

Correlational Studies in Typological and Historical Linguistics

D. Robert Ladd,¹ Seán G. Roberts,² and Dan Dediu^{3,4}

¹Department of Linguistics and English Language, University of Edinburgh, Edinburgh EH8 9AD, United Kingdom; email: bob.ladd@ed.ac.uk

²Language and Cognition Department, ³Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen 6500 AH, Netherlands

⁴The Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen 6525 EN, Netherlands

Annu. Rev. Linguist. 2015. 1:221–41

First published online as a Review in Advance on August 4, 2014

The *Annual Review of Linguistics* is online at linguistics.annualreviews.org

This article's doi: 10.1146/annurev-linguist-030514-124819

Copyright © 2015 by Annual Reviews. All rights reserved

Keywords

language universals, quantitative methods, phylogenetics, databases, statistical inference, color lexicon

Abstract

We review a number of recent studies that have identified either correlations between different linguistic features (e.g., implicational universals) or correlations between linguistic features and nonlinguistic properties of speakers or their environment (e.g., effects of geography on vocabulary). We compare large-scale quantitative studies with more traditional theoretical and historical linguistic research and identify divergent assumptions and methods that have led linguists to be skeptical of correlational work. We also attempt to demystify statistical techniques and point out the importance of informed critiques of the validity of statistical approaches. Finally, we describe various methods used in recent correlational studies to deal with the fact that, because of contact and historical relatedness, individual languages in a sample rarely represent independent data points, and we show how these methods may allow us to explore linguistic prehistory to a greater time depth than is possible with orthodox comparative reconstruction.

1. INTRODUCTION

The past several years have seen the publication of a number of studies, often based on large-scale databases of language data (see **Supplemental Appendix A**; follow the Supplemental Material link in the online version of this article or at <http://www.annualreviews.org>) such as *The World Atlas of Language Structures* (WALS; Dryer & Haspelmath 2013, first published in 2005), that use correlational techniques in support of hypotheses and claims about language typology and linguistic prehistory. These studies include (in no particular order) our own proposal of a link between population-level variability in the genes *ASPM* and *Microcephalin* and the geographical distribution of tone languages (Dediu & Ladd 2007); suggestions that the phonemic inventory of a language is related to the size of the population that speaks it (Hay & Bauer 2007) or to the distance of the language's homeland from Africa (Atkinson 2011); and claims either that computational phylogenetic techniques adapted from evolutionary biology allow inferences concerning proposed word-order universals (Dunn et al. 2011) or that they confidently place the Indo-European homeland in Anatolia (Bouckaert et al. 2012; R. Gray & Q. Atkinson, manuscript in preparation). In general, linguists have tended to be skeptical of such work, as the first author (who regards himself as a typical linguist in most respects) knows from the experience of discussing the Dediu & Ladd (2007) paper with colleagues when it first appeared. Many of the most controversial recent studies were carried out by scholars whose background gives them a firm grasp of statistical methods but who do not share the typical intellectual outlook or the background knowledge of someone trained in linguistics (the third author places himself in this category). Despite the potential for unproductive cross talk, however, we argue that this general line of work, although sometimes clearly flawed, is potentially very valuable, and that it is worth working toward collaboration and interdisciplinary understanding. Our goal in this review is to make these studies more accessible to those without an extensive background in statistics and to summarize some of the justified skepticism from within linguistics.

2. CORRELATION AND EXPLANATION

Correlation is simply a quantitative relation between the variability of two different phenomena. A trivial example (developed in **Supplemental Appendix B**) is the relation between the height and weight of professional baseball players: In general, baseball players who are taller (variability in one phenomenon) are also heavier (variability in another). A comparable linguistic example is the relation between word length and frequency of occurrence. Word length is variable (some words are longer than others), word frequency is variable (some words are more common than others), and the two kinds of variability are systematically and quantifiably related: Less frequent words tend to be longer, and longer words tend to be less frequent (Zipf 1936, 1949).

If we do observe a correlation between two variables, an obvious reaction is to wonder why. Most readers are aware of the well-known adage that “correlation is not causation” (http://en.wikipedia.org/wiki/Correlation_does_not_imply_causation), and correlations observed in nature are often part of complex networks of factors that can make it difficult to establish causes. Unless we can do an experiment—that is, manipulate a variable (or variables) in a controlled fashion and observe the effects of this manipulation on some other variable(s)—we may not be entitled to infer causal relationships. Unfortunately, of course, many domains of scientific interest are not amenable to experimental manipulation. For example, having observed a positive correlation between the presence of ejectives in a language and the altitude at which the language is spoken (Everett 2013), we cannot experimentally test the hypothesis that there is a causal relationship between altitude and phonology. We cannot move all the speakers of selected languages

to Tibet or the Andes to see whether over multiple generations they develop ejectives, or move the speakers of Quechua or Georgian to low-lying islands to see whether ejectives are lost. Instead, we can only seek additional data, often at different levels of explanation (e.g., in terms of an articulatory model in which low air pressure favors the production of ejectives). Research on many typological and historical aspects of language is affected by such limitations, although linguistics is hardly unique in this respect. Constraints on the availability of experimental data are present in all branches of science, from particle physics to molecular biology to cosmology, and correlational studies are found in many fields, especially in the historical and social sciences.

This is not to say that the use of correlational approaches is without problems. Most obviously, in any study with a correlational design, there will generally be multiple models that can reproduce the same empirical observations—that is, there will often be many potential answers to the question, “Why?” The concept of causality is actually very complex and widely debated (e.g., Beebe et al. 2009). Recent important advances hold the promise of both a better theoretical understanding and new quantitative methods for inferring causal models from correlational data (Pearl 2000, Hernan & Robins 2014). Most of the time, however, correlational studies can change only the relative likelihood of competing accounts. Yet in the long run, changing the relative probabilities of different models or theories is extremely informative. In this Bayesian view of science (see, e.g., Easwaran 2013, Talbott 2013, but also see Gelman & Shalizi 2013), well-collected, rigorously analyzed, and clearly reported correlational data can lower the probability of causal models that cannot generate the observed correlational structures, while increasing the probability of compatible models.

Determining what counts as well-collected and rigorously analyzed correlational data is not necessarily straightforward, though. The recent history of neuroimaging studies in cognitive neuroscience provides a useful case study here. An explosion of research early in this century showing correlations between specific cognitive or perceptual tasks and activity in specific brain areas was soon found to be based on statistical practices that led to unsound inferences (“voodoo correlations” in the memorable phrase of Vul et al. 2009; also see Carp 2012, Silver et al. 2011). At the same time, however, awareness of the problems quickly led to changes in practice. The field is now developing robust and standardized ways to deal with the issues that result from the use of correlational approaches, so that it can make progress in understanding the neuroanatomical and neurofunctional bases of human cognitive processes (e.g., Kriegeskorte et al. 2009).

A further issue, and one that is sometimes a legitimate basis for skepticism about correlational studies of language, is the extent to which correlations are valid and robust. Although these terms are sometimes confused, they refer to different concepts, and in theory both should be addressed by a correlational study. Validity can be ensured to a certain degree by being systematic and clear in laying out the assumptions behind the hypothesis proposed and behind the specific variables involved—what they mean, how are they defined, and how exactly they are measured and coded. However, if the different sets of assumptions available present different advantages and disadvantages, then researchers can at least test the robustness of the results by comparing the agreement between results derived from different data sets, different methods, and different assumptions and by systematically examining the differences and trying to understand their causes. A more detailed technical discussion of validity and robustness is presented in **Supplemental Appendix C**.

Validity is at the heart of many objections to correlational studies of language. In a large-scale test of implicational word-order universals, for instance, it is usual to assign every language a unique value as having either verb–object (VO) or object–verb (OV) order. What do we do with Dutch and German, the specialist wants to know: Do we count them as VO on the basis of main clauses or OV on the basis of subordinate clauses? This is unquestionably a genuine problem,

 [Supplemental Material](#)

although it should be considered a methodological issue, not a fatal flaw in the correlational approach. Specific decisions about doubtful cases can be made on the basis of theoretical work whose validity may be evaluated independently. Moreover, separate correlational analyses can be carried out on the basis of systematically different decisions about doubtful cases, and we can determine whether the conclusions are robust or whether they depend too delicately on the precise details of how the doubtful cases are coded. And in any case, correlational studies are only one tool among many for advancing our understanding: Many areas of linguistics are now seeing an expansion of experimental and modeling studies that can be used as a further check on hypotheses generated by correlational work.

3. CORRELATION AND UNIVERSALS OF LANGUAGE STRUCTURE

Several aspects of empirical work in linguistics do not require quantitative investigation. You do not need to do an experiment or carry out a survey to determine that the plural of *child* is *children* or that the Latin verb *esse* ‘be’ was irregular. This means that it is not only normal but frequently even appropriate for descriptive statements about language structure to be made without statistical support, and many areas of linguistics have consequently developed without a firm basis in quantitative approaches that are taken for granted in many fields. As a result, linguists are often ill prepared to evaluate the technical aspects of correlational studies. For many purposes, this is not a problem: As noted in the previous section, clear thinking about the assumptions and definitions underlying correlational studies is frequently as important as actually carrying out the quantitative work itself. Yet statistical facts about language structure have tended to remain marginalized, so the relevance of correlational findings is often not clear even if the findings themselves are not in doubt.

Consider our understanding of Zipf’s (1936, 1949) observation of the correlation between word length and frequency. Although the precise quantitative relationship varies between texts and between languages, the basic correlation is robust, in the specific statistical sense that correlational findings based on different data, measures, and assumptions tend to agree (for a review, see Piantadosi 2014). Moreover, the measures themselves—word length and frequency—are relatively easy to define and determine empirically, so the validity of the correlation is also not seriously in doubt. The question of causality, however, is much less clear. On the basis of his original findings, Zipf (1936, 1949) suggested that a frequently used word would adapt to become shorter in order to minimize production effort. Other theories suggest that, because short words tend to be more polysemous (Köhler 1999), they are more likely to be used in a greater number of contexts and, thus, more frequently (e.g., Köhler 1986). More recently still, Piantadosi et al. (2011a) demonstrated that a better predictor of a word’s length is not its frequency as such, but rather the information the word carries. This finding suggests that words may adapt under pressure to be informative, rather than to minimize production effort. However, the mechanism for any such adaptation is unclear, and it is probably relevant that word frequency is also correlated with many other measurable variables: with developmental factors such as the age of acquisition (Kuperman et al. 2013), processing factors such as the speed of lexical decision (Balota et al. 2007), semantic factors (see the sidebar titled *Correlations in Semantic Domains*) such as valence (Boucher & Osgood 1969, Kuperman et al. 2013), and diachronic factors such as the rate of lexical replacement over time (Pagel et al. 2007). Lexical decision time, meanwhile, is also correlated with word length (Hudson & Bergman 1985), the age of acquisition (Walker & Hulme 1999), and the physical size of the referent (Sereno et al. 2009, Yao et al. 2013), whereas the concreteness of a word’s referent is correlated with word length, informativeness (Piantadosi et al. 2011b), and the age of acquisition (Reilly & Kean 2007). This complex network of relationships has so far proved resistant to a comprehensive explanation (Figure 1), and at least partly as a consequence, the

CORRELATIONS IN SEMANTIC DOMAINS

Research on how semantic domains are divided into linguistic categories has tended to involve comparisons between specific languages rather than large-scale quantitative analyses. However, recent studies have used multidimensional scaling and clustering techniques, which help visualize distances between features so that common categories emerge (Croft & Poole 2008). For example, Majid et al. (2008) studied how different languages divide the space of cutting and breaking events. Participants described videos that showed people cutting, tearing, and breaking things in various ways. Dimensions were then identified along which languages tended to make lexical distinctions. For example, languages tended to distinguish precise cuts (e.g., *slice*) from imprecise breaks (e.g., *smash*). Majid et al. argue that languages have a “general solution” to conceptualizing this domain and that, although there is variation, there are also strong constraints. Similar quantitative studies have considered motion verbs (Malt et al. 2013) and adpositions for spatial relations (Levinson & Meira 2003), as well as morphological domains such as case (e.g., Clancy 2006). Malt et al. (2014) suggest that the potential benefits (but also the constraints) of quantitative methods ought to inform approaches to collecting linguistic data so that methods become more standardized, systematic, and ecologically valid.

Zipfian correlations and similar quantitative regularities have not really been integrated into mainstream linguistic theorizing. [Notable exceptions can be found in work by such scholars as Charles Yang (e.g., Yang 2013) and Janet Pierrehumbert (e.g., Pierrehumbert 2002).] A rapid survey of introductory texts finds Zipf mentioned only in Crystal’s (2010) *Encyclopedia of Language*.

There is, however, at least one type of correlational regularity about language that has played a key role in developing linguistic theory: what Greenberg (1966) called “implicational universals.” An implicational universal is a statement that if a language has structural property X, then it is likely to have structural property Y as well. Among the implicational universals now widely thought to be valid are correlations of word order in different types of phrases—for example, the claim that a language with OV word order will normally also place adjectives before the noun. This type of regularity was incorporated into mainstream theorizing in generative grammar in the early 1980s, and the idea that languages tend to be either head-final or head-initial rapidly became commonplace and is widely assumed today. In statistical terms, that is, the observed correlations are explained with reference to a single deeper principle—headedness—that governs the order of several different types of grammatical constituents.

For some time, many linguists simply took it for granted that word-order correlations are actually valid, and focused on the implications of that supposed fact for parsing and for grammatical theory [e.g., the “cross-category harmony” principle (Hawkins 1979, 1982)]. One of the first large-scale attempts to validate the correlations themselves was carried out by Matthew Dyer in *WALS*. His work (Dyer 2013a,b,c) suggests that consistent patterns of headedness are more frequent in certain types of phrases than in others, and that certain patterns may be restricted to certain parts of the world. In particular, a consistent association between the order of object and verb and the use of postpositions or prepositions seems to be universal. By contrast, there is little or no correlation between the order of OV and the order of adjective and noun (AN). The head-final OV+AN combination is found primarily in South Asia and Siberia, and the head-initial VO+NA combination primarily in Southeast Asia and sub-Saharan Africa. In Australia and the Americas, the most widespread pattern is the “mixed” OV+NA (Table 1) (Figure 2). The theoretical implications of such findings for characterizations of languages as head-initial versus head-final are outside the scope of our discussion, although we note that a considerable amount of theoretical

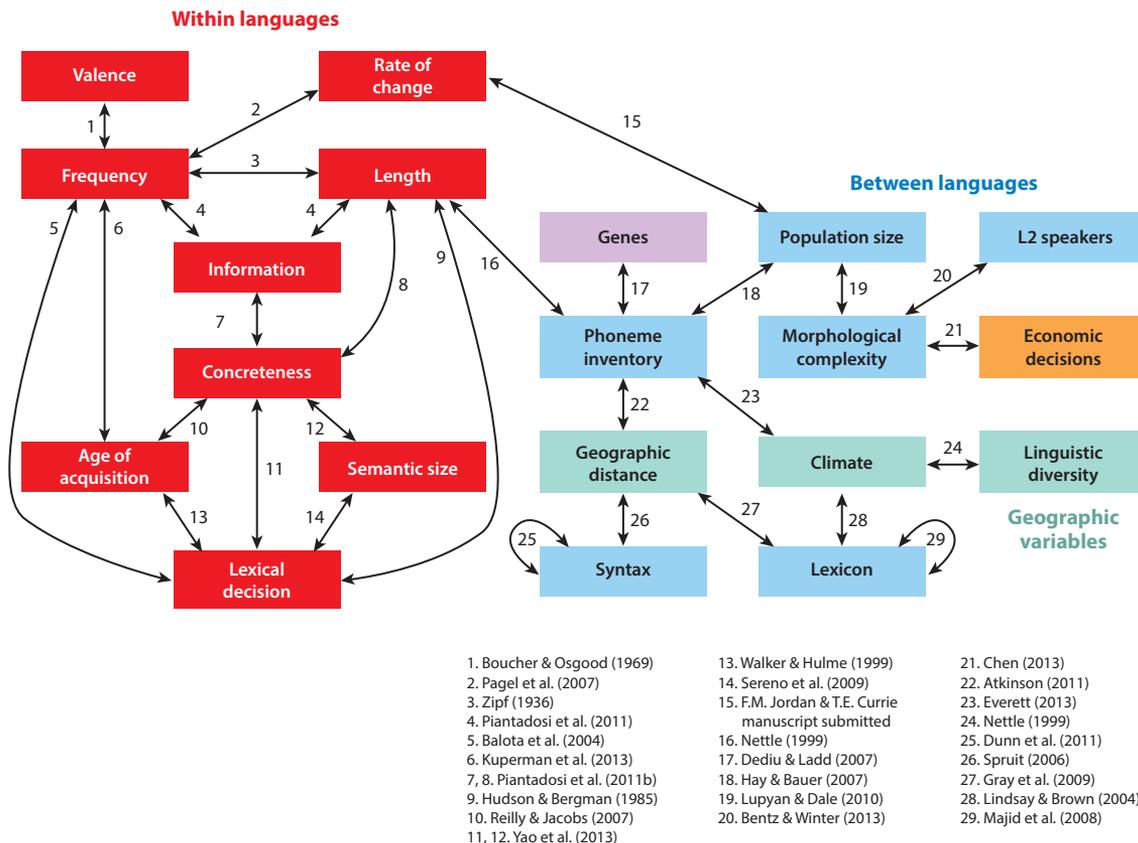


Figure 1

Correlations between various linguistic and nonlinguistic variables. The red boxes on the left show properties of words that have been studied within languages, and the boxes on the right show other properties that have been investigated between languages (green boxes denote geographic/environmental variables; orange and purple boxes denote extralinguistic variables).

work in linguistics continues to assume that languages tend to display a consistent “headedness direction” (e.g., Biberauer et al. 2014). Our general points are simply that implicational universals are a type of correlation and that large-scale correlational studies based on databases such as *WALS* are therefore directly relevant to progress in linguistics.

Actually, there is a more fundamental issue that arises from Dryer’s work in *WALS*, specifically from the facts that the consistently head-final *OV+AN* pattern is found primarily in a large contiguous area of the Eurasian landmass and that both Australia and sub-Saharan Africa exhibit substantial agreement on their own patterns. What if the observed word-order correlations are not in fact due to some underlying unity of headedness, but rather are simply the result of contact between neighboring languages or inheritance from a common ancestor language? That is, these correlations might be artifacts of the fact that the languages under consideration are not independent of each other. We return to this problem in Section 5, below. Here our point is the same general one about the value of correlational studies: A potential problem with the idea of implicational universals has come to light through the combination of large-scale, correlation-based research and explicit thinking about its validity.

Table 1 Data on implicational universals of word order

a	Order of object and verb		b	Order of object and verb	
	OV	VO		Order of adjective and noun	OV
Adpositions					
Postpositions	472	42*	AN	216	114*
Prepositions	14*	456	NA	332*	456

(a) The three left columns make clear that languages violating the implication universal (*indicated by asterisks*) are quite rare, whereas (b) the three right columns show that violations are widespread and, in particular, that there are considerably more OV languages that violate the implicational universal than ones that obey it. Data are from Dryer (2013a,b).

4. CORRELATIONAL STUDIES INVOLVING EXTRALINGUISTIC FACTORS

Although language-internal structural correlations such as implicational universals have long been treated as an interesting problem in linguistics, the same cannot be said of 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. For example, Chen (2013) reported that individuals who speak a language with no grammatical distinction between present and future tense are more likely to save money rather than spend. Similarly, Gay et al. (2013) showed that there are greater gender differences in political and economic opportunities among speakers of languages with gender distinctions in pronouns. Other proposed links between linguistic and nonlinguistic variables include phonetics and climate (Fought et al. 2004), ejectives and altitude (Everett 2013), and phonetics and sexuality (Ember & Ember 2007).

In some cases, reports of such relationships have been treated as interesting challenges for linguistics, and as potential catalysts for research questions (e.g. Hay & Bauer 2007, Roberts & Winters 2012). In general, however, attempts to link language structure with extralinguistic factors are almost intrinsically suspect: The Boasian tradition (e.g., Boas 1931) insists that there are no “primitive” languages and emphasizes the suitability of all languages to their speakers’ communicative needs; Whorfian ideas about the cognitive biases imposed by the native language are no longer widely credited (but see Carroll et al. 2004, Levinson 2012); Chomskyans assume the existence of universal (and probably innate) structural principles; and the Saussurean foundation of all modern linguistics means that linguists take for granted the arbitrariness of linguistic form almost from the first day of their first introduction to the subject. Not surprisingly, then, commentaries by linguists have often summarily dismissed claims like Chen’s or Gay et al.’s (e.g., Lewis & Pereltvaig 2013 and similarly skeptical comments by, e.g., Pullum 2012, Grossman 2013, Liberman 2013, Kiparsky 2014).

Although the linguist’s perspective is certainly a useful check on overenthusiastic acceptance of causal claims that have sometimes made it into big-name general science journals, blanket rejection of the idea that extralinguistic factors correlate with anything linguistic may lead us to throw away genuine explanatory insights into why languages are the way they are. As Östen Dahl (2013) puts it, “I am not sure this is the best way to respond to these ideas; it is not quite as easy to poke a hole in Chen’s argument as one might think, and to outsiders it may seem like an irrational knee-jerk reaction. It is better if we try to analyze the claims seriously; maybe we can learn something from it—especially if the results are spurious, it may make us be more careful about our own pet hypotheses.” In this section we consider two strands of research on correlations with extralinguistic factors that seem to us to have this kind of potential.

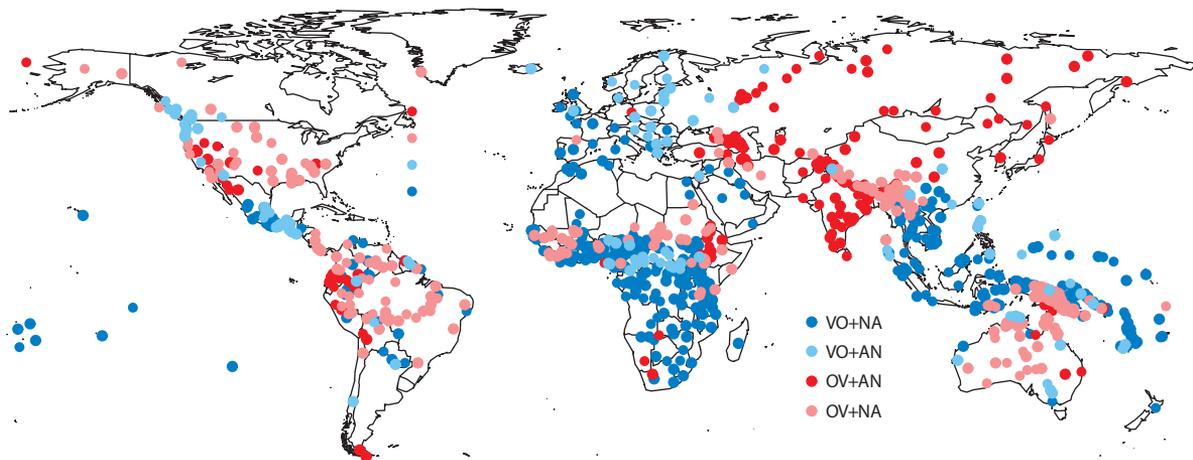


Figure 2

A map of the relationship between the order of object (O) and verb (V) and the order of adjective (A) and noun (N). Each point represents a language. Points in shades of blue indicate VO languages, and those in shades of red indicate OV languages. Within each of these two types, dark color denotes languages that conform to the proposed implicational universal, and light color represents languages that violate it. Figure based on data from Dryer 2013a.

4.1. Social Structure and Linguistic Complexity

A number of recent studies have considered the relation between social structure and linguistic complexity (see Nettle 2012 for a review). For example, languages spoken by larger populations have been found to have less complex morphology (Lupyan & Dale 2010) and slightly higher rates of lexical replacement or change (Wichmann & Holman 2009; Nettle 1999; F.M. Jordan & T.E. Currie, manuscript submitted). At the same time, they have also been reported to have larger phoneme inventories (e.g., Hay & Bauer 2007, Atkinson 2011, Wichmann et al. 2011), although Moran et al. (2012) and Donohue & Nichols (2011) have not found this correlation. There are also complex patterns in how social structure, demographics, and ecological factors influence rates of lexical borrowing (Bowern et al. 2011). All of this evidence suggests that there may be trade-offs in complexity between different levels of structure. Such an inference is consistent with correlations between the size of the phoneme inventory and the average word length (Nettle 1999, Wichmann et al. 2011) and between morphological complexity and phrase length (Juola 1998). Thus, it may seem appropriate to treat these findings as indications of language-internal structural tendencies rather than as direct effects of social structure (for work on dialects, see the sidebar titled Exploring Differences Between Dialects).

Nevertheless, recent comparative work (e.g., Trudgill 2002, McWhorter 2007) suggests that morphological simplification can result from language contact [and indeed, in specific cases this idea has been around for a long time; see, e.g., Thomason & Kaufman (1988, section 9.8) on the effects of contact with Scandinavian and Norman French on Old English]. If so, then we should expect languages with a more extensive history of contact to have simpler morphology than languages in less regular contact with others. Lupyan & Dale (2010) used a large-scale correlational study to quantify and test this hypothesis directly. They took the number of speakers of a language as a proxy for the probability of contact and showed that this number correlates with a composite morphological complexity score based on the number of morphological cases. They also showed that this correlation exists in many language families. Subsequently, Dale & Lupyan

EXPLORING DIFFERENCES BETWEEN DIALECTS

Most correlational studies are based on worldwide samples of languages, but correlational methods can also apply to more local analyses. Phonetic differences between dialects of English have been explored using clustering methods (e.g., McMahon et al. 2007, Sullivan & McMahon 2010). Most correlational studies also focus on continuous or countable variables such as population size, number of phonemes, or number of morphological cases. However, it is also possible to perform correlational studies on categorical variables such as differences in morphosyntax (Szmrecsanyi 2014). Geographical effects can also be observed on a local scale. Spruit (2006) demonstrated a correlation between syntactic differences between dialects of Dutch in the Netherlands and the geographic distance between those dialects. Correlational studies of dialects have proved useful for quantifying differences between varieties in an objective way that is free from typical ingrained social prejudices and that deals with complex relationships too numerous for a single human linguist to have intuitions about (Nerbonne 2009).

(2012) also demonstrated a correlation between a preference for regularized forms and children's contact with nonnative speakers. Although this finding demonstrates a general agreement (and hence suggests that the correlation is robust in the sense introduced in Section 2, above), there have been criticisms (e.g., Nettle 2012) that the proxy measures are too coarse and that therefore the inferences may not be valid. Moreover, other studies using different measures of grammatical complexity have mixed results: Sinnemäki (2009) has found that larger populations tend to have more systematic paradigms, whereas Nichols (1992) and Martowicz (2011) have found no evidence for a link between population size and the complexity of specific structures. Nevertheless, a more focused test of the hypothesis, carried out by Bentz & Winter (2013), demonstrated that languages with large proportions of L2 (nonnative) speakers have very restricted nominal case systems compared with languages with low proportions of L2 speakers. All of this evidence strengthens the idea that extensive language contact can result in morphological simplification and suggests that this process is one cause of the correlations that have been found.

As with the example of altitude and ejectives mentioned above, it is not feasible to test this causal hypothesis experimentally on real languages over multiple generations, but it is possible to demonstrate the robustness of the theory using alternative methodologies such as laboratory experiments with artificial languages and agent-based computational simulations. (Agent-based simulations are programs that simulate linguistic interactions within a population of artificial agents; researchers specify their assumptions about how interactions work and parameters such as the number of individuals, and they can then measure how global properties emerge in the model from individual interactions by varying parameters or comparing different assumptions.) For example, Dale & Lupyan (2012) provided an agent-based simulation as a proof of concept that morphological complexity decreases when a language is learned by adults rather than children. Around the same time, Little (2012) ran a laboratory experiment in which participants learned an artificial language, and found that the learners used morphologically simpler forms when they thought they were communicating with speakers of a different artificial language. That is, Little's result suggests that effects on morphological complexity could be driven by "foreigner-directed" speech, rather than adult learning. However, both results are consistent with the broader theory that contact is an important key to understanding the influence of social factors on language structure.

4.2. Geographical Factors in the Structure of the Color Lexicon

Berlin & Kay (1969) reported that there are universals in the structure of basic color words between languages, and suggested that there is a hierarchy of basic color terms, with more basic distinctions (e.g., dark/light) emerging in a language before others (e.g., pink/purple). When it appeared, this proposal was extremely influential in moving linguistics away from a strongly relativistic outlook toward greater acceptance of universals of various sorts, broadly in keeping with the Chomskyan ideas that were coming to dominate at that time. Berlin & Kay's approach has subsequently been subject to many criticisms, including objections to their methods, reports of exceptions to the proposed universals, and more general critiques of the notion that all languages have dedicated color terms that cover the whole visible spectrum (Rosch 1974, Saunders & van Brakel 1997, Davidoff et al. 1999, Levinson 2000, Lin et al. 2001). However, it has also inspired several correlational studies suggesting that universals of the color lexicon may derive from biological and environmental universals and from general principles of cultural evolution, rather than representing a specifically linguistic universal. In this literature, that is, there is a move from the observation of a pattern to a quantified correlation to controlled experiments and models that demonstrate causal links. Although the initial, qualitative work set the scene for the issues, more quantitative work has helped to evaluate the theories.

Bornstein (1973) noted that there is a correspondence between the size of the basic color vocabulary and the distance of the linguistic community from the equator, suggesting that exposure to greater amounts of ultraviolet light (UV) on the equator diminished the ability to perceptually discriminate colors. The quantitative correlation between these variables was confirmed by Ember (1978). Brown & Lindsey (2004) went further by showing a correlation between the size of the basic color lexicon and actual UV levels. They also performed a color discrimination experiment with participants who wore tinted glasses to simulate the effects of UV phototoxicity. The experiment allowed the first causal interpretation because there was a time dimension: Participants had naturally randomized color responses before the experiment, but their responses differed after wearing tinted glasses.

However, Hardy et al. (2005) found conflicting results when they added a condition in which older participants were given stimuli virtually manipulated to simulate less exposure to UV. Subsequent work suggests that the conflicting results may result in part from individual developmental factors based on differences in ambient light. Laeng et al. (2007) tested the perceptual color discrimination of Norwegians living in the same city but born in different seasons and at different latitudes. Those born at higher latitudes or in winter, when sunlight is shifted toward blue, had reduced sensitivity to yellow–green and blue–green contrasts but greater sensitivity to variations in the purple range. That is, sensitivity to particular color regions may be adapted to the properties of ambient light at birth. This finding suggests that there is a developmental aspect to the interaction between perception, latitude, UV radiation, and language.

Beyond these correlational and experimental studies based on Berlin & Kay's (1969) original book, agent-based computational simulations have also been used to research the evolution of the color lexicon (e.g., Steels & McIntyre 1999, Dowman 2007, Baronchelli et al. 2010). Loreto et al. (2012) demonstrated that an approximation of the Berlin & Kay hierarchy emerges in a model in which computational agents have human perceptual discrimination of color regions. Agreement on categories and labels emerges earliest in areas of low sensitivity, apparently because individuals have more difficulties in communicating these colors and negotiate them in interaction more regularly. This observation suggests that the structure of the color lexicon is influenced not only by geographically determined individual developmental factors, but also by universals of cultural interaction.

4.3. Graphic Summary

The correlations discussed in this and the previous section, along with others that space limitations prevent us from mentioning, are diagrammed in **Figure 1**. This graphic representation makes clear, among other things, the complexity of the interactions among the various factors. This complexity is both an important motivation for attempting large-scale quantitative studies and a significant barrier to making sure that the findings of such studies are valid and robust.

5. LANGUAGE RELATEDNESS IN CORRELATIONAL STUDIES

We now return to one of the central problems in applying correlational methods to language data, namely the assumption that separate languages are independent data points. We encountered this problem in Section 3, in connection with Dryer's (2013a,b,c) survey of VO and NA word order. The specific linguistic hypothesis under investigation there is that languages exhibit a consistent headedness direction in clauses and in noun phrases: Dryer's data raise the suspicion that this correlation may apply primarily to languages in specific parts of the world. Standard evaluations of whether a correlation is statistically significant (see **Supplemental Appendix B**) presuppose that the observations on which it is based are independent of one another; if they are not, the chance of reaching a false-positive conclusion (more formally, a Type I error) are increased. In our example of the relationship between the height and weight of professional baseball players, the physical traits of individual players are presumably independent, and the overall correlation is therefore valid. However, if our sample of players happened to include a large number of members of the same extended family, the correlation would presumably be strengthened by shared genetic and environmental features, but its validity as a general statement of the relationship between height and weight would be weakened.

In cross-linguistic studies there are several sources of nonindependence. The most important is sometimes referred to as Galton's problem (Mace & Pagel 1994): When languages are related through descent from a common ancestor (or protolanguage), some of their properties are more similar than might be expected by chance, simply because they reflect the properties of this common ancestor. This means that we cannot take related languages as completely independent data points, and any correlational study must therefore control or correct in some way for the nonindependence of the languages under consideration (Roberts & Winters 2013). However, nonindependence due to relatedness is clearly a matter of degree. We know that some languages are more closely related than others: If we observe a common typological tendency in Russian, Spanish, and Armenian (which are demonstrably but distantly related), this observation is more informative than finding shared features in Russian, Polish, and Bulgarian (which are so closely related that many similarities are evident even on superficial inspection). This observation suggests that in some way we need to quantify the degree of relatedness between the languages in our sample to control for it in our statistical analysis. Doing so unfortunately complicates the statistical analysis considerably, and immediately involves us in debates about the number, composition, and internal structure of language families.

Another important source of nonindependence between languages is contact. It has been known for a long time that languages whose speakers are in regular contact often exhibit a variety of similarities, sometimes to the extent of forming identifiable "linguistic areas" such as the Balkans, India, or Mesoamerica (e.g., Thomason 2000). Here, again, it is obviously inappropriate to treat the existence of a given typological feature in one language as entirely independent of its occurrence in a neighboring language. Unfortunately, for purposes of conducting correlational analyses, quantifying contact is even more difficult than quantifying genealogical

 [Supplemental Material](#)

relatedness. (A further complication arises from the fact that, by the nature of language splits, languages in contact also tend to be related.) As a proxy for contact, many studies have used geographical distance between the languages, on the assumption that languages that are closer together in space also have more opportunities for linguistic exchange—yet there are well-known counterexamples to the idea that geographical closeness implies linguistic and cultural contact. This is clearly a point on which historical linguistics and sociolinguistics can make meaningful contributions to assessing the validity of correlational studies, and can propose refinements to the kinds of factors used as proxies for contact.

5.1. Controlling for Language Relatedness and Contact

There are several ways of dealing statistically with these sources of nonindependence. A method used relatively commonly in the typological literature is to extract a subsample of languages that can be considered unrelated and unlikely to be in contact. (For specific approaches to constructing the subsample, see Dryer 1989, 2011; Janssen et al. 2006; Cysouw & Comrie 2008; Hammarström 2009.) This method has several disadvantages, most notably the drastic reduction in statistical power and the fact that it relies on relatively arbitrary judgments of independence. However, it also has the advantage that it can be easily integrated into a randomization approach, in which multiple subsamples are generated from the full sample, and the resulting distribution of results from all the subsamples can be used for statistical testing.

Another relatively common method is to compute distances between languages. These distances can be of different types. One can compute structural distances between languages, considering typological features as the dimensions of a very high dimensional space in which each language is a single point, and calculating the Euclidean distances between languages in this space. Somewhat more intuitively understandable are historical distances (computed as the distance between the languages on a linguistic family tree) and geographical distances (either simple great-circle distances between the places where two languages are spoken or more sophisticated distances that take geographical and ecological barriers into account as well). These distances are then analyzed using Mantel correlations (Mantel 1967), which express the relationship between distance matrices by taking into account the nonindependence of observations. Some examples of this approach are Dediu & Ladd (2007), which computes partial Mantel correlations between structural linguistic distances and human genetic distances while controlling for historical linguistic and geographic distances, and Dediu & Levinson (2012), which computes distances between the stability profiles of language families and correlates them with geographic distances. This method has the advantage that it is easy to apply but the disadvantage that the results may be difficult to interpret.

A very promising method is to control for historical and geographical relatedness using hierarchical regression, or so-called mixed-effects models. This is a general statistical approach to dealing with the fact that observations can be “nested” and that they tend to be more similar within higher-level units than between them. (For example, languages within families are generally more similar than those in different families, and languages within the same continent are more similar than those in different continents.) The mixed-effects approach is extremely flexible and powerful, and has recently come to be widely used in experimental linguistic work (e.g., Baayen et al. 2008, Jaeger 2008). It does not require any particular assumptions about the causes of similarity within units, and can deal with both multiple nested levels (e.g., genera within families and countries within continents) and crossed levels (e.g., historical relatedness and geographical proximity). However, it requires crisp boundaries between units, which are hard to justify, especially for geographical distance.

Another promising approach is to use phylogenetic methods inspired by evolutionary biology. These are specifically designed to deal with relatedness due to descent from a shared ancestor (see Huelsenbeck et al. 2001 and Felsenstein 2004 for introductions and discussions), and recent extensions of this approach can take geography and horizontal processes (e.g., contact) into account as well. Probably the best-known use of phylogenetic methods is to infer the internal structure of language families from various types of data, most often cognacy judgments on basic vocabulary (see McMahon & McMahon 2005 and Dunn et al. 2008 for introductions and discussions). This technique has been broadly validated on the basis of well-studied language families, in which phylogenetic methods yield family trees in rough agreement with those established by orthodox historical linguistic methods (McMahon & McMahon 2005). However, its real power, especially in a Bayesian framework, is that extremely complex models can be specified that use both vocabulary and typological data, as well as nonlinguistic information such as dates of attestation and geographic locations. This technique allows for the principled estimation of dates but does not, unlike glottochronology (Swadesh 1955), require a clocklike constant rate of change. It has been used to estimate the date of protolanguage splits (e.g., Gray & Atkinson 2003, Atkinson et al. 2005) and to estimate rates of change (Pagel et al. 2007, Atkinson et al. 2008, Dediu 2011) and rates of geographical spread (Bouckaert et al. 2012). Other important questions that can be dealt with in this way are those related to correlated evolution (Dunn et al. 2011) and the relative importance of vertical inheritance and language contact (Currie et al. 2010, Gray et al. 2010, Greenhill et al. 2010, Nelson-Sathi et al. 2010).

Space limitations do not permit us to discuss a number of other statistical techniques for dealing with issues of nonindependence. These include (a) the use of permutation/randomization techniques (Edgington & Onghena 2007; relevant work on language includes Janssen et al. 2006, Dediu & Ladd 2007, Dediu & Levinson 2012); (b) phylogenetic generalized least squares (Butler & King 2004; relevant work on language includes Currie et al. 2013, Torreira & Roberts 2014, Verkerk 2014, and S.G. Roberts, J. Winters & K. Chen, manuscript submitted); and (c) explicit models of spatial nonindependence developed in fields such as ecology and geology [e.g., kriging (Cysouw 2008), spatial autocorrelation (Knoolhuizen & Dediu 2012), and network modeling (Cysouw et al. 2012)].

5.2. Correlational Studies and Language Prehistory

In this final section we illustrate the use of two of the methods summarized above, namely mixed-effects modeling and phylogenetic approaches to inferring correlated evolution, to deal with the problem of language relatedness. In both cases we consider recent studies that have drawn both a good deal of interest and more than a little severe criticism.

The use of mixed-effects modeling to control for language relatedness attracted considerable attention in the wake of a proposal by Atkinson (2011), which was widely featured in mainstream media and generated heated scholarly discussion. Atkinson claimed that a measure of phonological complexity derived from *WALS* shows a decreasing statistical trend with greater distance from Africa, and suggested that this trend parallels the similar known trends of decreasing genetic and phenotypic diversity due to the spread of modern humans from their origin in Africa (Ramachandran et al. 2005, Betti et al. 2009). The subsequent discussion includes several comments in *Science* (e.g., Cysouw et al. 2012, Wang et al. 2012); a special issue of *Linguistic Typology* in 2011 (Bybee 2011), which contains 13 contributions; and two responses by Atkinson (2012a,b).

Atkinson's (2011) original study suffers from many shortcomings, and the current consensus is that its conclusions do not hold. For example, the phonological data in *WALS* are coded in terms of

fairly coarse categories, so Atkinson’s operationalization of phonological complexity lacks a great deal of validity (e.g., Cysouw et al. 2012, Wang et al. 2012). Also, again, Atkinson conducted a search for a single origin from which his measure of phonological complexity is best predicted by geographic distance; even taking various controls into account, this method is not robust to nonlinear relationships and tends to find such origins even when the process is random (e.g., Cysouw et al. 2012). Finally, the proposed causal mechanism—a serial “founder effect,” whereby daughter populations splitting off from an ancestral group retain only a subset of the source language’s phonological complexity, parallel to what happens in genetics—seems highly implausible for language (e.g., Hunley et al. 2012). Nevertheless, it is important to note also that the use of mixed-effects models to control for language relatedness works very well (see Jaeger et al. 2011 for a detailed discussion). More generally, it is worth highlighting that this flawed proposal has generated a lot of constructive interest in correlational methods in linguistics.

Turning from mixed-effects modeling to phylogenetic methods borrowed from biology, we consider an innovative study by Dunn et al. (2011), which used phylogenetic inference to look at the problem of implicational universals of word order. Instead of fighting against language relatedness—for example, by using a stratified sample of languages or mixed-effects models in ways summarized in the previous subsection—Dunn et al. argued that one can actually make use of this relatedness, and the historical processes that gave rise to it, to draw inferences about language typology. The principle is to treat each change in the value of a given typological feature as a historical event, which should predict the occurrence of other changes that are linked by the hypothesized implicational universal. For example, one can check whether changes in the VO order are correlated with changes in the NA order. If there are enough such events in a language family, then we can quantify the strength of the dependency—the correlation—between the typological features. Using this approach, Dunn et al. (2011) found that the Greenbergian word-order universals do not hold across the four large language families they tested; there are dependencies between word-order features, but they tend to be specific to specific language families. This finding seems to strengthen the doubts about implicational universals raised by Dryer’s work (Dryer 2013a,b,c), discussed in Section 3.

A further important property of Dunn et al.’s (2011) approach is that phylogenetic reconstructions of language family structure make an intrinsic assumption about time. The correlated evolution method estimates the probabilities with which pairs of features change values (Figure 3). If changes in one feature strongly tend to trigger changes in the other, we may be entitled to draw an inference of causality (for a detailed explanation and clarification, see Levinson et al. 2011). We may do so because time can be inferred on the family tree, flowing from the past (the protolanguage, from which relatedness and similarity ultimately emanate) toward the future (observed languages, which tend to diverge from each other due to language change), allowing the detection of the temporal ordering of changes in feature values. Thus, such methods allow one to draw causal inferences from apparently correlational data.

These two studies raise the issue of attempts to reconstruct the deep history of languages spoken today. Among the overall body of recent work we are reviewing here, the papers that have tended to attract the harshest criticism are those such as Gray & Atkinson (2003), Dunn et al. (2005, 2011), and Bouckaert et al. (2012), which attempt to use innovative quantitative methods to discover deep relationships between languages and to work out the age, location, and spread of language families such as Indo-European. Orthodox historical linguists have been scathing in their dismissal of the conclusions based on these methods, and quick to point out what seems to be rank ignorance of basic linguistic concepts, such as cognates, in this work (e.g., Thomason 2013). Again, it is clear that some of this work has significant flaws, and here, as in the other work we review, it is important for critiques based in linguistics to be heard.

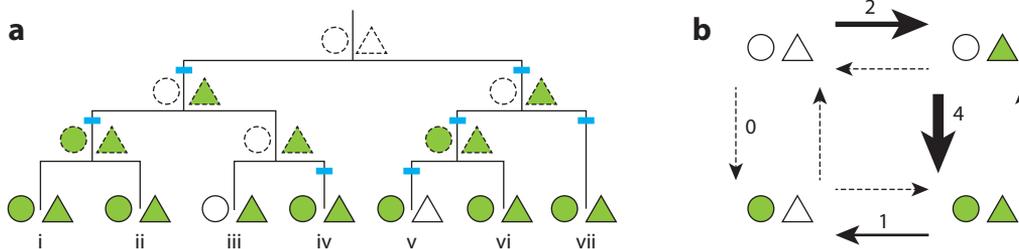


Figure 3

Correlated evolution. (a) A family with seven languages (*i–vii*) connected through a tree of common descent (*black lines*) from their protolanguage at the top. Also shown for each language, including the possibly unattested intermediate protolanguages (*dotted outline*), are two binary features (linguistic characteristics such as OV/VO and PrepN/NPostp), symbolized as a circle (*first feature*) and a triangle (*second feature*), each with two possible values (*white or green interior*). Each language in the family has a pair of values for these two features. Due to various processes, these feature values change through time (*top to bottom*). Each branch on which there is a feature change is marked by a short blue bar. (b) Summary of the history shown in panel *a*, showing all the possible changes of state from one pair of feature values to another. The probabilities of each change are indicated by width proportional to the actual number of changes that occur in the example tree in panel *a*. Unobserved or nonoccurring changes (those with count zero) are represented by thin dotted arrows. Clearly, the circle tends to become green following a similar change in the triangle.

However, these studies draw attention to an important dilemma. The comparative method is unquestionably the best way we currently have to reconstruct linguistic prehistory, but it requires very high standards that are not practically attainable for many parts of the world, and more importantly, it has an intrinsically limited time depth. Unless we want to declare that any history beyond the reach of the comparative method is permanently unknowable, we need to explore other ways of extending our ability to look into the past.

In saying this, we do not mean to suggest that anything goes; for example, we are well aware of the problems with what Greenberg called “mass comparison” in this approach to the linguistic prehistory of the Americas (Greenberg 1987, 1993). But the phylogenetic and other advanced statistical methods discussed here are much more principled, are more sophisticated in dealing with confounds, and can be validated on the basis of language families established by orthodox techniques. They also enable linguistic data and expertise to be effectively used in wider debates involving genetics, archeology, ecology, and economics so as to shed light on very important aspects of human history and evolution (e.g., Gray et al. 2009, Bouckaert et al. 2012, Sicoli & Holton 2014). We therefore believe that it makes sense to use them to explore further back than is possible on the basis of the orthodox linguistic reconstruction techniques alone.

6. SUMMARY AND PROSPECT

Because this article has been written as a contribution to the *Annual Review of Linguistics*, we have been at pains to address the relevance of the work we discuss for mainstream scholarly activity in linguistics and to highlight the importance of linguistically informed criticism to refining the work we discuss. We do not ignore the practical problem that few linguists have the necessary background in mathematics and statistics to evaluate some of the methods used in the studies we have reviewed here. Although it seems likely that acquiring a more sophisticated understanding of statistics will eventually become a part of ordinary training in linguistics, there will always be a gap between the kind of knowledge that is most useful for, say, a specialist in the history of the Indo-Aryan languages and that which might be needed to carry out, for example, a large-scale survey of the syntactic manifestations of aphasia in different languages of the world. For this reason we focus

on the value of theoretical critiques and emphasize the importance of interdisciplinary collaboration. Some of the correlational studies we review are manifestly flawed, and correlational work by its very nature must sometimes dispense with subtleties of specific languages in order to make the variables of interest mathematically tractable. Moreover, as we note at the beginning of Section 4, it seems clear that some of these studies implicitly question central tenets of modern linguistics, and it is likely that this is part of the reason for their skeptical reception. Nevertheless, we believe that correlational studies based on ever-larger and more accurate databases have the potential to make significant contributions to our understanding.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

S.R. is supported by the Interactional Foundations of Language project within the Language and Cognition Department at the Max Planck Institute for Psycholinguistics. D.D. is funded by the NWO (Netherlands Organisation for Scientific Research) through Vidi grant 276-70-022. D.R.L. is grateful for continued access to the University of Edinburgh's research infrastructure as Professor Emeritus. For comments on earlier drafts and pointers to literature, we thank Peter Ackema, Michael Dunn, Nik Gisborne, James Kirby, Bettelou Los, Antonella Sorace, and one anonymous reviewer.

LITERATURE CITED

- Atkinson QD. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332:346–49
- Atkinson QD. 2012a. Response to comment on “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa.” *Science* 335:1042
- Atkinson QD. 2012b. Response to comments on “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa.” *Science* 335:657
- Atkinson QD, Meade A, Venditti C, Greenhill SJ, Pagel M. 2008. Languages evolve in punctuational bursts. *Science* 319:588
- Atkinson QD, Nicholls G, Welch D, Gray RD. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Trans. Philol. Soc.* 103:193–219
- Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59:390–412
- Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, et al. 2007. The English lexicon project. *Behav. Res. Methods* 39:445–59
- Baronchelli A, Gong T, Puglisi A, Loreto V. 2010. Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci. USA* 107:2403–7
- Beebe H, Hitchcock C, Menzies P. 2009. *The Oxford Handbook of Causation*. Oxford, UK: Oxford Univ. Press
- Bentz C, Winter B. 2013. Languages with more second language learners tend to lose nominal case. *Lang. Dyn. Change* 3:1–27
- Berlin B, Kay P. 1969. *Basic Color Terms: Their Evolution and Universality*. Berkeley: Univ. Calif. Press
- Betti L, Balloux F, Amos W, Hanihara T, Manica A. 2009. Distance from Africa, not climate, explains within-population phenotypic diversity in humans. *Proc. Biol. Sci.* 276:809–14

- Biberauer T, Holmberg A, Roberts I. 2014. A syntactic universal and its consequences. *Linguist. Inq.* 45:169–225
- Boas F. 1931. *The Mind of Primitive Man*. Paris: Ed. Artis. Devereaux
- Bornstein MH. 1973. Color vision and color naming: a psychophysiological hypothesis of cultural difference. *Psychol. Bull.* 80:257–85
- Boucher J, Osgood CE. 1969. The Pollyanna hypothesis. *J. Verbal Learn. Verbal Behav.* 8:1–8
- Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337:957–60
- Bowern C, Epps P, Gray R, Hill J, Hunley K, et al. 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLOS ONE* 6:e25195
- Brown AM, Lindsey DT. 2004. Color and language: worldwide distribution of Daltonism and distinct words for “blue.” *Vis. Neurosci.* 21:409–12
- Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–95
- Bybee J, ed. 2011. The vanishing phonemes debate apropos of Atkinson 2011. *Linguist. Typol.* 15(spec. issue): 147–534
- Carp J. 2012. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Front. Neurosci.* 11:149
- Carroll M, von Stutterheim C, Nuese R. 2004. The language and thought debate: a psycholinguistic approach. In *Multidisciplinary Approaches to Language Production*, ed. T Pechmann, CHabel, pp. 183–218. Berlin: Mouton de Gruyter
- Chen MK. 2013. The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets. *Am. Econ. Rev.* 103:690–731
- Clancy SJ. 2006. The topology of Slavic case: semantic maps and multidimensional scaling. *Glossos* 7:1–28
- Croft W, Poole KT. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theor. Ling.* 34:1–37
- Crystal D. 2010. *The Cambridge Encyclopedia of Language*. Cambridge, UK: Cambridge Univ. Press. 3rd ed.
- Currie TE, Greenhill SJ, Mace R. 2010. Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philos. Trans. R. Soc. B* 365:3903–12
- Currie TE, Meade A, Guillon M, Mace R. 2013. Cultural phylogeography of the Bantu languages of sub-Saharan Africa. *Proc. Biol. Sci.* 280:20130695
- Cysouw M. 2008. Generalizing scales. In *Scales: A Cross-Disciplinary Perspective on Referential Hierarchies*, ed. I Bornkessel, A Malchukov, M Richard, pp. 379–96. Berlin: Mouton de Gruyter
- Cysouw M, Comrie B. 2008. How varied typologically are the languages of Africa? In *The Cradle of Language*, ed. R Botha, C Knight, pp. 189–203. Oxford, UK: Oxford Univ. Press
- Cysouw M, Dediu D, Moran S. 2012. Comment on “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa.” *Science* 335:657
- Dahl Ö. 2013. Stuck in the futureless zone. *Linguistics Diversity Comment Blog*, March 9. <http://dlc.hypotheses.org/360>
- Dale R, Lupyan G. 2012. Understanding the origins of morphological diversity: the linguistic niche hypothesis. *Adv. Complex Syst.* 15:1150017
- Davidoff J, Davies I, Roberson D. 1999. Colour categories in a stone-age tribe. *Nature* 398:203–4
- Dediu D. 2011. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proc. Biol. Sci.* 278:474–79
- Dediu D, Ladd DR. 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. USA* 104:10944–49
- Dediu D, Levinson SC. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLOS ONE* 7:e45198
- Donohue M, Nichols J. 2011. Does phoneme inventory size correlate with population size? *Linguist. Typol.* 15:161–70
- Dowman M. 2007. Explaining color term typology with an evolutionary model. *Cogn. Sci.* 31:99–132

- Dryer MS. 2013a. Relationship between the order of object and verb and the order of adjective and noun. See Dryer & Haspelmath 2013, <http://wals.info/chapter/97>
- Dryer MS. 1989. Large linguistic areas and language sampling. *Stud. Lang.* 13:257–92
- Dryer MS. 2011. The evidence for word order correlations: a response to Dunn, Greenhill, Levinson and Gray’s paper in *Nature. Linguist. Typol.* 15:335–80
- Dryer MS. 2013b. Relationship between the order of object and verb and the order of adposition and noun phrase. See Dryer & Haspelmath 2013, <http://wals.info/chapter/95>
- Dryer MS. 2013c. Relationship between the order of object and verb and the order of relative clause and noun. See Dryer & Haspelmath 2013, <http://wals.info/chapter/96>
- Dryer MS, Haspelmath M, ed. 2013. *The World Atlas of Language Structures Online*. Leipzig, Ger.: Max Planck Inst. Evol. Anthropol. <http://wals.info>
- Dunn M, Greenhill SJ, Levinson SC, Gray RD. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473:79–82
- Dunn M, Levinson SC, Lindström E, Reesink G, Terrill A. 2008. Structural phylogeny in historical linguistics: methodological explorations applied in island Melanesia. *Language* 84:710–59
- Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072–75
- Easwaran K. 2013. *Bayesianism*. Oxford Bibliographies Online: Philosophy. <http://www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-9780195396577-0204.xml>
- Edgington E, Onghena P. 2007. *Randomization Tests*. London: Chapman & Hall/CRC. 4th ed.
- Ember C, Ember M. 2007. Climate, econiche, and sexuality: influences on sonority in language. *Am. Anthropol.* 109:180–85
- Ember M. 1978. Size of color lexicon: interaction of cultural and biological factors. *Am. Anthropol.* 80:364–67
- Everett C. 2013. Evidence for direct geographic influences on linguistic sounds: the case of ejectives. *PLOS ONE* 8:e65275
- Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Assoc.
- Fought JG, Munroe RL, Fought CR, Good EM. 2004. Sonority and climate in a world sample of languages: findings and prospects. *Cross-Cult. Res.* 38:27–51
- Gay V, Santacreu-Vasut E, Shoham A. 2013. *The grammatical origins of gender roles*. Work. pap., Berkeley Econ. Hist. Lab. Pap. Ser.
- Gelman A, Shalizi CR. 2013. Philosophy and the practice of Bayesian statistics. *Br. J. Math. Stat. Psychol.* 66:8–38
- Gray RD, Atkinson QD. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435–39
- Gray RD, Bryant D, Greenhill SJ. 2010. On the shape and fabric of human history. *Philos. Trans. R. Soc. B* 365:3923–33
- Gray RD, Drummond AJ, Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323:479–83
- Greenberg JH. 1966. *Universals of Language*. Cambridge, MA: MIT Press
- Greenberg JH. 1987. *Language in the Americas*. Stanford, CA: Stanford Univ. Press
- Greenberg JH. 1993. Observations concerning Ringe’s “Calculating the Factor of Chance in Language Comparison.” *Proc. Am. Philos. Soc.* 137:79–90
- Greenhill SJ, Atkinson QD, Meade A, Gray RD. 2010. The shape and tempo of language evolution. *Proc. Biol. Sci.* 277:2443–50
- Grossman E. 2013. The costs of ignoring language change. *Diversity Linguistics Comment Blog*, March 10. <http://dlc.hypotheses.org/371>
- Hammarström H. 2009. Sampling and genealogical coverage in the WALS. *Linguist. Typol.* 13:105–19
- Hardy JL, Frederick CM, Kay P, Werner JS. 2005. Color naming, lens aging, and grue: what the optics of the aging eye can teach us about color language. *Psychol. Sci.* 16:321–27
- Hawkins JA. 1979. Implicational universals as predictors of word order change. *Language* 55:618–48
- Hawkins JA. 1982. Cross-category harmony, X-bar and the predictions of markedness. *J. Linguist.* 18:1–35
- Hay J, Bauer L. 2007. Phoneme inventory size and population size. *Language* 83:388–400

- Hernan MA, Robins JM. 2014. *Causal Inference*. London: Chapman & Hall/CRC. <http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
- Hudson PT, Bergman MW. 1985. Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks. *J. Mem. Lang.* 24:46–58
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–14
- Hunley K, Bowern C, Healy M. 2012. Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proc. Biol. Sci.* 279:2281–88
- Jaeger TF, Graff P, Croft W, Pontillo D. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguist. Typol.* 15:281–319
- Jaeger TF. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59:434–46
- Janssen DP, Bickel B, Zúñiga F. 2006. Randomization tests in language typology. *Linguist. Typol.* 10:419–40
- Juola P. 1998. Measuring linguistic complexity: the morphological tier. *J. Quant. Linguist.* 5:206–13
- Kiparsky P. 2014. *New Perspectives in Historical Linguistics: The Routledge Handbook of Historical Linguistics*. London/New York: Routledge
- Knooihuizen R, Dediu D. 2012. Historical demography and historical sociolinguistics: the role of migrant integration in the development of Dunkirk French in the 17th century. *Lang. Dyn. Change* 2:1–33
- Köhler R. 1986. *Zur linguistischen Synergetik*. Bochum, Ger.: Strukt. Dyn. Lexik
- Köhler R. 1999. Der Zusammenhang zwischen Lexemlänge und Polysemie im Maori. In *Pange lingua. Zborník na počest' Viktora Krupu*, ed. S Ondrejovič, J Genzor, pp. 27–33. Bratislava: Veda
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12:535–40
- Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. 2013. Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 44:978–90
- Laeng B, Brennen T, Elden K, Gaare Paulsen H, Banerjee A, Lipton R. 2007. Latitude-of-birth and season-of-birth effects on human color vision in the Arctic. *Vis. Res.* 47:1595–607
- Levinson SC, Meira S. 2003. ‘Natural concepts’ in the spatial topological domain—adpositional meanings in crosslinguistic perspective. An exercise in semantic typology. *Language* 79:485–516
- Levinson SC, Greenhill SJ, Gray RD, Dunn M. 2011. Universal typological dependencies should be detectable in the history of language families. *Linguist. Typol.* 15:509–34
- Levinson SC. 2000. Yéli Dnye and the theory of basic color terms. *J. Linguist. Anthropol.* 10:3–55
- Levinson SC. 2012. Foreword. In *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. JB Carroll, SC Levinson, P Lee, pp. vii–xxiii. Cambridge, MA: MIT Press. 2nd ed.
- Lewis MW, Pereltsvaig A. 2013. Ejectives, high altitudes, and grandiose linguistic hypotheses. *GeoCurrents Blog*, June 14. <http://languagesoftheworld.info/geolinguistics/ejectives-high-altitudes-grandiose-linguistic-hypotheses.html>
- Liberman M. 2013. High-altitude ejectives. *Language Log Blog*, June 14. <http://languageblog ldc.upenn.edu/nll/?p=4685>
- Lin H, Luo MR, MacDonald LW, Tarrant AW. 2001. A cross-cultural colour-naming study. Part I: Using an unconstrained method. *Color Res. Appl.* 26:40–60
- Little H. 2012. The role of foreigner-directed speech in language evolution. In *Proceedings of the 2012 European Human Behaviour and Evolution Association Annual Meeting*, p. 7. Durham, UK: Univ. Durham. http://community.dur.ac.uk/jeremy.kendal/EHBEA2012/Talks%20&%20Posters_files/PosterBookV2.pdf
- Loreto V, Mukherjee A, Tria F. 2012. On the origin of the hierarchy of color names. *Proc. Natl. Acad. Sci. USA* 109:6819–24
- Lupyan G, Dale R. 2010. Language structure is partly determined by social structure. *PLOS ONE* 5:e8559
- Mace R, Pagel M. 1994. The comparative method in anthropology. *Curr. Anthropol.* 35:549–64
- Majid A, Boster JS, Bowerman M. 2008. The cross-linguistic categorization of everyday events: a study of cutting and breaking. *Cognition* 109:235–50

- Malt BC, Ameer E, Imai M, Gennari SP, Saji N, Majid A. 2013. Human locomotion in languages: constraints on moving and meaning. *J. Mem. Lang.* 74:107–23
- Malt BC, Gennari SP, Imai M, Ameer E, Saji N, Majid A. 2014. Where are the concepts? What words can and can't reveal. In *Concepts: New Directions*, ed. E Margolis, S Laurence. Cambridge, MA: MIT Press. In press
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–20
- Martowicz A. 2011. *The origin and functioning of circumstantial clause linkers: a cross-linguistic study*. PhD thesis, Univ. Edinburgh. 419 pp.
- McMahon A, Heggarty P, McMahon R, Maguire W. 2007. The sound patterns of Englishes: representing phonetic similarity. *Engl. Lang. Linguist.* 11:113–42
- McMahon A, McMahon R. 2005. *Language Classification by Numbers*. Oxford, UK: Oxford Univ. Press
- McWhorter J. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. New York: Oxford Univ. Press
- Moran S, McCloy D, Wright R. 2012. Revisiting population size vs phoneme inventory size. *Language* 88:877–93
- Nelson-Sathi S, List J-M, Geisler H, Fangerau H, Gray RD, et al. 2010. Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proc. Biol. Sci.* 278:1794–803
- Nerbonne J. 2009. Data-driven dialectology. *Lang. Linguist. Compass* 3:175–98
- Nettle D. 1999. *Linguistic Diversity*. Oxford, UK: Oxford Univ. Press
- Nettle D. 2012. Social scale and structural complexity in human languages. *Philos. Trans. R. Soc. B* 367:1829–36
- Nichols J. 1992. *Linguistic Diversity in Space and Time*. Chicago: Univ. Chicago Press
- Page M, Atkinson QD, Meade A. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449:717–21
- Pearl J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge Univ. Press
- Piantadosi S, Tily H, Gibson E. 2011a. Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. USA* 108:3526–29
- Piantadosi S, Tily H, Gibson E. 2011b. Reply to Reilly and Kean: clarifications on word length and information content. *Proc. Natl. Acad. Sci. USA* 108:e109
- Piantadosi S. 2014. Zipf's word-frequency law in natural language: a critical review and future directions. *Psychon. Bull. Rev.* 21:1112–30
- Pierrehumbert J. 2002. Word-specific phonetics. In *Laboratory Phonology 7*, ed. C Gussenhoven, N Warner, pp. 101–39. Berlin: Mouton de Gruyter
- Pullum GK. 2012. Keith Chen, Whorfian economist. *Language Log Blog*, Feb. 9. <http://languageblog.ldc.upenn.edu/nll/?p=3756>
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. USA* 102:15942–47
- Reilly J, Kean J. 2007. Formal distinctiveness of high- and low-imageability nouns: analyses and theoretical implications. *Cogn. Sci.* 31:157–68
- Roberts SG, Winters J. 2012. Social structure and language structure: the new nomothetic approach. *Psychol. Lang. Commun.* 16:89–112
- Roberts SG, Winters J. 2013. Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLOS ONE* 8:e70902
- Rosch E. 1974. Linguistic relativity. In *Human Communication: Theoretical Perspective*, ed. E Silverstein, pp. 95–121. Hillsdale, NJ: Erlbaum
- Saunders BA, van Brakel J. 1997. Colour: an exosomatic organ? *Behav. Brain Sci.* 20:212–20
- Sereno SC, O'Donnell PJ, Sereno ME. 2009. Size matters: Bigger is faster. *Q. J. Exp. Psychol.* 62:1115–22
- Sicoli MA, Holton G. 2014. Linguistic phylogenies support back-migration from Beringia to Asia. *PLOS ONE* 9:e91722

- Silver M, Montana G, Nichols TE. 2011. False positives in neuroimaging genetics using voxel-based morphometry data. *Neuroimage* 54:992–1000
- Sinnemäki K. 2009. Complexity and size of speech community. In *Language Complexity as an Evolving Variable*, ed. PT Sampson, D Gil, pp. 126–41. Oxford, UK: Oxford Univ. Press
- Spruit MR. 2006. Measuring syntactic variation in Dutch dialects. *Lit. Linguist. Comput.* 21:493–506
- Steels L, McIntyre A. 1999. Spatially distributed naming games. *Adv. Complex Syst.* 1:301–24
- Sullivan J, McMahon A. 2010. Phonetic comparison, varieties, and networks: Swadesh’s influence lives on here too. *Diachronica* 27:325–40
- Swadesh M. 1955. Towards greater accuracy in lexicostatistic dating. *Int. J. Am. Linguist.* 21:121–37
- Szmrecsanyi B. 2014. Forests, trees, corpora, and dialect grammars. In *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*, ed. B Szmrecsanyi, B Wälchli, pp. 89–212. Berlin: Mouton de Gruyter
- Talbott, W. 2013. Bayesian epistemology. In *The Stanford Encyclopedia of Philosophy*, ed. EN Zalta. Stanford, CA: Stanford Univ. <http://plato.stanford.edu/archives/fall2013/entries/epistemology-bayesian/>
- Thomason SG. 2000. Linguistic areas and language history. *Stud. Slav. Gen. Linguist.* 28:311–27
- Thomason SG. 2013. Ultraconserved words? Really?? *Language Log Blog*, May 8. <http://languageblog.ldc.upenn.edu/nll/?p=4612>
- Torreira H, Roberts S. 2014. *Functional trade-off between lexical tone and intonation: typological evidence from polar-question marking*. Presented at Int. Symp. Tonal Asp. Lang. (TAL 2014), 4th, Nijmegen, Neth.
- Trudgill P. 2002. *Sociolinguistic Variation and Change*. Washington, DC: Georgetown Univ. Press
- Verkerk A. 2014. The correlation between motion event encoding and path verb lexicon size in the Indo-European language family. *Folia Linguist. Hist.* 35:307–58
- Vul E, Harris C, Winkielman P, Pashler H. 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* 4:274–90
- Walker I, Hulme C. 1999. Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* 25:1256–71
- Wang C-C, Ding Q-L, Tao H, Li H. 2012. Comment on “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa.” *Science* 335:657
- Wichmann S, Holman EW. 2009. Population size and rates of language change. *Hum. Biol.* 81:259–74
- Wichmann S, Rama T, Holman EW. 2011. Phonological diversity, word length and population sizes across languages: the ASJP evidence. *Linguist. Typol.* 15:177–97
- Yang C. 2013. Ontogeny and phylogeny of language. *Proc. Natl. Acad. Sci. USA* 110:6324–27
- Yao B, Vasiljevic M, Weick M, Sereno ME, O’Donnell PJ, et al. 2013. Semantic size of abstract concepts: It gets emotional when you can’t see it. *PLOS ONE* 8:e75000
- Zipf G. 1936. *The Psychobiology of Language*. London: Routledge
- Zipf G. 1949. *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley



Contents

Suppletion: Some Theoretical Implications <i>Jonathan David Bobaljik</i>	1
Ditransitive Constructions <i>Martin Haspelmath</i>	19
Quotation and Advances in Understanding Syntactic Systems <i>Alexandra D'Arcy</i>	43
Semantics and Pragmatics of Argument Alternations <i>Beth Levin</i>	63
Events and Situations <i>Sandro Zucchi</i>	85
Vagueness and Imprecision: Empirical Foundations <i>Stephanie Solt</i>	107
Cross-Linguistic Temporal Reference <i>Judith Tonhauser</i>	129
Variation in Information Structure with Special Reference to Africa <i>Tom Güldemann, Sabine Zerbian, and Malte Zimmermann</i>	155
Diachronic Semantics <i>Ashwini Deo</i>	179
The Indo-European Homeland from Linguistic and Archaeological Perspectives <i>David W. Anthony and Don Ringe</i>	199
Correlational Studies in Typological and Historical Linguistics <i>D. Robert Ladd, Seán G. Roberts, and Dan Dediu</i>	221
Advances in Dialectometry <i>Martijn Wieling and John Nerbonne</i>	243

Sign Language Typology: The Contribution of Rural Sign Languages <i>Connie de Vos and Roland Pfau</i>	265
Genetics and the Language Sciences <i>Simon E. Fisher and Sonja C. Vernes</i>	289
Language Abilities in Neanderthals <i>Sverker Johansson</i>	311
How Nature Meets Nurture: Universal Grammar and Statistical Learning <i>Jeffrey Lidz and Annie Gagliardi</i>	333
Bringing Machine Learning and Compositional Semantics Together <i>Percy Liang and Christopher Potts</i>	355
Bilingualism, Mind, and Brain <i>Judith F. Kroll, Paola E. Dussias, Kinsey Bice, and Lauren Perrotti</i>	377
Taking the Laboratory into the Field <i>D.H. Whalen and Joyce McDonough</i>	395

Errata

An online log of corrections to *Annual Review of Linguistics* articles may be found at <http://www.annualreviews.org/errata/linguistics>