14

ON THE NECESSITY OF AN INTERDISCIPLINARY APPROACH TO LANGUAGE UNIVERSALS

Florencia Reali and Morten H. Christiansen

14.1. Toward an Interdisciplinary Theory of Language Universals

There is considerable variation across the languages of the world, nonetheless it is possible to discern common patterns in how languages are structured and used. The underlying source of this variation as well as the nature of crosslinguistic universals is the focus of much debate across different areas of linguistics. Some linguists suggest that language universals derive from the inner workings of Universal Grammar (UG)—a set of innate constraints on language acquisition (e.g., see Bever, Chapter 6; Hornstein & Boeckx, Chapter 5; and Pinker & Jackendoff, Chapter 7). Others see universals as emerging from patterns of language use, primarily because of processes taking place over diachronic rather than synchronic timescales (Bybee, Chapter 2; Hawkins, Chapter 4). Yet, other linguists propose that universals may derive from some combination of language acquisition and use (Hurford, Chapter 3). Even within the same theoretical linguistic framework, there is often little agreement about what the exact universals are. For example, when surveying specific universals proposed by different proponents of UG, Tomasello (2004) found little overlap between proposed universals.

We believe that a resolution to this debate is unlikely to be forthcoming from within linguistics itself; instead, it must be sought by adopting an interdisciplinary approach to language universals, integrating linguistic insights with those of other relevant disciplines. Thus, we need to understand the possible biological bases for language universals (see Clark & Misyak, Chapter 12; Finlay, Chapter 13; and Müller, Chapter 11, for discussion), their potential psychological underpinnings (Bever, Chapter 6), and how they may relate to semantics, computation, and learnability

December 8, 2008 20:0 OUP/LAUN Page-266 9780195305432_0266-0277_Christiansen_LAUN_ch14

(Bach & Chao, Chapter 8; Stabler, Chapter 10; and Steedman, Chapter 9), just to mention a few key constraining factors on a broad theory of language universals. Importantly, though, such a theory will also have to take seriously the widespread diversity that can be observed across the languages of the world in terms of phonology, morphology and syntax (Evans & Levinson, to appear). In this chapter, we discuss a single case in which a broader interdisciplinary approach may help shed additional light on the UG perspective on language universals. We note, however, that other approaches to universals are likely also to benefit from a broader interdisciplinary perspective.

14.2. Language Universals and Universal Grammar

Boeckx (2006) describe the research agenda of generative linguistics, including the ways in which its specific aims have evolved over the past five decades up to the current Minimalist Program (see also Bever, Chapter 6). Since Chomsky (1965), generative linguistics has been explicitly grounded on the assumption of an innate linguistic endowment, providing the basis for language acquisition, a UG. On this account, language universals derive from the properties of UG. The necessity of UG rests primarily on the Poverty of Stimulus (POS) argument for innateness of linguistic-specific constraints. Originally proposed by Chomsky (1980a, b), POS is based on the assumption that the information in the linguistic environment is too impoverished for language to be learnable. As noted by Boeckx (2006) and others, the logic of the argument is powerful: If the data in the primary linguistic input is insufficient for correct grammatical generalization, then language acquisition requires an endogenous biological explanation. If the premises are valid, the conclusion seems unavoidable. However, a critical appraisal of POS inevitably brings up the crucial question raised by Boeckx (2006): How good are the premises?

Until recently, the POS premises have been taken for granted based on intuitive observations. Here, however, we argue that one of the weaknesses of the argument stems from the difficulty in assessing the informativeness of the input, and from the imprecise and intuitive definition of what counts as "insufficient information" available to the learner. Moreover, we underscore the need for an interdisciplinary approach to POS, where no discipline is primary. Along these lines, we describe recent research in cognitive science that has begun to posit serious theoretical challenges to the fundamental assumptions of POS. In particular, we discuss studies that have contested the traditional views by focusing on the paradigmatic linguistic example used by Boeckx (2006) to illustrate POS: Auxiliary fronting in complex Yes/No interrogatives.

14.3. Learning Structure from Regularities

Recent work in cognitive science has begun to call POS assumptions into question, including its underlying assumptions about the nature of the linguistic input and the learning abilities of young infants. Much of this research has contributed to a substantial reappraisal of the role of statistical learning in language acquisition.

The ability to infer structure from statistical regularities in the input is a ubiquitous strategy throughout cognition (e.g., Goldstone, 2000; Markman & Gentner, 1993). Despite the growing bulk of work underscoring the role of statistical learning in perception and cognition, traditional generative linguistic approaches have argued over the past five decades that probabilistic information—including distributional, phonological, prosodic, and semantic cues—may be insufficient for acquisition of the rules of grammar (e.g., Chomsky, 1957; Crain & Pietroski, 2001; Fodor & Crowther, 2002; Hornstein & Lightfoot, 1981; Laurence & Margolis, 2001; Legate & Yang, 2002). Recent research in psycholinguistics, however, has started to demonstrate that distributional regularities may provide an important source of information for bootstrapping syntax (e.g., Mintz, 2002, 2003; Reali & Christiansen, 2005; Redington, Chater, & Finch, 1998; Solan, Horn, Ruppin, & Edelman, 2005). Moreover, distributional information is especially useful when it is integrated with other probabilistic cues such as prosodic or phonological information (e.g., Monaghan, Christiansen, & Chater, 2007; Morgan, Meier, & Newport, 1987; Reali, Christiansen, & Monaghan, 2003).

Behavioral studies over the last decade have shown that young infants are quite competent statistical learners (for reviews, see Gómez & Gerken, 2000; Saffran, 2003). For example, 8-month-old infants have access to powerful mechanisms to induce statistical regularities between linguistic elements (e.g., Gómez, 2002; Saffran, Aslin, & Newport, 1996; Saffran & Wilson, 2003), and by 1 year, children's perceptual attunement is likely to allow them to use language-internal probabilistic cues (Jusczyk, 1997). A recent line of research in natural language processing and connectionist modeling has revealed many properties of statistical learning of potential relevance for language acquisition (e.g., Christiansen & Chater, 1999; Elman, 1993; Lewis & Elman, 2001; Manning & Schütze, 1999). For instance, even though the primary linguistic input may be primarily characterized by a lack of *explicit* negative evidence, computational work suggest that learners could rely on *implicit* negative evidence, which may result from predictive learning algorithms (e.g., Elman, 1993; Rohde & Plaut, 1999). For example, Spivey-Knowlton and Saffran (1995) proposed a learning method that employs a type of feedback overlooked in most discussions on language learnability. In principle, a child could evaluate a general hypothesis about the target language by observing whether the *predictions* generated by the hypothesis are borne out in the speech she hears. As the child listens to others speak, she predicts that certain elements will follow one another. Thus, the child learns by listening to utterances rather than by producing them, and generates his or her own negative evidence by comparing the predicted input with the actual input (for further discussion, see Rohde & Plaut, 1999). Connectionist models such as Simple Recurrent Networks (SRNs; Elman, 1990) employ learning techniques that are consistent with prediction-feedback learning. When SRNs are trained to predict the next element in a sequential input (Elman, 1990, 1993), they produce implicit predictions regarding upcoming materials. By comparing a given prediction to the actual incoming input, the network produces an immediate error signal that can be functionally interpreted as implicit negative evidence derived from incorrect predictions.

During the 1980s, generative analysis of language learnability emphasized the unavailability of positive examples, shifting away from the focus on negative evidence (for further discussion, see MacWhinney, 2004). Thus, Chomsky's (1980a, b) statement of the poverty of stimulus argument applied to the case of multiclausal Yes/No questions relies on the notion of learning in the absence of positive evidence: "A person might go through much or all of his life without ever having been exposed to the relevant evidence, but he will nevertheless unerringly employ the structure-dependent generalization, on the first relevant occasion" (Chomsky, 1980a, p. 40). However, the notion of absence of positive evidence has been seriously contested by recent studies that indicate that child-directed speech may contain sufficient statistical information to distinguish between grammatical and ungrammatical multiclause Yes/No questions.

14.4. A Case Study: Auxiliary Fronting in Yes/No Questions

Auxiliary fronting in Yes/No questions is one of the most often cited examples used to illustrate the logic of POS argument. The ubiquity of this example is partly motivated by the study of Crain and Nakayama (1987), which provided empirical evidence suggesting that children only entertain structure-dependent hypotheses when they are prompted to produce multiclausal Yes/No questions. Moreover, Legate and Yang (2002) present corpus analyses of child-directed-speech indicating that relevant examples of grammatical Yes/No questions—that is, interrogatives containing an embedded "competing" auxiliary—appear to be extremely infrequent in the primary linguistic input. Specifically, they found that core examples constitute less than 1% of all sentences, and conclude that the information does not suffice for generalization, partly because the numbers suggest that examples may not be available to *every* child. Assuming that every child is capable of correct generalization, the necessity for a more endogenous, biological explanation seems to be needed.

The vast majority of literature discussion on POS has concentrated on whether examples of multiclause interrogatives such as, *Will the man who is tall leave now?*, are available to the child (e.g., Boeckx, 2006; Crain & Nakayama, 1987; Legate & Yang, 2002; cf. Pullum & Scholz, 2002; see Scholz & Pullum, 2002, for discussion). However, it has been recently proposed that such a characterization of what counts as relevant evidence may be too narrow, failing to take into account the possibility of implicit statistical information in the primary linguistic input (e.g., Lewis & Elman, 2001; Reali & Christiansen, 2005). These studies suggest that more *indirect sources* of statistical information may provide additional cues for making the appropriate grammatical generalizations. For example, Lewis and Elman (2001) trained SRNs on data from an artificial grammar that generated questions of the form, "auxiliary noun-phrase adjective?," and sequences of the form, "A_i noun phrase B_i" (where A_i and B_i represent a variety of different material). Crucially, the networks were not trained with core examples of multiclausal Yes/No interrogatives. Lewis and Elman found that the networks were better at making predictions for grammatical multiclause questions compared to ungrammatical ones involving incorrect auxiliary fronting. A possible caveat in this study, however, is that the networks were trained using an artificial grammar lacking the complexity of an actual child-directed speech corpus.

Recently, Reali and Christiansen (2005) conducted a series of corpus analyses of child-directed speech showing that there is indirect statistical information sufficient for distinguishing between grammatical and ungrammatical generalizations in multiclause Yes/No questions. First, they trained simple statistical models based on pairs (bigrams) and triples (trigrams) of words drawn from the Bernstein-Ratner (1984) corpus of child-directed speech. The Bernstein-Ratner corpus contains transcripts of speech from nine mothers to their children. The speech was recorded over a 4–5-month period when children were between 13 and 21 months of age. This corpus is relatively small and very noisy, mostly containing short sentences with simple grammatical structure. Importantly, there are no explicit examples of multiclause interrogatives in the corpus.

Bigram and trigram models (e.g., Jurafsky & Martin, 2000) measure how frequently pairs or triples of adjacent words occur in a linguistic corpus. Based on the probability of its fragments, the probability of a sentence can be calculated (for a more detailed description of the methods, see Reali & Christiansen, 2005). After training the bigram and trigram models, the authors compared the probabilities of a hundred test sentences that consisted of correct multiclause interrogatives (e.g., *Is the bunny that is on the chair sleeping?*) and their ungrammatical counterpart (e.g., **Is the bunny that on the chair is sleeping?*). Reali and Christiansen hypothesized that indirect statistical information in the form of word co-occurrences provided sufficient evidence for distinguishing grammatical from ungrammatical multiclause questions. In line with their prediction, they found a significant difference in the likelihood of these two alternative hypotheses: Grammatical versions were more probable than ungrammatical versions in more than 95% of the cases.

In a second series of analyses, they tested the bigram and trigram models trained on the Bernstein-Ratner corpus on the same sentences used in the Crain and Nakayama original study. In the Crain and Nakayama study, 3- to 5-year old children spontaneously produced sentences like, *Is the boy who is watching Mickey Mouse happy?*, and never produced sentences like, **Is the boy who watching Mickey Mouse is happy?*. Reali and Christiansen found that according to the bigram and trigram models the grammatical versions of the multiclause interrogatives were significantly more probable than their ungrammatical counterparts. In a subsequent series of simulation studies, they showed that simple learning devices, such as SRNs, were capable of exploiting the statistical cues captured by the bigram and trigram models. When trained on the full-blown child-directed speech corpus, the networks produced a bias toward grammatical multiclause questions when compared to their ungrammatical counterparts. The results indicate that a noisy child-directed speech corpus contains enough indirect statistical information to distinguish between grammatical and ungrammatical multiclause Yes/No questions.

Reali and Christiansen (2005) argue for a possible way in which exposure to fine-grain statistical information may translate into production biases. Crucially, the pattern of network predictions they found can be interpreted as providing statistical constraints on real-time production. Following previous connectionist work (Christiansen & Chater, 1999), they propose that the SRN's output predictions could be construed as a set of possible sentence *continuations* during production. For example, semantic factors being equal, after the speaker produces the fragment, *The boy who* . . . , she would be biased to continue the sentence using an auxiliary (e.g., *is*) rather than a verb in progressive form or an adjective (e.g., *watching* or *happy*). This is because chunks of the form "who is" are considerably more frequent than chunks of the form "who watching" or "who happy." They found that the Yes/No questions generated in this fashion are consistent with children production data found in Crain and Nakayama (1987).

Importantly, statistical regularities result from the nonrandomness in the distribution of linguistic elements. Thus, the importance of the underlying syntactic structure should not be underestimated. Linguistic structure is a prerequisite for statistical learning because it is the constituent properties of well-formed sentences that make distributional cues useful in the first place. Therefore, a language without reliable structural regularities would not be learnable from a statistical perspective. Reali and Christiansen concluded that sequential statistics could help explain why children tend not to make many auxiliary-fronting errors. Moreover, the model predicted that children should make fewer errors involving high-frequency word chunks compared to low-frequency ones. Interestingly, this prediction has been confirmed by a recent question elicitation study (Ambridge, Rowland, & Pine, 2008). For example, they found higher rates of auxiliary-doubling error for questions where such errors involved high-frequency word category combinations (e.g., more errors such as **Is the boy who is washing the elephant is tired?* than **Are the boys who are washing the elephant are tired?*).

Although these results only pertain to a single linguistic construction, on the theoretical side they point toward the necessity of a serious reassessment of the type of information that should be considered useful for learning a particular linguistic structure. More generally, the POS assumption may have to be revisited in the light of the statistical richness present in the primary linguistic input.

14.5. A Broader Perspective

A remaining question is where universal patterns of language structure derive from. It seems clear that at least of some aspects of language universals are determined by innate constraints. The key question, however, is whether these constraints are best characterized as being specifically linguistic[1] in nature, or whether they may derive from more general cognitive and perceptual constraints on learning and processing. Interdisciplinary work on the evolution of language supports the latter view (e.g., Batali, 2002; Brighton, 2002; Christiansen & Chater, 2008; Christiansen, Dale, Ellefson, & Conway, 2002; Christiansen, Reali, & Chater, 2006; Deacon, 1997; Kirby & Christiansen, 2003). According to this perspective, most language universals may derive from nonlinguistic constraints on the statistical learning mechanisms themselves and from general functional and pragmatic properties of communicative interactions. Additional common features of language might have emerged through processes of cultural transmission across generations of human learners and through grammaticalization (e.g., Bybee, Chapter 2; Hurford, Chapter 3; Givón, 1998; Heine & Kuteva, 2002). Thus, language could be regarded as "piggy-backing" on more general cognitive mechanisms adapted for other functions. These mechanisms, in turn, determine the cognitive constraints that are brought to bear on language acquisition.

The cognitive mechanisms used to learn language may be not qualitatively different from those used to learn other aspects of cognition and perception. According to this view, complex linguistic tasks would be performed using similar machinery to that used by other cognitive systems, providing a possible framework for a unification of theories of representation. This view is supported by research in neuroanatomy and neurophysiology, suggesting that similar architectures underlie language and other cognitive processes. For example, the study of cortical areas indicates that brain structures are quite homogenous across different functional areas. In fact, the cortex has been compared with a multidimensional plaid (Kingsbury & Finlay, 2001) that is more suitable for the implementation of fine-grained distributed architectures than for the implementation of computer-like modules functioning independently and interchanging discrete packets of information (see also Finlay, Chapter 13; Müller, Chapter 11; and Clark & Misyak, Chapter 12).

In sum, a remaining challenge for the language sciences is the question of whether our innate language-acquisition biases are better characterized as part of domain-specific or domain-general cognitive mechanisms. An effective research program designed to investigate the nature and constraints of cognitive mechanisms should be grounded within an interdisciplinary approach in which no single discipline is primary.

Key Further Readings

For reappraisals of POS arguments and the logical problem of language acquisition, we recommend the articles by Scholz and Pullum (2002) and Pullum and Scholz (2002). These two articles are part of *The Linguistic Review* special issue: A review of "The Poverty of stimulus argument." This special issue consists of a discussion paper by Geoffrey Pullum and Barbara Scholz with responding articles by various contributors. Pullum and Scholz provide an insightful reevaluation of POS argument when applied to various frequently used examples, including the case of plurals in noun–noun compounding, auxiliary sequences, anaphoric one, and Yes/No questions in English. We also recommend the discussion article by Brian MacWhinney (2004), *A multiple process solution to the logical problem of language acquisition* (and associated peer commentaries). In this paper, MacWhinney discusses alternatives to the UG hypothesis in the context of language acquisition, including item-based learning, indirect negative evidence, and multiple-cue integration.

For an empirical reevaluation of children's spontaneous production of multi-clause Yes/No questions, we recommend a recent article by Ambridge, Rowland, and Pine (2008). They found that children's errors in auxiliary fronting were consistent with input-based learning predictions.

For an introduction to linguistic alternatives to the generative framework, we recommend the book *Probabilistic Linguistics* edited by Bod, Jay, and Jannedy (2003). This book comprises the contribution of various authors, providing empirical and computational evidence for the probabilistic nature of linguistic behavior at various levels of representation, ranging from phonetics to discourse. The work covered in the book indicates that linguistic competence is far from discrete, challenging core assumptions of generative approaches and providing a new probabilistic framework for the study of language.

Acknowledgments

MHC was supported by a Charles A. Ryskamp Fellowship from the American Council of Learned Societies.

Note

1 The term "specifically linguistic innate knowledge" refers to *representational innateness* as defined in Elman et al. (1996). Representational innateness is the strongest and most specific form of linguistic nativism. It allows for an innately specified encoding of detailed grammatical knowledge (for discussion, see Chapter 7, Elman et al., 1996).

References

Ambridge, B., Rowland, C., & Pine, J. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex question production. *Cognitive Science*, *32*, 222–255.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among examples. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 111–172). Cambridge: Cambridge University Press.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language*, *11*, 557–578.

Bod, R., Jay, J. H., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods, and aims*. New York: Oxford University Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*, 25–54.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton & Co.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980a). The linguistic approach. In M. Piatelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Chomsky, N. (1980b). *Rules and representations*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–558.

Christiansen, M. H., Dale, R. A. C., Ellefson, M. R., & Conway, C. M. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp.165–187). London: Springer-Verlag.

Christiansen, M. H., Reali, F., & Chater, N. (2006). The Baldwin effect works for functional, but not arbitrary, features of language. In A. Cangelosi, A. Smith, & K. Smith (Eds.), *Proceedings of the sixth international conference on the evolution of language* (pp. 27–34). London: World Scientific Publishing.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, *63*, 522–543.

Crain, S., & Pietroski, P. (2001). Nature, nurture, and universal grammar. *Linguistics and Philosophy*, *24*, 139–186.

Deacon, T. W. (1997). *The symbolic species*: *The coevolution of language and the brain*. New York: WW Norton.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.

Evans, N., & Levinson, S. (to appear). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*.

Fodor, F. D., & Crowther, C. (2002). Understanding stimulus poverty arguments. *The Linguistic Review*, *19*, 105–145.

Givón, T. (1998). On the co-evolution of language, mind, and brain. *Evolution of Communication*, *2*, 45–116.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology*: *Human Perception and Performance*, *26*, 86–112.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

Heine, B., & Kuteva, T. (2002). On the evolution of grammatical forms. In A. Wray (Ed.), *Transitions to language* (pp. 376–397). Oxford, UK: Oxford University Press.

Hornstein, N., & Lightfoot, D. (Eds.). (1981). *Explanation in linguistics*. London: Longman.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.

Jusczyk, P. (1997). The discovery of spoken language. Cambridge, MA: MIT Press.

Kingsbury, M. A., & Finlay, B. L. (2001). The cortex in multidimensional space: Where do cortical areas come from? *Developmental Science*, *4*, 125–157.

Kirby, S., & Christiansen, M. H. (2003). From language learning to language evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 272–294). Oxford, UK: Oxford University Press.

Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for the Philosophy of Science*, *52*, 217–276.

Legate, J. A., & Yang, C. (2002). Empirical re-assessment of stimulus poverty arguments. *Linguistic Review*, *19*, 151–162.

Lewis, J. D., & Elman, J. L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the twenty-sixth annual Boston University conference on language development* (pp. 359–370). Somerville, MA: Cascadilla Press.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, *31*, 883–914.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, *30*, 678–686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child-directed speech. *Cognition*, *90*, 91–117.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Crosslinguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*, 498–550.

Pullum, G. K., & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review*, *19*, 9–50.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structural dependence and indirect statistical evidence. *Cognitive Science*, *29*, 1007–1028.

Reali, F., Christiansen, M. H., & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the twenty-fifth annual conference of the Cognitive Science Society* (pp. 970–975). Mahwah, NJ: Lawrence Erlbaum.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–469.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67–109.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

Saffran, J. R., Aslin, R., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy*, *4*, 273–284.

Scholz, B., & Pullum, G. K. (2002). Searching for arguments to support linguistic nativism. *Linguistic Review*, *19*, 185–223.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences*, *102*, 11629–11634.

Spivey-Knowlton, M., & Saffran, J. R. (1995). Inducing a grammar without an explicit teacher: Incremental distributed prediction feedback. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the seventeenth annual conference of the cognitive science society* (pp. 230–235). Hillsdale, NJ: Erlbaum.

Tomasello, M. (2004). What kind of evidence could refute the UG hypothesis? *Studies in Language*, *28*, 642–644.