Complementing quantitative typology with behavioral approaches: Evidence for typological universals

HARRY TILY and T. FLORIAN JAEGER

1. Introduction

Two main classes of theory have been advanced to explain correlations between linguistic features like those observed by Greenberg (1963). ARBITRARY CONSTRAINT theories argue that certain sets of features patterm together because they have a single underlying cause in the innate language faculty (e.g., the Principles and Parameters program; see Chomsky & Lasnik 1993). FUNCTIONAL theories argue that languages are less likely to have certain combinations of properties because, although possible in principle, they are harder to learn or to process, or less suitable for efficient communication (Hockett 1960, Bates & MacWhinney 1989, Hawkins 2004, Dryer 2007, Christiansen & Chater 2008; for further discussion see Hawkins 2007 and Jaeger & Tily 2011). The failure of Dunn, Greenhill, Levinson & Gray (2011) to find systematic feature correlations using their novel computational phylogenetic methods calls into question both of these classes of theory.

Dunn et al.'s methodology is a principled and powerful new way of evaluating change-based theories of language typology, but it faces fundamental challenges common throughout quantitative typology. Typological data is usually sparse, and additional data hard or impossible to obtain. The statistical power to detect an effect (e.g., to detect universal tendencies) is reduced by uncertainty about the genealogical relations between languages, the period of time during which a language was spoken, and the amount of contact with other languages. It is hence possible that the failure to find support for word order universals reported in Dunn et al. 2011 is a spurious null effect (see Croft et al. 2011, in this issue, who also discuss other potential methodological issues with that study). To some extent, this is a critique that can be leveled at much work on quantitative typology: sparse and uncertain data weakens the conclusions that can be drawn from it. If we desire a higher level of certainty when evaluat-

Linguistic Typology 15 (2011), 497–508
DOI 10.1515/LITY.2011.033
1430–0532/2011/015-0497

ing our theories, we must necessarily corroborate inferences from typological distribution with other sources of evidence.

Here, we discuss an alternative BEHAVIORAL approach to the study of linguistic universals that offers a powerful way to detect subtle universal biases on acquisition, where typological data is perhaps too sparse to detect such differences under appropriate statistical control. Experiments employing so-called Artificial Language Learning (ALL: MacWhinney 1983, Morgan et al. 1987, Hudson Kam & Newport 2005) and the recently flourishing Iterated Artificial Language Learning paradigm (IALL: Esper 1966, Kirby et al. 2008, Griffiths et al. 2008) speak to the question that Dunn et al. set out to answer. As we describe below, these two paradigms make it possible to test hypothetical universals directly against data from language acquisition and language transmission over generations. Both the computational phylogenetic method and (I)ALL methods have their strengths and weaknesses, some of which are complementary. We argue that the best way to adjudicate between competing theories will be to consider evidence from multiple methodologies together. We first introduce the ALL paradigm, then consider its extension IALL. Finally, we discuss the complementary strengths and weaknesses of quantitative typology and (I)ALL relative to the question of interest.

2. Artificial Language Learning

During ALL experiments, infants, children, or adults learn artificially constructed miniature languages by playing a game or watching videos described in the language. During and after training in these languages, a variety of measures are collected that allow researchers to assess both receptive and the productive competence. For example, to assess receptive competence, learners might have to judge which of two scenes corresponds to a sentence they hear or read (e.g., Wonnacott et al. 2008). Some experiments have employed self-paced reading (e.g., Amato & Macdonald 2010) or eye-tracking during auditory input (e.g., Wonnacott et al. 2008) in the artificial language. Similarly, to assess productive competence, learners might see a video which they are asked to describe in the artificial language (e.g., Hudson Kam & Newport 2005, 2009). Figure 1 shows example trials for exposure and receptive tests from an ALL experiment. The constructed languages under study may vary in complexity from some tens of morphemes/words describing visual properties of a set of referents to somewhat larger grammars containing different verb argument frames and other complex syntax. Training and test typically takes place in sessions lasting around an hour, often over several days (one to six days, commonly four for larger languages).

There are two common designs used in ALL experiments to identify an acquisition bias for or against a particular language feature. The first is a between-

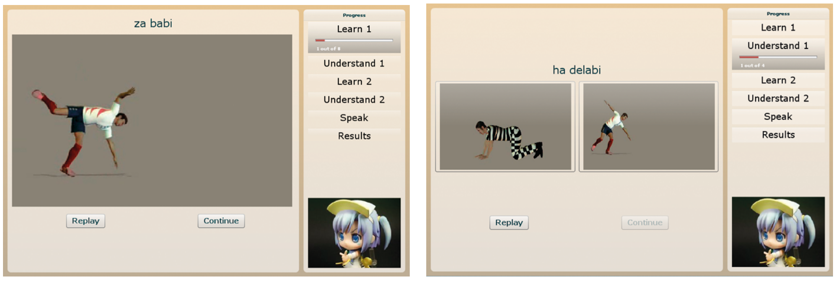

Figure 1. *A screenshot from an ALL experiment (see Tily et al. 2011) showing a training trial (left) and a receptive competence trial (right). Videos were accompanied by audio presentation of a description in the test language. Participants responded in the latter by clicking on the video that corresponded to the spoken description.*

participants design in which one group of participants is taught each language, and differences in their performance are compared between groups (e.g., Christiansen 2000). In the other, the quantity of interest is manipulated within participants, in a design that is sometimes referred to as the "mixture shift" paradigm (Culbertson & Smolensky in press). Here, participants are taught variable languages containing a mixture of both forms, and any tendency to perform better in comprehension of one of the forms or favor that form in production is interpreted as evidence for a bias (e.g., Culbertson & Smolensky in press; Hudson Kam & Newport 2005, 2009). The second design is often preferable where possible because it controls for individual differences between participants.

ALL methods have a long history in psychology, and many ALL studies have identified processes familiar from language change in the lab. This confirms the validity of the method for studying the role of the learner in change. For instance, children tend to regularize inconsistency in their input in much the same way as is observed in the first generation of creole languages (Hudson Kam & Newport 2005, 2009). As observed in natural language use and change, ALL learners can acquire probabilistic tendencies for specific verbs to appear in certain argument frames ("verb biases") but regularize these depending on their frequency and the frequency of their verb class (Wonnacott et al. 2008).

Several recent ALL studies have been motivated by the question of typological distribution. Most relevant to the feature correlations described by Dunn et al. is Experiment 1 in Christiansen 2000, which explored the learnability of languages with consistent and inconsistent branching. Participants in Christiansen's "consistent" condition learned a constructed language with OV order, where adpositions followed the noun. That is, both verb phrases and adposition phrases were consistently left-branching. The "inconsistent" language was also OV but had adpositions before the noun. The training phase consisted of simple visual exposure to strings generated from the grammars, without any accompanying meanings. In subsequent testing, participants in the consistent condition were significantly better at identifying grammatical sentences than inconsistent condition participants, who were no better than a third group of control participants receiving no training at all. These results support the hypothesized universal correlation between verb-object and adposition-noun order, which arises through acquisition biases or constraints. Crucially, these results cannot be due to superficial transfer from the native language of participants into the artificial language: all participants in the experiment were monolingual speakers of English, in which verb and adposition phrases are consistently right-branching.

More recent work has examined similar questions, using larger grammars, video stimuli to provide meaning, and audio presentation of the language. Culbertson & Smolensky (in press) test Greenberg's Universal 18, which predicts word order harmony in numeral-noun and adjective-noun order. Teaching variable order languages using a video game-like interface, they find that learners are likely to regularize harmonic patterns (noun-adjective & noun-numeral or adjective-noun & numeral-noun) and particularly disprefer the typologically most infrequent feature combination (adjective-noun & noun-numeral).

In Tily et al. 2011, we tested hypothesized universals about both basic verb-argument ordering (Greenberg's Universal 1, Greenberg 1963; see also Dryer 2007) and the relation between basic verb-argument ordering and the relative order of determiners and nouns (similar to Greenberg's Universal 4 on adposition ordering; see also Hawkins 1994, 1999, 2004). Participants were randomly assigned to learn one out of twelve artificial languages. The twelve languages crossed the six theoretically possible basic verb-argument orders (VSO, VOS, SVO, OVS, SOV, OSV) with the two possible orderings of determiners and nouns (Det-N, N-Det). Hence languages either had consistent headedness (V-O and Det-N, or O-V and N-Det) or inconsistent headedness. Participants performed better at acquiring languages with subject-object order, supporting Greenberg's Universal 1. In line with the results reported in Christiansen 2000 and Culbertson & Smolensky (in press), there also was evidence that languages with consistent headedness are more easily acquired. This latter evidence was, however, relatively weak. In ongoing work, we are testing whether a strong preference for consistent headedness emerges if inconsistent headedness results in considerably longer dependency lengths (as predicted by Hawkins 1994, 2004).

Several related studies have found that non-word-order typological universals, whether absolute or implicational, are reflected in learnability in the lab. In particular, there are now a large number of ALL explorations of phonological universal patterns. Velar palatalization (e.g., the change from [k] to [tʃ]) is more likely to be generalized by participants in phonological contexts where that change is more frequently attested in the world's languages (Wilson 2006).

Participants will learn non-adjacent dependencies between segments more easily than between entire syllables (Newport & Aslin 2004). Participants are worse at learning left-to-right vowel harmony systems than (typologically more frequent) right-to-left systems (Finley & Badecker 2009), and vowel harmony is learned more readily than consonant harmony overall (Nevins 2010). Morphosyntactic universals have also been tested. MacWhinney (1983) and Hupp et al. (2009) show that it is easier to learn suffixation systems than prefixation systems, mirroring the typologically asymmetry observed by Sapir (1921). Culbertson & Legendre (in press) find that learners generalize variable agreement marking from definite to indefinite NPs but not vice versa, mirroring the implicational typological universal saying that, in languages where the verb agrees with one type of subject NP, it must agree with all other types that are higher on a scale of definiteness. Fedzechkina et al. (2011) find that learners are more likely to regularize the word order of a variable order language if it does not have case marking, again mirroring typological distributions.

The co-dependence observed in ALL experiments for, for example, different aspects of word order (Christiansen 2000, Culbertson & Smolensky in press) and case and word order (Fedzechkina et al. 2011) is exactly the type of co-development of grammatical features that typologists refer to as implicational universals. Hence, the experiments reported here provide evidence for the type of implicational word order universals that Dunn et al. failed to find in their data.

3. Iterated Artificial Language Learning

A natural extension of the ALL method is to move from a single "generation" of participants learning and using a language to an iterated procedure in which participants' productions are used as the input for the next generation of language learners. Iterated Artificial Language Learning (IALL) makes it possible to study how biases affecting language acquisition accumulate over generations, leading to language change. Such methods for studying the role of the individual in change date back at least to Esper 1966, but its recent resurgence as a method for exploring typological patterns originates in Kirby's computer simulations (e.g., Kirby 1999; see also Kirby 1997 in the very first issue of *Linguistic Typology*). In those studies, processing biases like those discussed in Hawkins 1994 are implemented in simulated agents, resulting in more frequent language change towards typologically attested languages.

In modern IALL studies with human participants, it is standard to record the productions of a single participant in one generation and feed them to a single participant in the next, resulting in independent "chains" of participants. By running several chains, a distribution of resulting language types is obtained, which can be used to estimate the probability of the input language changing over generations to one of the languages observed in the final iteration. In fact, under certain assumptions, this distribution will eventually converge on the language learner's prior over possible languages (Griffiths & Kalish 2007, Kirby et al. 2007, Griffiths et al. 2008), where the prior can be understood as whatever predisposition humans have for certain language configurations. If some languages are disfavored by the learner (e.g., due to conflicting parameter settings or being functionally suboptimal) then they will be less frequently represented in this distribution. IALL therefore provides a method to directly map the space of possible languages.[1] IALL is particularly valuable because it is more sensitive than ALL to small effects, detecting biases that are too weak to be observed in a single generation but which could still have a substantial influence on language change over time (Kirby et al. 2007, Reali & Griffiths 2009, Smith 2011).

As yet, little work has explicitly explored typological universals of the type discussed by Dunn et al. using IALL. Nevertheless, the applicability of this method is confirmed by studies that have documented the emergence of natural language-like properties in lab experiments.

Reali & Griffiths (2009) use IALL to explore processes of regularization similar to those observed in creole formation. When adult participants are taught variable names for objects, they tend to approximately reproduce that variability, as in other ALL studies with inconsistent input (e.g., Hudson Kam & Newport 2009). However, when this procedure is iterated over generations, a gradual tendency towards regularization emerges: higher frequency forms displace lower frequency forms. Comparable results are observed by Smith & Wonnacott (2010) in a study of plural markers: over generations, variability in marking is regularized, and different markers become associated with different noun classes. Cornish (2006) and Kirby et al. (2008) find that the languages that emerge in their experiments are increasingly adaptive for communication (error rate decreases over generations). They present evidence that this is due to emergent compositionality in the languages, similar to the morphosyntactic compositionality of natural languages.

Gutman (2011) reports one preliminary application of IALL to a word order universal, viz., the typological correlation between word order fixedness and case marking. Some fixing of word order and loss of case is observed over

1. A related line of research explores the emergence of (not necessarily linguistic) communication systems. Participants in collaborative tasks have been shown to invent functioning communication systems using abstract symbols (Galantucci 2005), pictures (Garrod et al. 2007), or restricted strings of symbols (Selten & Warglien 2007). In certain situations, these emergent systems come over time to take on language-like properties, such as conventionalization, abstraction, and compositionality. Such studies may shed light on constraints or biases on language that are rooted in human interaction and communication more generally.

generations, but it is hard to interpret due to short training periods and correspondingly noisy transmission in the task. This question remains under active research.

4. Comparing quantitative typology and (Iterated) Artificial Language Learning

To summarize, behavioral data elicited from ALL and IALL experiments can provide a viable complement to historical and typological data when evaluating theories of language change and typological universals. In particular, ALL experiments have provided evidence for implicational word order universals of the type for which Dunn et al. (2011) fail to find evidence (Christiansen 2000, Culbertson & Smolensky in press). IALL arguably has tremendous potential as a method for the study of universals. In their application to the study of universals, both ALL and IALL are recent and as of yet under-utilized methodological innovations with their own set of challenges. We end with a discussion of some strengths and weaknesses of the (I)ALL paradigm which particularly complement phylogenetic analysis.

Typological methods – particularly those which make use of phylogenetic approaches – often need to make substantial simplifications to the representation of languages, and this may limit the conclusions that can be drawn from them (e.g., Wälchli 2009). Although existing (I)ALL studies have used relatively simple languages, in principle languages of any complexity could be used. In particular, variable or mixed language configurations and probabilistic tendencies are often eliminated in the featural representations used in typological studies. As suggested by the behavioral studies summarized above, understanding when variation is regularized or preserved may be important in understanding language change.

The biased typological frequency distributions of linguistic features can mean that in extant languages, there will not be not enough data to test certain theories. For instance, strictly OSV order languages are vanishingly rare, meaning that no theories making specific predictions about which features should correlate with OSV could be tested. In contrast, hypotheses can be tested against constructed languages with any imaginable combination of language features.

All historical studies necessarily rely on some kind of representation of the past states of dead languages which must be reconstructed given limited evidence. Even where phylogenetic trees can be inferred with reasonable certainty, the time-course of divergence can be difficult to estimate, and system-external factors like contact may complicate the picture. IALL allows the state of a language at each generation to be observed and analyzed exactly. While this information is not useful for reconstructing actual language histories, artificial language histories enable the study of exactly which properties correlate with language stability, and which to gradual or rapid change.

Naturally, (I)ALL is not a replacement for analysis of historical and typological data. Artificial language studies face their own challenges. Most obviously, language learning in the lab is not the same as first language learning of a natural language. Children may differ from adults, and participants may have differing biases depending on their first language. Understanding the potential consequences of these limitations has been an active area of research. For example, early comparison of the learning behavior of child and adult participants in ALL experiments suggested that children and adults differ in what they do when confronted with variable input (e.g., Austin & Newport 2011; Hudson Kam & Newport 2005, 2009) like that in the mixture shift paradigm. In particular, adult learners generally match input patterns rather closely and only show small deviations. For researchers interested in the nature of language acquisition this is a desirable property. For researchers interested in detecting potentially subtle universal biases, this can pose a challenge: in the mixture shift paradigm, it is the DIRECTION OF THE DEVIATION FROM THE INPUT that is used to assess whether a hypothesized universal bias on acquisition exists or not. If no deviation is found, one is left with a null result. More recent work, however, suggests that adult participants in ALL experiments resemble the learning behavior of children, when confronted with more complex languages (Hudson Kam & Newport 2009). By carefully constructing artificial languages that are sufficiently complex to lead to deviations from the input, but sufficiently simple to be learned, it is then possible to draw conclusions about the direction of learning biases, such as implicational word order universals. Even if qualitative differences between infant and adult learners exist, this apparent confound would offer the potential to study the different roles of child and adult language users on language change, as well as the importance of language contact. Since the best source of hypotheses for these (I)ALL studies will remain historical and typological research conducted using real language data, we believe that the most productive research program will involve close collaboration between typologists and behavioral researchers.

Another challenge to IALL experiments is of a more technical nature. IALL experiments can be time-consuming and expensive to conduct, as they require a large number of participants and tend to require a relatively long test time even for each participant. Given that ALL experiments on syntactic biases typically involve three to four days of training per participant, and given that even small IALL experiments typically involve four independent chains of ten generations of learners, IALL experiments on syntactic universals could easily require 160 participant visits to the lab per experiment. This can easily become infeasible if more chains or additional conditions (i.e., separate languages) are required. Some of these costs may be offset in the near future by technological advances, however. For instance, Tily et al. (2011) introduce a web-based experimental platform designed to look like commercial language teaching software. This has been used to recruit hundreds of participants for ALL experiments in just a few days. Experiments conducted with this software platform replicate findings from the lab (Frank et al. 2010, Tily et al. 2011). Gutman (2011) demonstrates that this technology can be extended to conduct IALL experiments over the web. This shortens the time and reduces the costs for an IALL experiment by several orders of magnitude (the experiment reported in Gutman 2011 was completed in two weeks at a cost of less than USD100). This is not merely convenient. By dramatically reducing the investment necessary to conduct IALL experiments, it is possible to teach more complex languages. By conducting experiments over the web, it also becomes more feasible to systematically compare and control for effects of properties of the learners' native languages.

5. Conclusion

(Iterated) Artificial Language Learning is a powerful research method that can complement quantitative typology, and vice versa. Some of the ALL experiments summarized above (in particular Christiansen 2000, Culbertson & Smolensky in press, Tily et al. 2011) provide an independent source of evidence that the failure to find a significant effect in Dunn et al. 2011 might be a spurious null effect (Type II error). Artificial Language Learning and, in particular, Iterated Artificial Language Learning are relatively recent research paradigms, with extensions that are still being explored. These paradigms offer a great opportunity for collaborations between typologists and behavioral scientists. Work underway is extending the experiments reported above to other word order phenomena. Many of these studies could be conducted as Iterated ALL experiments, permitting an even more direct investigation of the hypothesis that biases operating during language acquisition lead to co-development of word order features, and hence implicational universals in the sense of Greenberg (1963).

Received: 6 June 2011
Revised: 14 July 2011

Massachusetts Institute of Technology
University of Rochester

Correspondence address: (Tily, corresponding author) MIT Brain and Cognitive Sciences, 43 Vassar St, Building 46, Room 3037, Cambridge MA 02139, U.S.A.; e-mail: hjt@mit.edu; (Jaeger) Brain and Cognitive Sciences, University of Rochester, Rochester NY 14627, U.S.A.

Acknowledgements: We thank Jennifer Culbertson, Masha Fedzechkina, Kyle Mahowald, Elissa Newport, and the anonymous *LT* reviewers for their helpful comments on this commentary. This work was partially supported by an Alfred P. Sloan Research Fellowship to TFJ.

References

Amato, Michael S. & Maryellen C. MacDonald. 2010. Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition* 116. 143–148.

Austin, Alison C. & Elissa L. Newport. 2011. When children learn more than they are taught: Regularization in child and adult learners. Unpublished manuscript, University of Rochester.

Bates, Elizabeth A. & Brian MacWhinney,. 1989. Functionalism and the competition model. In Brian MacWhinney & Elizabeth A. Bates (eds.), *The crosslinguistic study of sentence processing*, 3–73. Cambridge: Cambridge University Press.

Chomsky, Noam & Howard Lasnik. 1993. The theory of Principles and Parameters. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 506–569. Berlin: Walter de Gruyter.

Christiansen, Morten H. 2000. Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In Jean-Louis Dessalles & Ladeh Ghadakpour (eds.), *Proceedings of the 3rd Conference on the Evolution of Language, Paris, France*, 45-48.

Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31. 489–509.

Cornish, Hannah. 2006. Iterated learning with human subjects: An empirical framework for the emergence and cultural transmission of language. Edinburgh: University of Edinburgh master's thesis.

Croft, William, Tanmoy Bhattacharya, Dave Kleinschmidt, D. Eric Smith & T. Florian Jaeger. 2011. Greenbergian universals, diachrony, and statistical analysis. *Linguistic Typology* 15. 433–453.

Culbertson, Jennifer & Géraldine Legendre (in press). Investigating the evolution of agreement systems using an artificial language learning paradigm. *Western Conference on Linguistics* 40.

Culbertson, Jennifer & Paul Smolensky (in press). Testing Greenberg's Universal 18 using an artificial language learning paradigm. *North East Linguistic Society* 40.

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Language typology and syntactic description*, Vol. 1: *Clause structure*, 61–131. 2nd edn. Cambridge: Cambridge University Press.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

Esper, Erwin A. 1966. Social transmission of an artificial language. *Language* 42. 575–580.

Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2011. Functional biases in language learning: Evidence from word order and case-marking interaction.*Cognitive Science Society* 33. 318–323. http://palm.mindmodeling.org/cogsci2011/papers/0062/paper0062.pdf

Finley, Sara & William Badecker. 2009. Right-to-left biases for vowel harmony: Evidence from artificial grammar. *North East Linguistic Society* 38(1). 269–282.

Frank, Michael C., Inbal Arnon, Harry Tily & Sharon Goldwater. 2010. Beyond transitional probabilities: Human learners impose a parsimony bias in statistical word segmentation. *Cognitive Science Society* 31. 760–765. http://palm.mindmodeling.org/cogsci2010/papers/0272/index.html

Galantucci, Bruno. 2005. An experimental study of the emergence of human communication systems. *Cognitive Science* 29. 737–767.

Garrod, Simon, Nicolas Fay, John Lee, Jon Oberlander & Tracy MacLeod. 2007. Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science* 31. 961–987.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 110–113. Cambridge, MA: MIT Press.

Griffiths, Thomas L., Brian R. Christian & Michael L. Kalish. 2008. Using category structures to test iterated learning as a method for revealing inductive biases. *Cognitive Science* 32. 68–107.

Griffiths, Thomas L. & Michael L. Kalish. 2007. Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31. 441–480.

Gutman, Jacqueline. 2011. Cultural transmission of grammatical structure: Using a web-based iterated learning paradigm to investigate the emergence of systematicity in syntax. Rochester: University of Rochester master's thesis.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 1999. The relative order of prepositional phrases in English: Going beyond Manner–Place–Time. *Language Variation and Change* 11. 231–266.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, John A. 2007. Processing typology and why psychologists need to know about it. *New Ideas in Psychology* 25. 87–107.

Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203. 88–96.

Hudson Kam, Carla & Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1. 151–195.

Hudson Kam, Carla & Elissa L. Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59. 30–66.

Hupp, Julie M., Vladimir M. Sloutsky & Peter W. Culicover. 2009. Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes* 24. 876–909.

Jaeger, T. Florian & Harry Tily. 2011. On language "utility": Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(3). 323–335.

Kirby, Simon. 1997. Competing motivations and emergence: Explaining implicational hierarchies. *Language Typology* 1. 5–32.

Kirby, Simon. 1999. *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS* 105(31). 10681–10686.

Kirby, Simon, Mike Dowman & Thomas Griffiths. 2007. Innateness and culture in the evolution of language. *PNAS* 104(12), 5241–5245.

MacWhinney, Brian. 1983. Miniature linguistic systems as test of the use of Operating Principles in second-language learning by children and adults. *Journal of Psycholinguistic Research* 12. 467–478.

Morgan, James L. , Richard P. Meier & Elissa L. Newport. 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.

Nevins Andrew. 2010. Two case studies in phonological universals: A view from artificial grammars. *Biolinguistics* 4. 217–232.

Newport, Elissa L. & Richard N. Aslin. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48. 127–162.

Reali, Florencia & Thomas L. Griffiths. 2009. The evolution of linguistic frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111. 317–328.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace.

Selten, Reinhard & Massimo Warglien. 2007. The emergence of simple languages in an experimental coordination game. *PNAS* 104(18). 7361–7366.

Smith, Kenny. 2011. Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology* 83. 261–278.

Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116. 444–449.

Tily, Harry, Michael C. Frank & T. Florian Jaeger. 2011. The learnability of constructed languages reflects typological patterns. *Cognitive Science Society* 33. 1364–1369. http://palm.mindmodeling.org/cogsci2011/papers/0317/paper0317.pdf

Wälchli, Bernhard. 2009. Data reduction typology and the bimodal distribution bias. *Linguistic Typology* 13. 77–94.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.

Wonnacott, Elizabeth, Elissa L. Newport & Michael K. Tanenhaus. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology* 56. 165–209.