target product and the diversifiability of products by the reuse of parts of manufacturing processes that are already acquired. Two factors promote these adaptabilities: (1) extending the time available for making products, and (2) decreasing the cost of object manipulation. As a possible evolutionary process, it is necessary to increase the opportunity for production and reduce the manipulation cost before the evolution of recursive combination.

The rest of this paper is organized as follows: (1) The simulation model to examine whether agents evolve to be capable of recursive combination is described in section Materials and Methods. (2) The simulation results and resulting considerations for the model are presented in section Results. (3) A discussion based on the simulation results in consideration with other results is delivered in section Discussion. (4) The conclusion is delivered in section Conclusion.

MATERIALS AND METHODS

In this section, (1) the concepts and mechanisms of genetic algorithm (GA) and evolutionary simulation are introduced; (2) the model of object manipulation used in this paper is explained; (3) we describe how recursive and non-recursive combinations are modeled; (4) to illustrate evolutionary simulation of object manipulation, we describe the encoding of a state transition table onto a gene and also the simulation flow; and finally, (5) three fitness functions in the evolutionary simulation are posited.

Evolutionary Simulation for Investigating Adaptability

Evolution has three basic factors; (1) *Variation* meaning that there are groups with different traits. (2) *Selection* meaning that variation causes differences of survival probability depending on the environment. (3) *Inheritance* meaning that the traits aiding in the survival of individuals will be passed on to the next generation. These mechanisms can be written as a sequential procedure that is the genetic algorithm.

The genetic algorithm is constructed from the following processes:

- Generation of population (variation): Generate individuals having different genes representing different traits.
- Evaluation depending on fitness function (selection): Evaluate the genes and give them fitness values according to fitness function. The fitness function is formed and abstracted from the ecological environment.
- Reproduction with crossover and mutation (Inheritance with modification): Pass on the genes of individuals with a high fitness value to the next generation. Genes in the next generation are modified by the process of gene crossover in the parents, and mutation.

If the fitness function is presented as a problem, then the genes are the optimized solution to this problem by a cumulative process.

Typically, a genetic algorithm is used to search for (quasi-)optimal solutions according to a fitness function representing an optimization problem. However, we intend to identify fitness functions having recursive combination (as an abstract operation) as their solution. Therefore, we define the candidates for the fitness functions by considering the ecological meanings of recursive combination, i.e., the evolutionary processes and adaptability are examined by evolutionary simulations. It is not our intention to model biological evolution directly, and this simulation does not reproduce the process of human evolution.

Model of the Object Combination Operation
Abstraction of Recursive Combination and Non-recursive Combination

Prior to designing the model, we considered the computational difference between recursive combination and non-recursive combination. The crafting of a stone spear from diverse materials such as wood for the shaft, a chiseled stone for the head and adhesives used to bind everything together is a good example. Such tools had been made in 0.2 mya (Wymer, 1984). When non-recursive combination is performed, one object is combined repeatedly, i.e., the builder attaches the base of the stone edge to the wooden shaft, and fixes it using an adhesive. Thus, this operation needs both a finite set of states that is expressed as an object and a transition function that is expressed as a combination. When recursive combination is performed, combined objects combine to form another object, i.e., the builder attaches the base of the stone edge to the part where the

adhesive was applied beforehand on the wooden shaft. Therefore, this operation needs two finite sets of states (the state for combining and the state for storing) and the transition functions that are expressed as storing and retrieving.

Agent Performing Object Manipulation

An agent performing object manipulation to manufacture products is modeled using an automaton with a stack. The aim of the agent is to make products by combining the objects (hereinafter, an elemental object is represented by a letter such as *A* or *B* and a combined object by concatenating letters, such as *AB* or *ABC*). An agent is equipped with a workspace in which objects are combined and a stack in which objects are stored temporarily from the workspace. The objects correspond to the cups in the experiments of Greenfield (1991) and Matsuzawa (1986); two or more objects cannot exist in the workspace simultaneously, and this is true for the stack as well. There are any number of objects of the same type in a set of elemental objects; thus, it is possible to make a product including multiple instances of the same type of object, such as *AAB* or *AAA*. Once combined, the objects are treated as one object and cannot be separated into two objects.

In this simulation, in order to clarify the difference between recursive combination and non-recursive combination, both combinatorial operations can produce the same set of objects by assuming that a combined object has a linear structure with directionality. Therefore, an object is added at the end of another (elemental or combined) object.

The agent performs the following four actions, depending on the state of its workspace and stack:

- Get: Combine an elemental object at the end of an object in the workspace (when there is an object in the workspace) or place an elemental object in the workspace (when the workspace is empty). The elemental object is given randomly from the set of elemental objects.
- Stop: Designate an object in the workspace as a finished product. This can be executed only when the stack is empty.
- Push: Store an object in the stack; the workspace becomes empty. For simplicity, this can be executed only when the stack is empty.
- Pop: Retrieve an object from the stack. The retrieved object is combined at the end of an object in the workspace (when there is an object in the workspace) or the retrieved object is placed in the workspace (when the workspace is empty). The stack becomes empty after this action.

If multiple actions are possible in a state, one action is randomly chosen.

The initial state for the agent features an empty workspace and stack. Product-making is the process of state transitions of combined objects from the initial state to the final state. If there is an object in the stack, the agent is accepted as being in the process of production; the stack must be empty at the final state. There are *k* types of elemental objects, and an agent can make products composed of any number of elemental objects up to the maximum length, *l*, hereinafter, the maximum length of the product. The two combining actions, Get and Pop, are limited to avoid producing a combined object longer than *l*. If an agent cannot perform the Stop action when the length of the combined objects in the workspace becomes *l*, this production process is a failure, and a new production process begins from the initial state. An agent can make any number of products within the upper limit of the number of manipulation steps, which sets the agent's lifetime.

In this model, two strategies, non-recursive combination and recursive combination, are formalized, respectively, as follows:

- *Non-recursive Combination*: An agent combines an elemental object with another elemental object in the workspace. The stack is not used or stores an uncombined object only.
- *Recursive Combination*: An agent combines an elemental or combined object with a combined object that has been combined in advance and stored in the stack. Stack operations to store and retrieve the combined object are necessary.

Note that the following operations are not recursive combinations:

- Pop an elemental object stored in the stack to combine with an elemental or combined object in the workspace.
- Pop a combined object stored in the stack to place it in the empty workspace.

State Transition Table

A state transition, effected by performing an action, is expressed as:

$$\left(\text{stack, workspace}\right) \overset{\text{action}}{\rightarrow} \left(\text{stack}', \text{ workspace}'\right). \quad (1)$$

The behavior of a particular agent is defined by the state transition table shown in **Figure 4**. The state transition table describes a transition of a finite number of states, in our paper, workspace and stack, of the agents. In **Figure 4**, the two columns on the left are the state of the stack and of the workspace, that is, the left-hand side of (1). The five columns on the right are the actions. The destination of the transition after each action, corresponding to the right-hand side of (1), is indicated in each box as the states of the stack and the workspace. The symbol "ε" signifies nothing in the stack or in the workspace; that is, it represents an empty state. An instance of "–" indicates that the agent cannot perform this transition, while "n/a" indicates that the transition is forbidden due to a non-empty stack. If more than one destination is provided, one is selected randomly. Both the number of workspace states and stack states are

$$1 + S,$$

where

$$S = \sum_{l'=1}^{l} k^{l'}$$

is the size of the combinatorial space, and the number of actions is $(k + 3)$. The number of *n/a*'s is

$$2S(1 + S).$$

Therefore, the total size of the state transition table is

$$(k+3)(1+S)^2 - 2S(1+S) = (1+S)\{(1+k)+k+3\}.$$

Figure 5 provides examples of state transitions corresponding to the state transition table in **Figure 4**. For an example of the state transitions, when an agent has states where the workspace is ε and the stack is also ε (as seen in columns 1 and 2, row 1 in **Figure 4**; as at the top of **Figure 5**), if the agent performs Get A, the agent will have a state where the workspace is A and the stack is ε (as seen in columns 1 and 2, row of workspace 2 and stack 1 in **Figure 4**; as seen at the left top of **Figure 5**). Then, if the agent performs Pop, the agent will have states where the workspace is ε and the stack is A (as seen in columns 1 and 2, row of workspace 1 and stack 2 in **Figure 4**; as seen under the top left of **Figure 5**). The same product can be manufactured either by using or by not using stacks, but production using stacks require more steps than the latter process.

Model for Evolutionary Simulation
Gene Encoding of Transition Table
The state transition table of the agent is encoded into a gene with a binary string, as shown in **Figure 4**. If a transition is possible, the corresponding box in the state transition table is filled; in such a case, the locus is one. If a transition is impossible, the box is "–" then the locus is zero. Boxes showing "n/a" are not encoded into a gene. There is a regulatory locus for stacks. If it is zero, agents cannot use any stacks even if loci for Push and Pop are on[1]. As can be seen from the figure, for an agent to be equipped with a stack that can store all possible objects, all loci corresponding to Push and Pop and the regulatory loci must be turned on in the agent's gene.

Simulation Flow and Selection Mechanism
In an evolutionary simulation, the initial population's gene is generated as all loci are zero for all agents. Each agent performed production according to the state transition table encoded in its gene; the fitness of each agent is evaluated depending on the results of its production. The fitness function is defined in the following subsection.

For generation turnover, two parents are selected from the top 10% with a rank selection according to fitness values, and two offspring are produced using a one-point crossover. This process of selection and reproduction is repeated until the number of offspring reached a predefined population. Thereafter, bit inversions occur as mutations with a locus in each agent's gene in the next generation.

Although this is not a biologically plausible implementation, this design is adopted because the aim is to identify the role of recursive combinations.

---

[1]The regulatory locus was introduced to reduce computational time. We confirmed that there is no change in simulation results using a model without the regulatory locus for stacks.

Fitness Function
The evolutionary process and evolvability of recursive combinations under each fitness function were examined by evolutionary simulations. The following three fitness functions were set.

- Making any product:

$$F_{\mathrm{I}}(t) = \sum_{\mathrm{all}\,x} n_x^i(t),\qquad(2)$$

where x represents a product composed of up to l elements and $n_x^i(t)$ is the number of times the product x is produced by agent i at generation t. The fitness function F_{I} is based on the expectation that recursive combination is used in making many products.

- Making a specific product:

$$F_{\mathrm{II}}(t) = n_x^i(t),\qquad(3)$$

where x represents a product which is the longest, that is, l, and consists of the most number of types of elemental objects, namely, k. This fitness function is based on the fact that human made products have become increasingly complex in structure (Stout et al., 2008; Arthur, 2009). We choose a target product such as $ABAB$ ($k=2, l=4$) or $ABCABC$ ($k=3, l=6$).

- Making products as diverse as possible:

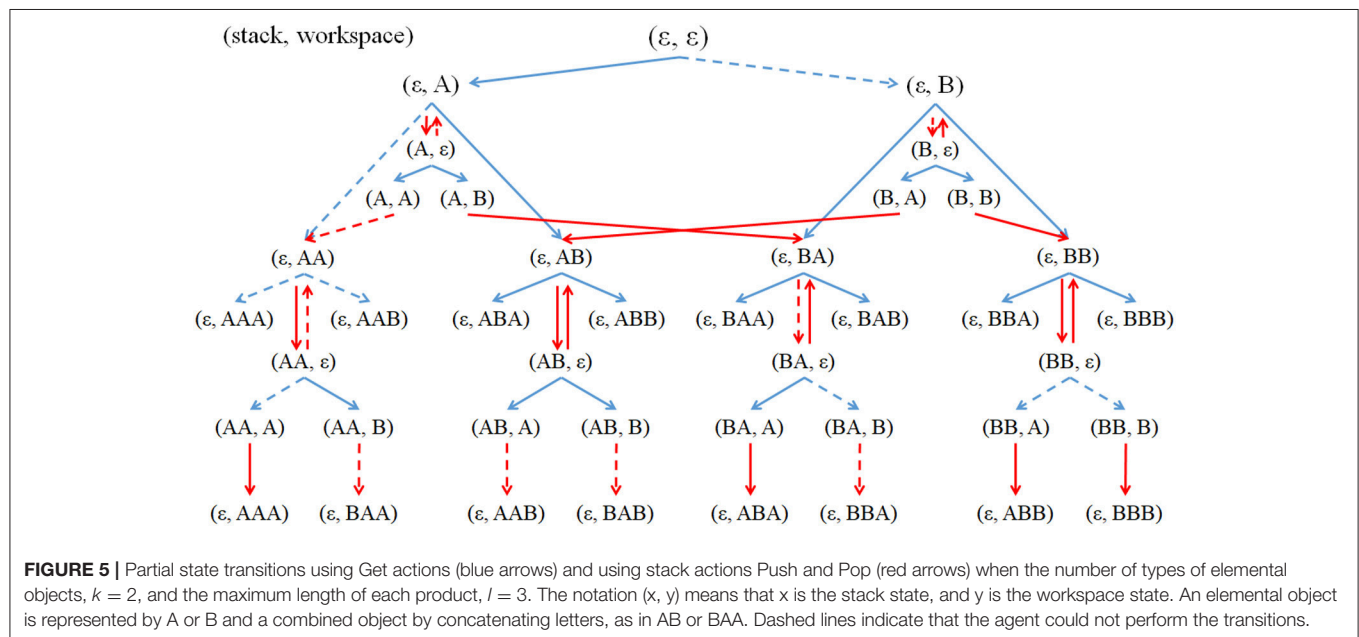$$F_{\mathrm{III}}(t) = \sum_{\mathrm{all}\,x} \delta\left(n_x^i(t)\right),$$

$$\delta\left(n_x^i(t)\right) = \begin{cases} 1, & n_x^i(t) \geq 1 \\ 0, & n_x^i(t) = 0 \end{cases}.\qquad(4)$$

This fitness function is based on the fact that humans make increasingly diverse products (Arthur, 2009). We expect that manufacturing many types of products encourages an agent's survival and reproduction, while manufacturing the same product does not.

Although the manipulation steps for making one product are not explicitly expressed in these fitness functions, they nevertheless indirectly influence agent fitness because an upper limit of the number of manipulation steps is set. Thus, when an agent requires a considerable number of manipulations to make one product, the number of products made decreases and the agent's fitness is reduced.

RESULTS

The purpose of this evolutionary simulation is to clarify the adaptability of recursive combination to demonstrate the conditions of the ecological environment and the process of evolution. In the first subsection, we show the simulation results in the three fitness functions introduced above, at first by setting

FIGURE 4 | Example of part of a state transition table. The number of the types of elemental object, $k = 2$. A corresponding gene code is shown above the table. The first bit of gene is a regulatory locus for stacks.

FIGURE 5 | Partial state transitions using Get actions (blue arrows) and using stack actions Push and Pop (red arrows) when the number of types of elemental objects, $k = 2$, and the maximum length of each product, $l = 3$. The notation (x, y) means that x is the stack state, and y is the workspace state. An elemental object is represented by A or B and a combined object by concatenating letters, as in AB or BAA. Dashed lines indicate that the agent could not perform the transitions.

the number of types of elemental object $k = 2$ and the maximum length of product $l = 6$. Then, the dependencies of these results on the parameters, k and l, are illustrated. These analyses suggest that recursive combination has two kinds of adaptabilities. In the second subsection, considerations based on the adaptabilities are used to modify the fitness functions to add cost factors that may affect the evolution of recursive combination. It is expected

that the cost of manipulation influences negatively the evolution of recursive combination because it requires a greater number of manipulation steps than non-recursive combination. We also investigated the influence of a possible failure of operation on the evolution of recursive combination. We considered the evolutionary mechanism of recursive combination only on the simulation in this section of the paper. The cognitive or linguistic

interpretations about the simulation results are considered in the Discussion section.

The parameters are summarized as shown in **Table 1**. The population size is 100, and the upper limit of manipulation steps is set at 10,000, which does not influence the results unless it is too small. Simulation results were taking 200 runs in each parameter. In this section, hereinafter, recursive combination, non-recursive combination, and the agent using recursive combination are called *RC*, *non-RC*, and *RC agent*, respectively.

The Fitness Function for Which the Recursive Combination Is Adaptive
Making Any Products

With the fitness function F_I, RC agents did not evolve in all the 200 runs as shown in **Figure 6A**. Since the fitness function F_I encourages the act of making any product, agents gained fitness by repeatedly making specific simple products over many production trials. The average fitness is 5,000 with the upper limit of manipulation steps set at 10,000. This fitness value indicates that agents make products containing only one element such as *A* or *B*, using Get and Stop actions, that is, two manipulations, and RC is not used as shown. The number of types of product is one with slight fluctuations. This means that the population is mostly occupied by agents making products with one elemental object. This result suggests one reason that RC is observed only in humans. In human activity, the typical case of product manufacturing is tool-making for resource acquisition. This notable human behavior requires the combination of elemental objects or units made from elemental objects. In contrast, animals other than humans develop survival strategies without tool-making, in which object combination is not necessary.

Making a Specific Product

With the fitness function F_{II}, **Figure 6B** demonstrates that the RC has appeared; it increased the average fitness when it appeared. It disappeared, however, with increasing the average fitness as shown in **Figure 7** that depicts an example of the transition of the population share of RC agents in a typical run under F_{II}. This phenomenon implies that RC makes it easier to discover a specific product than non-RC (a detailed explanation of this point is in the next paragraph). An agent using non-RC for a product obtains more fitness value than an agent using RC for the same product because RC requires longer manipulation steps than non-RC; and the opportunity for making products is limited by the upper limit of manipulation steps. Therefore, after the product is discovered, RC agents are taken over by non-RC agents. When the length of the gene (which is determined by *k* and *l*) is too long, it is hard for non-RC agents to take over from RC agents because the mutation is one locus per agent per generation. For example, converting an agent that performs a state transition shown in column 2 row 3 in **Figure A1** in Appendix (RC) to one shown in column 1 row 1 (non-RC) needs to switch four loci.

The fitness landscape of F_{II} makes hill-climbing evolution virtually impossible and makes it hard to discover a specific product *x* for earning fitness. We employed the adaptability of
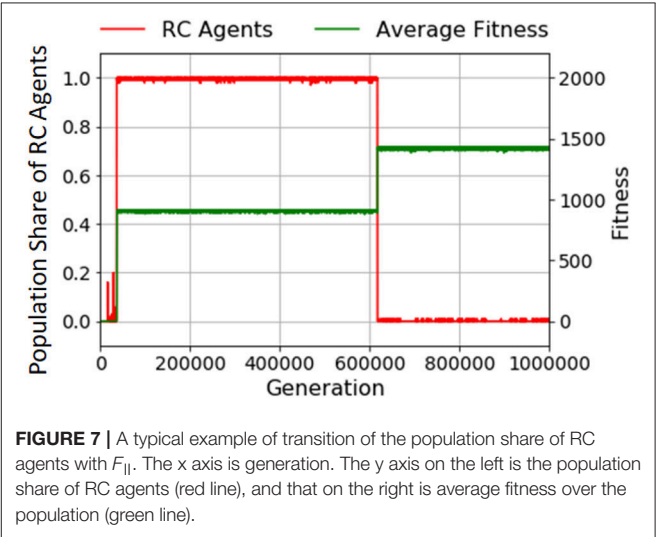

FIGURE 7 | A typical example of transition of the population share of RC agents with F_{II}. The x axis is generation. The y axis on the left is the population share of RC agents (red line), and that on the right is average fitness over the population (green line).

TABLE 1 | List of simulation parameters.

Name of parameter	Symbol	Value
Population size	–	100
Upper limit of manipulation	–	10,000
Number of types of elemental object	k	1∼4
Maximum length of product	l	3∼8

FIGURE 6 | Transitions of the population share of RC agents in **(A)** F_I, **(B)** F_{II}, and **(C)** F_{III} (average of 200 runs). The x axes denote generation. The y axes on the left denote the population share of RC agents (red line), and those on the right denote average fitness over the population (green line).
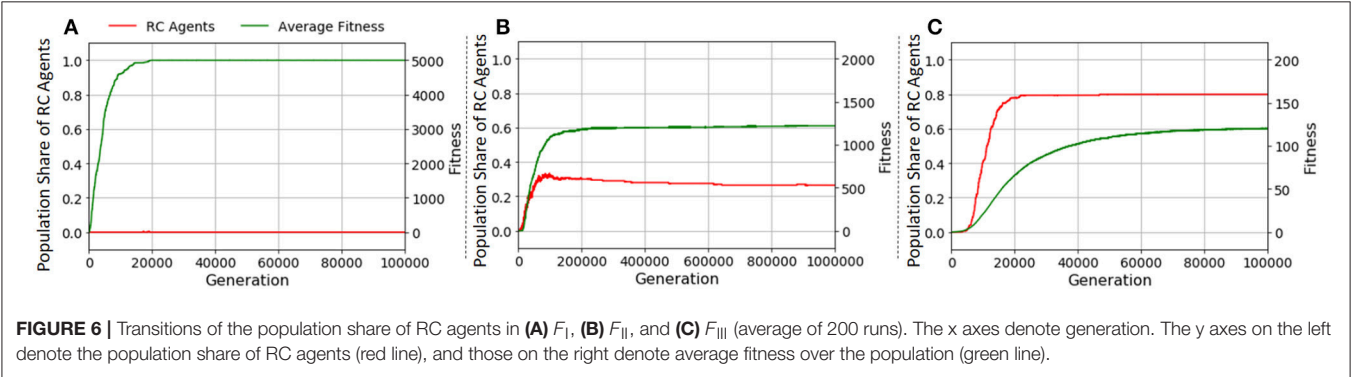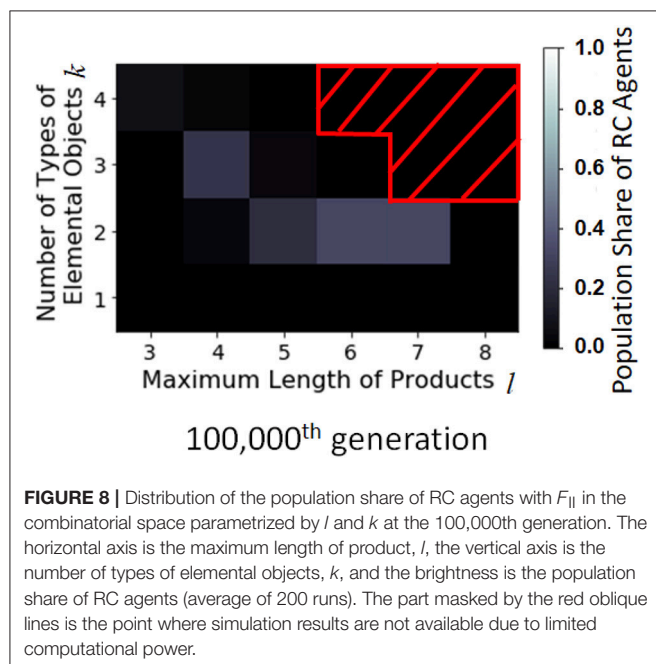
RC by providing it with multiple routes to increase the discovery rate of a specific product. When the agent makes a specific product *ABABAB* (if only non-RC agents without stacks exist) the production of this specific product is unique because the elements must be obtained in exactly the same order from left to right of the specific product, as shown in the top left of **Figure A1** in Appendix. Therefore, the discovery rate of making a specific product is very low. In contrast, if RCs are possible, at most 25 methods for making the product are available. Thus, the discovery rate greatly increases. Additionally, multiple methods to make a specific product promote robustness against failure in making processes (for which a detailed explanation is in section Effect of Failure Rate of Combination on Recursive Combination). In summation, the first adaptability of RC is diversifiability of production methods.

The number of production methods using RC depends on the size of the combinatorial space. **Figure 8** shows the population share of RC agents in a combinatorial space parametrized by the number of types of elemental objects k (vertical axis) and the maximum length of products per product l (horizontal axis). When $k = 2$, the combinatorial space is larger than when $k = 1$, the RC agents evolve more frequently than when $k = 1$; however, if the combinatorial space is too large, the agents cannot discover the production process of a specific product until the 100,000th generation.

Making Products as Diversified as Possible

In an environment fostering diversified products, RC evolves most in the three fitness functions as shown in **Figure 6C** compared with other cases (**Figures 6A,B**). **Figure 9** shows typical examples of the transition of the population share of RC agents in two runs with F_{III}. In this fitness function, the maximum fitness depends on the size of combinatorial space. If

FIGURE 8 | Distribution of the population share of RC agents with F_{II} in the combinatorial space parametrized by l and k at the 100,000th generation. The horizontal axis is the maximum length of product, l, the vertical axis is the number of types of elemental objects, k, and the brightness is the population share of RC agents (average of 200 runs). The part masked by the red oblique lines is the point where simulation results are not available due to limited computational power.

the upper limit of manipulation steps is sufficient for making all types of products, both RC and non-RC can earn the maximum fitness. Therefore, the RC agents or the non-RC agents can be maintained once either achieved the maximum fitness.

RC agents more frequently appears than other fitness functions because the production method using RC to make new products can evolve by less loci change than that using only non-RC. We explain this difference using **Figure 10**. For example, when an agent can already make *BABAB* as shown by solid arrows in the left branch, the agent evolves to make *ABABAB* by three loci changes represented by the broken arrows which depict the RC production method. These changes are much fewer than evolving to make the product only with the non-RC making method as shown in the right branch (6 loci changes). Therefore, agents to make new products using RC method are more easily attainable than those using non-RC method in evolutionary process. Further those that make new products earn more fitness than their ancestral agents. Thus, RC agents can appear and spread more rapidly than non-RC agents with F_{III}. The second adaptability of RC is diversifiability of product.

The effect of the size of combinatorial space was investigated. Since the RC production method is more effective in searching production space than non-RC, the RC agents are more likely to evolve when the combinatorial space is large enough as shown in the center part of **Figure 11**. However, if the combinatorial space is very large, such as $k = 3$ and $l = 6$, the making processes of products are difficult to find, and the RC agents are not likely to appear by the 100,000th generation.

Factors Affecting the Evolution of Recursive Combination
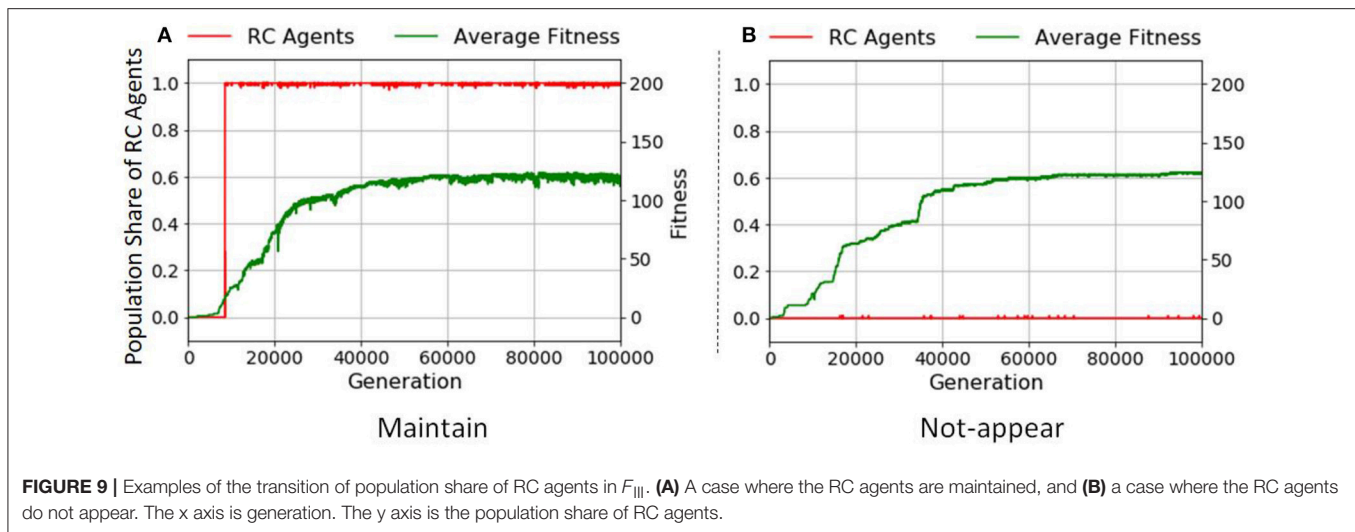
In the previous settings of the fitness functions, we identified two adaptabilities of RC: the diversifiability of production methods and the diversifiability of product. From these results, in this section, several factors that may affect the evolution of RC are introduced. The factors are the cost of manipulation and the failure of combination. RC exhibits a disadvantage when tool-making requires energy. In contrast, the diversification of production methods is useful for failure in object combination. As a result, these factors affect the evolution of RC. The evolutionary scenario of RC is expected from these effects.

Effect of Manipulation Cost on Recursive Combination

RC requires more manipulation steps than non-RC. We did not consider the cost incurred to perform operations in the simulation described in the previous section. If RC is costlier than non-RC, how does their evolution change? In order to find answers, we modified the fitness functions F_{II} and F_{III} as follows:

$$F'_{II} = \sum_x \frac{n^i_x(t)}{m^i_x(t)^c}, \tag{5}$$

$$F'_{III} = \sum_x \frac{\delta\left(n^i_x(t)\right)}{m^i_x(t)^c}, \tag{6}$$

FIGURE 9 | Examples of the transition of population share of RC agents in F_{III}. **(A)** A case where the RC agents are maintained, and **(B)** a case where the RC agents do not appear. The x axis is generation. The y axis is the population share of RC agents.
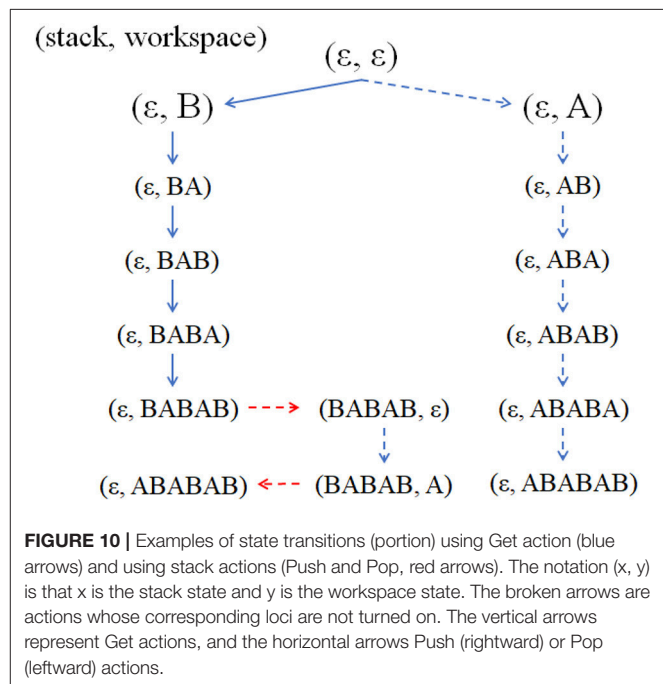
where $m_x^i(t)$ is the manipulation steps required to make the product x at each production for (5) and at its first production for (6) and the parameter c regulates the effect of the cost.

The agents incur the manipulation cost when they perform Get, Push, and Pop actions. **Figure 12A** illustrates the effect of the manipulation cost on the population share of RC agents. It is naturally understandable that increasing the manipulation cost made the evolution of RC more difficult with F'_{III} since RC requires more manipulation steps than non-RC. Even if an agent makes many types of products, the fitness is discounted at the cost of production depending on manipulation steps. However, with F'_{II}, the manipulation cost does not influence the evolution of RC. Since the fitness landscape of F'_{II}, and F_{II} as well, is not a hill-climb type but discrete, the difference of fitness values of the fitted traits is hard to affect the possibility of takeover from RC to non-RC agents (A detailed explanation is provided in section Making a Specific Product, paragraph 1).

Effect of Failure Rate of Combination on Recursive Combination

In the fitness function F_{II}, we expected that the multiple production methods by RC would promote robustness against failures in production processes. We introduced the failure of combination action into the model to confirm this expectation. With a constant probability, the agents fail to combine objects using Get or Pop action, and the state of the workspace becomes empty. This modeling expresses that a product is broken due to a failure of combination. The fitness functions are the same as the Equations (3) and (4). **Figure 12B** shows that the probability of appearance of RC increase gradually with increase in the failure rate of the fitness function F_{II}. In F_{III}, the population share of RC agents rise when the failure rate is not zero but decrease slightly with a larger rate of failure. These increases are explained by the function of stack to keep a combined object. If an agent fails to make a product on the way of production, the agent does not have to return to the initial state but can restart from a production step when a partial product is kept in the stack. This function of stack

FIGURE 10 | Examples of state transitions (portion) using Get action (blue arrows) and using stack actions (Push and Pop, red arrows). The notation (x, y) is that x is the stack state and y is the workspace state. The broken arrows are actions whose corresponding loci are not turned on. The vertical arrows represent Get actions, and the horizontal arrows Push (rightward) or Pop (leftward) actions.

realizes the diversification of production methods, but is not so strongly effectual for robustness. Actually, it is not so successful for higher failure rate in F_{III}, the higher is the failure rate and the longer is the manipulation steps for a product, the more difficult to complete the production process of the product. Thus, the population share of RC agents decreases with larger failure rate in F_{III}.

DISCUSSION

In this section, we mainly discuss the implication of each simulation result and its application to human evolution and language from the viewpoint of producing action sequences such

as making tools. First, from the simulation results of F_{II} and F_{III}, the adaptability and evolvability of recursive combination are considered. Next, a possible evolutionary scenario of recursive combination in human history is provided and supported with evidence from anthropology and archeology. Then, we speculate that recursive combination realizes flexibility of interpretation (that corresponds to diversifiability of production methods and of language products or expressions) and a driving force to diversify concepts and culture. Finally, we discuss the origin of recursive combination and recursive syntax by comparing two hypotheses (1) evolution of recursive combination via action control and (2) boot-strapping of recursive syntax via recursive intentionality.
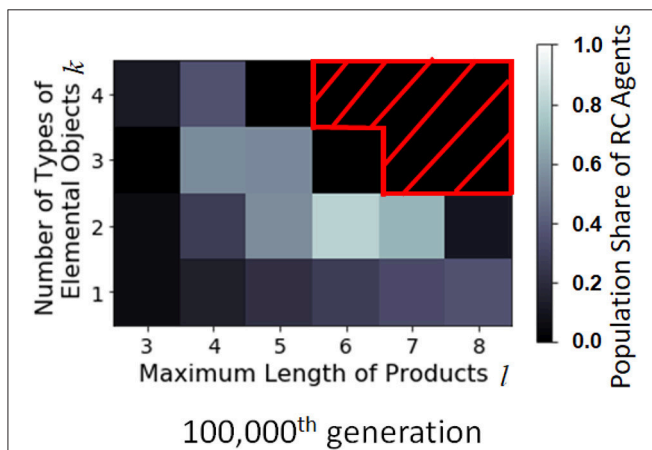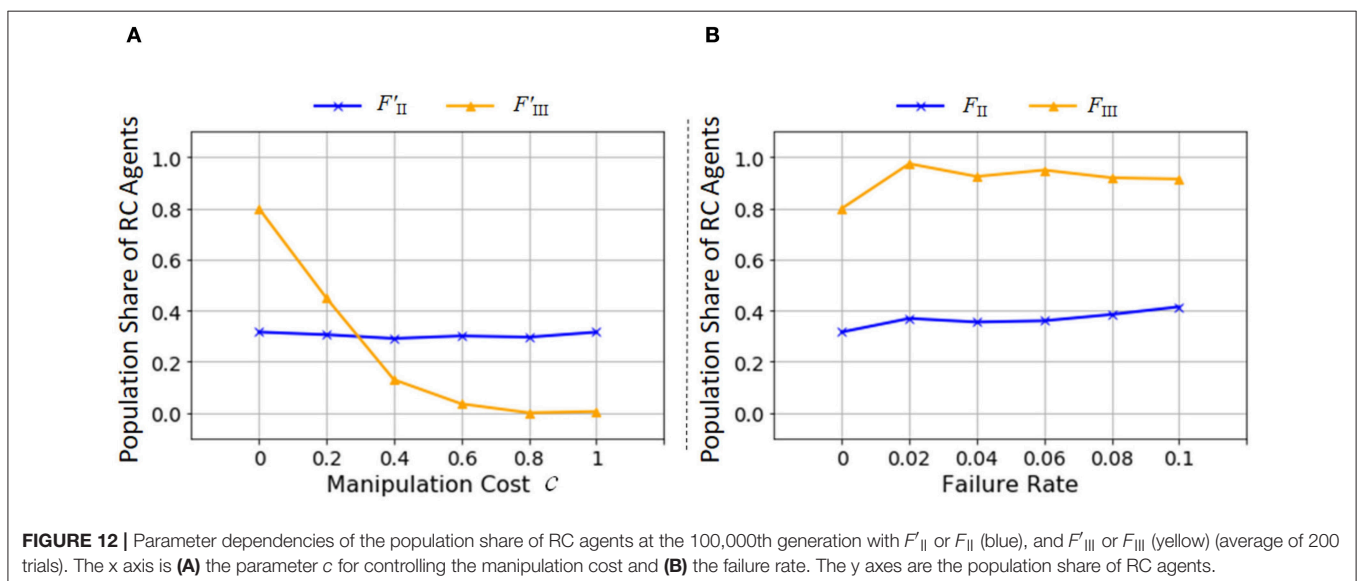

FIGURE 11 | Distribution of the population share of RC agents with F_{III} in the combinatorial space parametrized by l and k at the 100,000th generation. The horizontal axis is the maximum length of products, and the vertical axis is the number of types of elemental objects, and the brightness is the population share of RC agents (average of 200 runs). The part masked by the red oblique lines is the point where simulation results are not available due to limited computational power.

Adaptability of Recursive Combination

In an environment in which making a specific product with a complicated sequence is adaptive, production methods using recursive combination are discovered frequently (section Making a Specific Product). Additionally, the availability of multiple production methods for one tool is a workaround for inaccurate and/or irreversible manipulation. The greater access an agent has to multiple production methods the better that agent can make tools with increased stability (section Making a Specific Product); therefore, agents using recursive combination evolve faster than those that do not. When an agent must use many types of objects for product-making or must undergo a long process to make products (section Effect of Failure Rate of Combination on Recursive Combination), the frequency of failure derives from increase in inaccurate or irreversible manipulation; thus, diversifiability of production methods using recursive combination is effective.

In an environment in which making products as diversified as possible is adaptive, an agent searching for a production method that reuses existing methods can obtain relatively larger fitness than those who search for an all-new production method (section Making Products as Diversified as Possible). Therefore, the agent using recursive combination passes on its gene more easily than others. This adaptability is the diversifiability of products. In other words, recursive combination may have diversified the types of product in material culture beginning from stone tools. Human beings have diversified and complexified technology from the early stone age to the present. Arthur and Polak (2006) show that recursive combination of modularized technologies helped to identify more complex structures in a vast searching space. If the agents incur high manipulation costs, the adaptability of the diversifiability of products does not work (section Effect of Manipulation Cost on Recursive Combination).

Although we have already attempted other variants of this model, the approximate results of simulation (adaptability of recursive combination) did not change. The adaptability of

FIGURE 12 | Parameter dependencies of the population share of RC agents at the 100,000th generation with F'_{II} or F_{II} (blue), and F'_{III} or F_{III} (yellow) (average of 200 trials). The x axis is **(A)** the parameter c for controlling the manipulation cost and **(B)** the failure rate. The y axes are the population share of RC agents.

recursive combination will not be altered by adopting a learning algorithm such as a neural network instead of GA, since learning algorithms do not influence the size of the learning space for the production procedure. This expectation, however, has to be checked in future research.

The Evolution of Recursive Combination in Human History

How is a condition formed in which recursive combination is adaptive? From the results of the simulation with F'_{II} and F'_{III} (**Figure 12A**), when the manipulation cost is applied, recursive combination is used more easily and with a lower cost. As we introduced in section Results, recursive combination is not common in animal behavior; we assume that this strategy is costly and not adaptive in most environments. Consequently, we must identify the environmental conditions that promote the evolution of recursive combination while considering the existence of manipulation cost. Manual dexterity may be a key factor to performing significant object manipulations with decreased cost. Development of dexterity can lower manipulation cost at product-making.

Is there any archeological evidence in human evolutionary history corresponding to our proposal? In fact, the morphology of the early hominin's hand 3.00 mya acquired forceful opposition of the thumb, that is, an opposable thumb with the ability to exert forceful precision and power "squeeze" gripping (Skinner et al., 2015). Moreover, by 1.42 mya the hominin's hand had essentially evolved into the form of the modern human hand (Ward et al., 2013), in particular in terms of the distinctively human arrangement of the wrist associated with enhanced hand function when making and using tools. This evidence implies that early hominins might have been able to use their hands as dexterously as modern humans. According to other archeological evidence, tool use started around 3.39 mya (McPherron et al., 2010); tool-making around 2.60 mya (Plummer, 2004); and the recursive combination of objects around 0.28 mya (Moore, 2010). When the cost of object manipulation was high, recursive combinations could not have been maintained (section Effect of Manipulation Cost on Recursive Combination); this parallels the reasons that recursive combination is difficult to observe in animals, that is, its disadvantages (energy loss, manipulation injuries due to mistakes, etc.) are greater than its benefits.

Based on this account, we speculate on the possible evolutionary process of recursive combination. First, hominins came to use stone tools more frequently. This led to the evolution of hands and fingers to become dexterous enough to make superior tools that could survive repeated use. This dexterity helped decrease the cost of object manipulation and increase the chance of tool-making by reducing the steps to make each tool. When certain complicated tools were produced, recursive combination emerged as an adaptability to avoid failure in making these tools through diversification of production methods. Finally, these agents used their developing ability of recursive combination to develop various new tools, showing adaptability by diversifiability of products.

Diverse products can be made without recursion, and the recursive and non-recursive combinations can produce the same set of products. We argue, however, that recursive combination can increase the efficiency of product-making. If agents use non-recursive combination only, they make products through specific procedures. If they use recursive combination as well, they can create a variety of products from the combination of partial modules, and the creation procedure becomes flexible; thus, the success and discovery rates of production are improved. We showed that improving the success and discovery rates contributes to the successful diversification of products. Hominins could create a variety of products from the combination of partial modules or procedures in actual behavior of making stone tools (Moore, 2010, 2011; Stout, 2011).

Recursive Combination in Language

Let us now consider whether the adaptability of recursive combination (shown by this simulation and explained by the speculative evolutionary account above) can also be demonstrated in language. Recursive combination in language, that is, a syntactic operation, is used to generate hierarchically structured symbol sequences. In our simulation, object manipulation and product manufacturing are modeled on the lines of an agent combining elemental objects represented by a letter such as *A* and *B*, or a combined object by concatenating letters, such as *AB* or *ABC*. If this model applies to language, elemental objects are lexical items, and products are sentences. For instance, when non-recursive combination is performed, words are combined repeatedly, e.g., the agent combines a word *book* and a word *club* to a word *child*. When recursive combination is performed, combined words (phrase) combine to form another word or phrase, e.g., the agent combines words *child* and *book* and then combines it with *club* to form *child book club*.

Diversifiability of production methods by recursive combination in language is presumed to encompass the making of multiple hierarchical structures, because various combination procedures can be of utility. This diversifiability assists plentiful interpretations to one expression. In linguistic communication, the interpretations of a sentence depend not only on sequential order but also on hierarchical structures that are not directly disclosed to receivers. The multiple hierarchical structures may cause ambiguity in meaning sharing when hierarchical structures represent meanings as the notable characteristic of human language, which is known as structural dependency.

Diversifiability of products by recursive combination in language then entails generating various expressions or ideas, because various possible combinations of lexical items can be assumed by this adaptability. In this way, recursive combination enables and requires the creation of new expressions and concepts by combining symbols.

Taking together the two types of diversifiability described above, we introduce a concept called *co-creation*. Making a hierarchical structure by combining symbols does not merely produce an internal expression but constructs a hierarchically structured concept that leads to the creation of a new, sometimes fictitious, concept that can attain a socially shared reality via linguistic communication. At the same time, however, the interpretation of these hierarchically structured sequences

remains potentially ambiguous, enabling message receivers (as well as senders) to produce personal, sometimes creative, conceptual structures. In short, the interaction between senders and receivers promotes creativity in both parties. Our premise is that the adaptability of language is in co-creation. Co-creation is not necessarily a creative activity through actual collaboration. The viewpoint of co-creation, integrating two different functions (communication and thinking) can explain the reality and nature of humans and the human cultures they have cumulatively created (or, that have cumulatively evolved). Humans create and share new concepts via linguistic communication and produce higher-level concepts. *Money*, a symbolic concept socially created and shared, is a good example. We mutually believe that it mediates exchange among us, measures value, and makes it possible to store wealth—and so it does, based on this belief and on the new conceptual structures supported by this belief, such as banks, bonds, capital markets, and the global economy. In this way, novel concepts emerge and are *realized* through the interaction of the thinking function and the communication function. The cultural explosion and the spread of mankind all over the world around 50–100 Kya (Mithen, 1996) can be considered as having been brought about by co-creation through linguistic communication.

On the other hand, if new concepts and expressions continue to be created only in a certain group, cultural isolation may occur between that group and other groups. In particular, higher-level, abstract concepts that do not have concrete existence and are not grounded in any physical object, are often very difficult to interpret and share due to lack of appropriate underlying concepts and linguistic means to convey their meaning. This difficulty of mutual understanding is probably a major cause of cultural conflict.

Origin of Recursive Combination and Recursive Syntax

In the introduction, we mentioned two reasonable hypotheses, origin of recursive combination via action control (Fujita, 2009, 2016) and boot-strapping of recursive syntax via recursive intentionality (Stiller and Dunbar, 2007; Oesch and Dunbar, 2017). In this subsection, the possibility of integrating these two hypotheses will be discussed as a future research. Recursive combination in object manipulation is to combine combined objects. Recursive intentionality has a structure that embeds a subject into a subject. It might be that these two hypotheses describe similar evolutionary scenarios of two different abilities.

In our simulation model, recursive combination needs a stack to store an object temporarily. In human cognition, this function for temporal storing is implemented by working memory (Baddeley, 2000, 2007). Working memory is an important faculty for higher order general cognition and behavior in humans, i.e., complicated action planning, presence of intentionality, and generation or recognition of other physical or conceptual structures. Therefore, we should consider an evolutionary process of working memory in human history.

Stout (2011) analyzed the production methods of stone tools that required complicated action planning, both the Oldowan and Acheulian types, and illustrated the methods using a tree diagram (Stout, 2011, **Figure 1**). The analyses of stone tool-making in Moore (2010, 2011) and Stout (2011) are almost the same. The notable point of Stout's (2011) analysis is using a tree diagram with dominance relationship in hierarchical structures. According to this analysis, the process of production of Oldowan tools required several steps of action: procurement of materials (for stone core and hammer stone) of appropriate size, shape, and composition; examination of the core; selection of target point to strike; positioning and fixing of the core; selection of hammerstone grip; and finally, accurate striking. These manipulations can be expressed by a tree diagram that has sixth order nesting. Unlike Moore (2010, 2011), Stout (2011) argued that the production method of Oldowan tools has discrete infinity that leads to the hierarchical structure of language. In the production method of the Acheulian type, Stout pointed out that the action sequence for achieving sub-goals was incorporated recursively into a higher order goal since the process of making a stone flake was included in the higher order intention of making stone flakes.

Arbib (2011) simplified the analysis of Stout's (2011) tree diagram from the viewpoint of working memory and re-interpreted the sixth order tree diagram in the production of Oldowan tools to five working processes. The five processes correspond to the following questions that stone tool-makers must answer: (1) Do I have a hammerstone? (2) Do I have a core? (3) Is there an available affordance for flake detachment? (4) If so, proceed with flake detachment. (5) If not, back up as far as needed. For Acheulian tools, Arbib insisted that automatization of the action sequence (working memory becomes needless) was essential because a complicated action sequence for stone flaking was incorporated into a subordinate component of the production of a stone tool. These studies argued that maintaining and combining sub-goals or sub-ordinate processes were essential for goal-directed action sequences that was a remarkable feature of Acheulian stone tools. Therefore, it is highly possible that the ability of recursive combination appeared in the age of Acheulian at the latest. In our simulation, a learning process such as Arbib's "automatization" is not implemented. We will clarify the relation of recursive combination and automatization as a future work by employing simulations with learning algorithms.

Mentalizing also needs working memory to maintain the mental state of others who have intentionality, such as *Simon believes that Martin thinks that Charlotte supposes that Jane knows that Simon thinks . . .* . Some studies show that mentalizing is limited to around the fourth or fifth order by working memory requirements (Stiller and Dunbar, 2007; Oesch and Dunbar, 2017). Oesch and Dunbar (2017) experimentally suggest that from first to fifth-order intentionality is necessary to assist the processing of simpler syntactic structures, but beyond fifth-order intentionality the cognitive scaffolding is provided by recursive syntax. We may apply this suggestion to the hypothesis of the origin of recursive combination via action control. Namely, lower-order recursive combination is necessary to assist the processing of simpler syntactic structures, but for more complicated action planning the cognitive scaffolding is

provided by recursive syntax. It is assumed that two cognitive abilities, recursive combination and inference of intentionality, evolved separately then they were integrated to create diverse and complicated hierarchical structures.

We do not claim to know the origin of recursive syntax. However, we argue that, if diversity, novelty, and robustness of production are required to survive or reproduce, recursive combination has adaptability in the various domains, and the ability of recursive combination needs working memory. It does not matter whether it originates from action control, social cognition, or others.

CONCLUSION

Adopting the hypothesis that recursive combination of object manipulation is the precursor of the syntactic ability intrinsic to human language, we developed an evolutionary simulation of product-making to clarify the adaptability of recursive combination in human evolution. In our study, a recursive combination, which is considered as a unique human ability, was modeled as a recursive combination in action grammar.

The main finding reported by this study, as evidenced by an evolutionary simulation, *is that the adaptability of recursive combination increased the rate of discovery and success at product making by diversifying production methods* and therein *increased fitness by diversifying products*. We argue that recursive combination may have evolved to become a consistent feature of human nature, through the production and use of tools that was

later generalized to many aspects of human cognition, including human language. Effectually, this may be part of the explanation as to how and why recursive combination evolved to become a consistent feature of human language, and not of other animal communication systems.

REFERENCES

Arbib, M. (2011). From mirror neurons to complex imitation in the evolution of language and tool use. *Annu. Rev. Anthropol.* 40, 257–273. doi: 10.1146/annurev-anthro-081309-145722

Arthur, B. (2009). *The Nature of Technology: What It Is and How it Evolves.* New York, NY: The Free Press; Simon & Schuster.

Arthur, B., and Polak, W. (2006). The evolution of technology within a simple computer model. *Complexity* 11, 23–31. doi: 10.1002/cplx.20130

Asano, R., and Boeckx, C. (2015). Syntax in language and music: what is the right level of comparison? *Front. Psychol.* 6:942. doi: 10.3389/fpsyg.2015.00942

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. (2007). *Working Memory, Thought, and Action.* Oxford: Oxford University Press.

Boeckx, C. (2017). A conjecture about the neural basis of recursion in light of descent with modification. *J. Neurolinguist.* 43, 193–198. doi: 10.1016/j.jneuroling.2016.08.003

Bouwer, F., Zuijen, T., and Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PLoS ONE* 9:e97467. doi: 10.1371/journal.pone.0097467

Brown, R. (1973). *A First Language: The Early Stages.* London: George Allen & Unwin.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton & Co.

Chomsky, N. (1993). "A minimalist program for linguistic theory," in *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger,* eds K. Hale and S. Keyser (Cambridge, MA: MIT Press), 1–52.

Chomsky, N. (2013). The dewey lecture 2013: what kind of creature are we? *J. Philos.* 110, 645–700. doi: 10.5840/jphil2013110121

Chomsky, N., and Berwick, R. C. (2015). *Why Only Us: Language and Evolution.* Cambridge, MA: MIT Press.

Conway, C. M., and Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends Cogn. Sci.* 5, 539–546. doi: 10.1016/S1364-6613(00)01800-3

Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529

Dunbar, R. (2009). The social brain hypothesis and its implications for social evolution. *Annu. Human Biol.* 36, 562–572. doi: 10.1080/03014460902960289

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., and Bolhuis, J. J. (2015). Structures, not string: linguistics as part of the cognitive sciences. *Trends Cogn. Sci.* 19, 729–743. doi: 10.1016/j.tics.2015.09.008

Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380. doi: 10.1126/science.1089401

Fujita, K. (2009). A prospect for evolutionary adequacy: merge and the evolution and development of human language. *Biolinguistics* 3, 128–153.

Fujita, K. (2016). On the parallel evolution of syntax and lexicon: a Merge-only view. *J. Neurolinguist.* 43, 178–192. doi: 10.1016/j.jneuroling.2016.05.001

Greenfield, P. (1991). Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav. Brain Sci.* 14, 531–595. doi: 10.1017/S0140525X00071235

Greenfield, P., Nelson, K., and Saltzman, E. (1972). The development of rule bound strategies for manipulating seriated cups: a parallel between action and grammar. *Cogn. Psychol.* 3, 291–310. doi: 10.1016/0010-0285(72)90009-6

Hashimoto, T. (2001). "The constructive approach to the dynamic view of language," in *Simulating the Evolution of Language,* eds A. Cangelosi and D. Parisi (London: Springer Verlag), 307–324.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1568–1579. doi: 10.1126/science.298.5598.1569

Jackendoff, R. (2002). *Foundations of Language (Brain, Meaning, Grammar, Evolution).* Oxford: Oxford University Press.

Jackendoff, R. (2011). What is the human language faculty? Two views. *Language* 87, 586–624. doi: 10.1353/lan.2011.0063

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychon. Bull. Rev.* 24, 118–137. doi: 10.3758/s13423-016-1166-7

Matsuzawa, T. (1986). Comparative cognitive development in primatological perspective. *Hattatsu* 28, 58–70.

Maynard Smith, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution.* Oxford: Oxford University Press.

McPherron, S. P., Alemseged, Z., Marean, C. W., Wynn, J. G., Reed, D., Geraads, D., et al. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature* 466, 857–860. doi: 10.1038/nature09248

Mithen, S. J. (1996). *The Prehistory of the Mind: A Search for the Origins of Art, Religion, and Science.* London: Thames and Hudson.

Moore, W. (2010). "'Grammars of action' and stone flaking design space," in *Stone Tools and the Evolution of Human Cognition*, eds A. Nowell and I. Davidson (Boulder, CO: University Press of Colorado), 13–43.

Moore, W. (2011). The design space of stone flaking: implications for cognitive evolution. *World Archaeol.* 43, 702–715. doi: 10.1080/00438243.2011.624778

Oesch, N., and Dunbar, R. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *J. Neurolinguist.* 43, 95–106. doi: 10.1016/j.jneuroling.2016.09.008

Plummer, T. (2004). Flaked stones and old bones: biological and cultural evolution at the dawn of technology. *Am. J. Phys. Anthropol.* 39, 118–164. doi: 10.1002/ajpa.20157

Skinner, M., Stephens, N., Tsegai, Z., Foote, A., Nguyen, H., Gross, T., et al. (2015). Human-like hand use in *Australopithecus africanus*. *Science* 347, 395–399. doi: 10.1126/science.1261735

Stiller, J., and Dunbar, R. (2007). Perspective-taking and memory capacity predict social network size. *Soc. Netw.* 29, 93–104. doi: 10.1016/j.socnet.2006.04.001

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philos. Trans. R. Soc. B. Biol. Sci.* 366, 1050–1059. doi: 10.1098/rstb.2010.0369

Stout, D., Toth, N., Schick, K., and Chaminade, T. (2008). Neural correlates of early stone age toolmaking: technology, language and cognition in human evolution. *Philos. Trans. R. Soc. B* 363, 1939–1949. doi: 10.1098/rstb.2008.0001

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Ward, C. V., Tocherib, M. W., Plavcand, J. M., Browne, F. H., and Manthif, F. K. (2013). Early Pleistocene third metacarpal from Kenya and the evolution of modern human-like hand morphology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 121–124. doi: 10.1073/pnas.1316014110

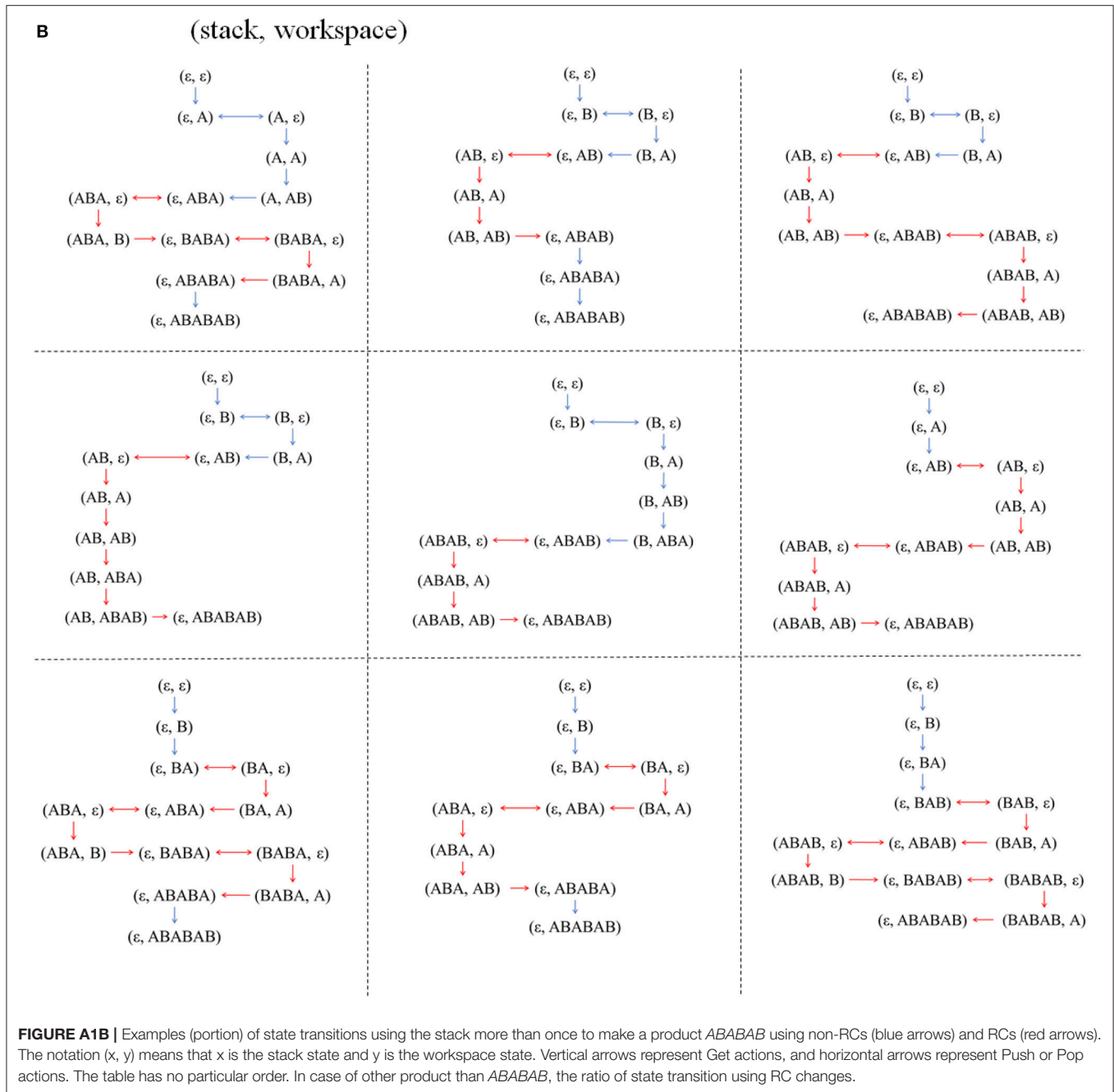Wymer, J. (1984). *The Palaeolithic Age.* London: Croom Helm.

APPENDIX

Figure A1 is a summary of state transitions of workspace and stack to make a specific product *ABABAB*.

FIGURE A1A | Examples (portion) of state transitions to make a product *ABABAB* using non-RCs (blue arrows) and RCs (red arrows). The notation (x, y) means that x is the stack state and y is the workspace state. Vertical arrows represent Get actions, and horizontal arrows represent Push (Rightward) or Pop (Leftward) actions. The table has no particular order. In case of other product than *ABABAB*, the ratio of state transition using RC changes. State transitions using the stack more than once are omitted.

FIGURE A1B | Examples (portion) of state transitions using the stack more than once to make a product *ABABAB* using non-RCs (blue arrows) and RCs (red arrows). The notation (x, y) means that x is the stack state and y is the workspace state. Vertical arrows represent Get actions, and horizontal arrows represent Push or Pop actions. The table has no particular order. In case of other product than *ABABAB*, the ratio of state transition using RC changes.

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: info@frontiersin.org | +41 21 510 17 00

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