

Communicative biases shape structures of newly acquired languages

Maryia Fedzechkina (mashaf@bcs.rochester.edu), T. Florian Jaeger (fjaeger@bcs.rochester.edu)

Department of Brain and Cognitive Sciences, University of Rochester
Rochester, NY 14627 USA

Elissa L. Newport (eln10@georgetown.edu)

Department of Neurology, Georgetown University,
Washington, DC 20007 USA

Abstract

Languages around the world share a number of commonalities known as *language universals*. We investigate whether the existence of some recurrent patterns can be explained by the learner's preference to balance the amount of information provided by the cues to sentence meaning. In an artificial language learning paradigm, we expose learners to two languages with optional case-marking – one with fixed and one with flexible word order. We find that learners of the flexible word order language, where word order is uninformative of sentence meaning, use significantly more case-marking than the learners of the fixed word order language, where case is a redundant cue. The learning outcomes in our experiment parallel a variety of typological phenomena, providing support for the hypothesis that communicative biases can shape language structures.

Keywords: Language acquisition; learning biases; language universals; efficient communication.

Introduction

In his seminal paper, the American linguist Joseph Greenberg (Greenberg, 1963) noticed that the vast majority of patterns that recur in apparently unrelated languages, also known as *language universals*, take the form of implicational statements: If a language has property A, then it will most likely have property B. Language universals point towards constraints on the space of structures possible in natural language since some of the theoretically possible feature combinations are cross-linguistically observed more frequently than others.

The nature of these recurrent patterns has sparked a debate in the cognitive sciences: Are language universals due to constraints specific exclusively to language which are not shared by other aspects of human cognitive systems (Chomsky, 1965), or are they due to general cognitive constraints such as constraints on perception, memory and learning (Hawkins, 2004; E Newport, 1981; Slobin, 1973)?

In this work, we explore the long-standing hypothesis that domain-general pressures associated with human communication can shape languages over time (Bates & MacWhinney, 1989; Slobin, 1977; Zipf, 1949). Support for this hypothesis comes from recent studies that apply mathematical theories of communication to the study of language structures. This work has found that speech has many properties that strike an efficient balance between successful and fast communication (Jaeger, 2010). Recent cross-linguistic studies have further found that languages

across the world share many properties that facilitate such efficient information transfer (S.T. Piantadosi, H Tily, & E Gibson, 2011; S.T. Piantadosi, H. Tily, & E. Gibson, 2011; Qian & Jaeger, 2012).

The study of language universals has primarily relied on typological and diachronic data, which has several major limitations. First, typological and diachronic studies suffer from data sparsity since only a small fraction of 6909 known languages (Lewis, 2009) have been sufficiently documented. This led some researchers to suggest that there is no evidence for language universals once common ancestry and geographical proximity between languages are taken into account (Dunn, Greenhill, Levinson, & Gray, 2011). Second, typological and diachronic studies do not provide an insight into the mechanism of *how* the hypothesized constraints come to shape language over time.

The aim of this study is three-fold. First, we set out to provide direct behavioral evidence for the existence of cross-linguistic universals. Second, we investigate the cause of these universals. We ask, in particular, whether some typologically frequent phenomena can be explained by domain-general biases associated with considerations about human communication. Finally, we explore whether learning can provide a potential mechanism through which these biases come to shape language structures.

We employ an artificial language learning paradigm, where learners are exposed to miniature languages designed to have certain properties of interest. This method has been used to study learning biases in adults and children (Culbertson, Smolensky, & Legendre, 2012; Hudson Kam & Newport, 2009) and has provided behavioral evidence for typological universals (Christiansen, 2000; Culbertson et al., 2012; Finley & Badecker, 2008; E. Newport & Aslin, 2004). Of particular interest is a recent study by Fedzechkina, Jaeger, and Newport (2012), who have used this paradigm to investigate the impact of learners' communicative preferences on language structure and shown that language learners are biased towards efficient case systems.

Here we focus on the correlation in a language between the flexibility of word order and the presence of a case system. It has long been observed that languages with rich case-marking typically allow more word order freedom than languages with no case-marking (Blake, 2001; Sapir, 1921). Languages like Russian or Latin, which allow sentential subjects and objects to be placed in a variety of positions

with respect to each other and the verb, tend to have rich case systems. However, languages that enforce strict order of subject and object (e.g., English and French) typically have no or only rudimentary case-marking.

Additional evidence for the correlation between word order flexibility and the presence of case-marking comes from studies of language change. For instance, Old English allowed the permutation of subject and object while having rich case-marking. This relationship between word order and case-marking substantially changed during the history of English, and Modern English became a language with a fixed word order and no case-marking.

We explore whether this correlation between word order freedom and the presence of a case system in a language can result from a trade-off between the information content of a cue and the amount of effort necessary to produce this cue. Word order is highly informative of sentence meaning (i.e., grammatical function assignment can be successfully recovered based on word order alone) in a fixed word order language. Case-marking is thus redundant in such language and can be omitted to conserve effort without hindering robust communication. In a flexible word order language, however, word order is less informative of grammatical function assignment, and situations can occur when sentence meaning cannot be successfully recovered based on the linear order of elements alone. The relative lack of informativity of word order is compensated for by case-marking, which, when present, provides crucial information about grammatical function assignment.

We expose learners to languages with optional case-marking that have either fixed or flexible word order. If learners indeed balance the amount of information provided by cues to sentence meaning, we predict that the relative lack of one cue will make it more likely that learners recruit alternative cues. In particular, we expect learners of the flexible word order language to use significantly more case-markers than the learners of the fixed word order language.

Experiment

Participants

Participants in the experiment were monolingual native English speakers recruited from the undergraduate students at the University of Rochester or their age-matched peers from the surrounding community. All participants were compensated \$25 for their time. Participants were pseudo-randomly assigned to one of the two languages: *variable word order* or *fixed word order* language (described below). Recruitment continued until the number of participants who successfully learned the language reached 20 in each of the two languages. 52 volunteers participated in the experiment. 12 participants were excluded from the analysis for the following reasons: failure to achieve 65% accuracy on unambiguous trials during the comprehension test (10 participants in the flexible word order language, see below), computer error (1 participant), or being a bilingual (1

participant). This left 40 participants for analysis, 20 in each of the two languages.

The Languages

Lexicon Verbs There were four verbs (*geed, kleidum, shen, zamper*) that denoted simple transitive actions (HUG, KNOCKOVER, ROCK, KICK). All verbs occurred equally frequently in the input overall and with each word order variant allowed by the language.

Nouns There were six nouns (*glim, flugit, bliffen, norg, spad, melnawg*), all of which denoted male referents (MOUNTIE, CHEF, REFEREE, CONDUCTOR, HUNTER, BANDIT). There were no restrictions on nouns. All nouns occurred equally often as subjects and objects of each of the four verbs.

Case-marker There was one case-marker ('*kah*') that optionally marked the object of the action.

Grammar There were two language conditions in the experiment.

Fixed word order language did not contain word order variation – subject-object-verb (SOV) occurred in 100% of the input sentences. The language had optional case-marking – 67% of the input sentences contained a case-marker that marked the object of the action.

Since grammatical function assignment could be unambiguously identified by word order in this language, case-marking added little information to successful recovery of sentence meaning.

Flexible word order language contained word order variation – subject-object-verb (SOV) and object-subject-verb (OSV) word orders occurred equally frequently in the input. The language contained optional case-marking – 67% of sentential objects were case-marked regardless of sentence word order.

In this language, word order was uninformative about grammatical function assignment. Case-marking, when present, provided important information about sentence meaning.

Head-final languages were chosen for both language conditions since they are cross-linguistically more likely to have case-marking systems (Greenberg, 1963).

The Procedure

The procedure builds on the method developed by Hudson Kam and Newport (2009). Participants were trained and tested on one of the two languages during three 30-35 minute visits to the lab spread over three consecutive days, with at most one day between the visits. The same procedure was used on all three visits. During each session, participants were presented with a mixture of training and test blocks that fell into two broad categories: noun training and sentence training.

Noun Exposure and Tests During *noun exposure* participants were presented with pictures of each of the characters accompanied by their label in the novel language

(12 trials total). Participants were instructed to repeat the names of the characters aloud to help them learn. The initial noun presentation was followed by a *noun comprehension test* where participants were presented with pictures of two characters and asked to choose the correct match for the name they heard (12 trials total). Feedback was provided after each trial. After completing the noun comprehension test, participants were presented with the *noun production test* where they were asked to provide the name of the character shown on the screen (6 trials). Feedback on performance was provided after each trial. The three noun training blocks were repeated immediately upon completion of the noun production test. On Day 1 noun exposure and noun production blocks were also presented before the sentence production test. On Days 2 and 3 participants were only given the noun production block before the sentence production test. Noun exposure and comprehension blocks were also shorter on Days 2 and 3 (6 trials each).

Sentence Exposure and Tests During the *sentence exposure phase*, participants viewed short computer-generated videos and heard their descriptions in the novel language. Participants were asked to repeat the sentences out loud to facilitate learning. On all days, exposure sentences were presented in sets of two training blocks (24 trials each). During the first sentence exposure block on Day 1 participants were allowed to replay the videos and sounds as many times as they wanted; replay was disabled for all other exposure blocks throughout the study. Sentence exposure was followed by a *sentence comprehension test* (24 trials total). Participants were shown two videos involving the same action where the order of the doer and undergoer was reversed and were asked to choose the video best matching the sentence they heard. No feedback on performance was provided during the sentence comprehension test. After the sentence comprehension test, participants completed two more sentence exposure blocks and one more sentence comprehension block. Each experimental session ended with a *sentence production test* (48 trials) where participants were asked to describe a previously unseen video using the provided verb prompt. No feedback on performance was provided.

Scoring

In the comprehension test, participants' responses were scored as 'correct' if they matched the intended sentence interpretation. This was based only on case-marked (unambiguous) trials for both languages. Participants who failed to achieve 65% accuracy were excluded from all analyses. The results reported below still hold, however, if these participants are included as well.

In the production test, we scored the word order used in the sentence, the presence of case-marking on the object as well as lexical (using the wrong word for a referent or an action) and grammatical mistakes (using a word order not allowed by the grammar or using the case-marker incorrectly). If the name of only one referent was incorrect

and it was still possible to determine sentence word order, productions were scored as overall correct but containing a lexical error. Such productions were included in the analyses below. Productions containing grammatical errors were excluded from all analyses. The results presented below still hold if productions containing lexical errors are excluded as well.

Results

Accuracy of Acquisition Both languages were acquired with a high degree of accuracy. On the final day of training learners of the fixed word order language made less than 1% lexical mistakes and no grammatical mistakes, while learners of the flexible word order language made 1.6% lexical mistakes and 6.2% grammatical mistakes in their productions. These data suggest that the task was feasible for our participants.

Word Order Use in Production One way learners of the flexible word order language can ensure robust communication is by fixing word order and dropping case-marking. Thus we first analyzed participants' word order use in the flexible word order condition, asking whether there was a tendency to regularize word order. If learners of the flexible word order language behave just like fixed word order learners who use SOV in all their productions, there will be little reason to expect differential case-marker use between the two language conditions.

Overall, learners of the flexible word order language maintained word order flexibility: There was no word order regularization in participants' productions on any day of training (Day 1: 49% SOV word order in production, not significantly different from the 50% input proportion [$\chi^2(1)=.15, p=.7, ns$]; Day 2: 45% SOV word order in production, not significantly different from the input [$\chi^2(1)=2.66, p=.1, ns$]; Day 3: 49% SOV word order in production, not significantly different from the input [$\chi^2(1)=.17, p=.7, ns$]).

Case-Marker Use in Production We now turn to our main question: Do learners balance the amount of cues to the meaning of the sentence, recruiting additional cues in those cases when existing cues do not provide enough information to successfully decode the intended meaning?

We used a mixed logit model to predict the use of case-marking in participants' productions based on language condition (flexible/fixed word order language) and day of training (1, 2, 3) as well as the interactions between these two factors. The model included the maximal random effects structure justified by the data based on model comparison. The results reported below hold when the model with the full random effects structure was used. There was a significant main effect of language (see Figure 1): Learners of the flexible word order language used significantly more case-markers in their productions than the learners of the fixed word order language throughout the experiment ($\beta=1.45, z=2.24, p<.05$). There was a significant

interaction between Day 2 of training and language condition ($\beta=.46$, $z=3.4$, $p<.001$) and Day 3 of training and language condition ($\beta=.25$, $z=2.75$, $p<.01$). Simple effects test shows that learners of the flexible word order language used significantly more case-markers than the learners of the fixed word order language on Day 2 ($\beta=1.65$, $z=2.5$, $p<.05$) and Day 3 ($\beta=1.94$, $z=2.72$, $p<.01$) of training.

As expected under our hypothesis, then, learners used significantly more case-marking when they learned a language where word order was uninformative of grammatical function assignment (flexible word order language) as compared to the language where referent-to-grammatical-function assignments were reliably identified by word order (fixed word order language).

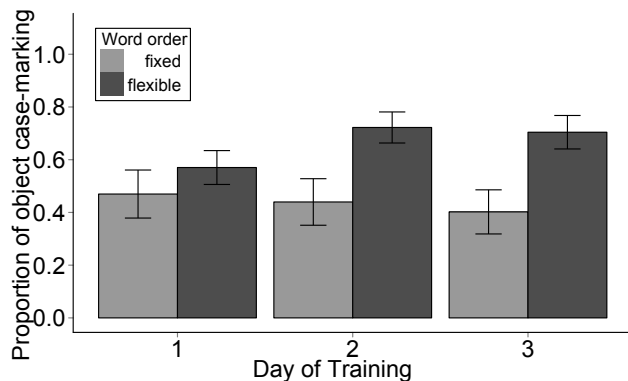


Figure 1: Case-marker use in production

What is driving the observed difference in case-marker use between the two language conditions? Under our hypothesis, we expect learners of the fixed word order language to gradually lose case-marking, producing fewer case-markers than the input proportion, since case-marking is a redundant cue to sentence meaning in a language that does not allow word order variability. Given the design of our flexible word order language, learners could take advantage of case-marking to facilitate successful decoding of the intended meaning in two ways. First, learners could regularize case-marking in the language overall and use more case-markers in their own productions than in the input. Alternatively, learners could condition case-marking on word order and use significantly more case-markers with one word order variant than with the other. Both possibilities will increase successful recovery of the intended meaning, but the latter will minimize effort at the same time (since not all sentential objects will need to be case-marked) and would thus be a more efficient option. In the following sections, we explored these predictions in more detail.

Case-Marker Use in the Fixed Word Order Language

Do learners of the fixed word order language deviate from the input they receive and reduce the amount of case-marking in the newly acquired language? Indeed, they showed a strong tendency to drop case-marking and used

significantly fewer case-markers in their own productions compared to the input starting on the first day of training (Day 1: 50% case-marking in production, significantly lower than the 67% input proportion [$\chi^2(1)=23.51$, $p<.001$]; Day 2: 45% case-marking in production, significantly lower than the input [$\chi^2(1)=40.6$, $p<.001$]; Day 3: 41% case-marking in production, significantly lower than the input [$\chi^2(1)=61.87$, $p<.001$]).

The behavior of the majority of individual subjects followed our prediction. Out of 20 participants, 14 participants produced fewer case-markers than the input on the final day of training; 8 of these did not use case-marking at all in their own productions; and only 3 participants produced substantially more case-markers than the input proportion (see Figure 2).

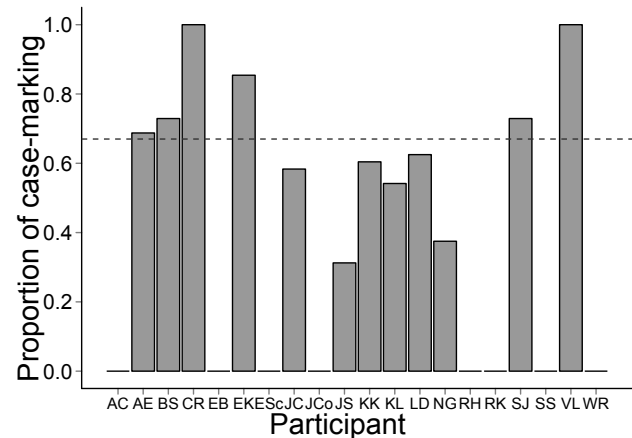


Figure 2: Individual preferences in case-marker use in the fixed word order language on the final day of training. The dashed line indicates the input proportion of case-marking.

Case-Marker Use in the Flexible Word Order Language

Next we took a closer look at the learning outcomes in the flexible word order language. Do learners increase communicative success by favoring robustness but sacrificing efficiency and regularize case-marking in the language overall? Or do they favor efficiency to achieve the same goal, conditioning case-marking on sentence word order?

Learners of the flexible word order language did not use case-markers significantly more frequently than in the input language (Day 1: 55% case-marking in production, significantly below 67% input proportion [$\chi^2(1)=9.5$, $p<.01$]; Day 2: 72% case-marking in production, not significantly different from the input [$\chi^2(1)=1.78$, $p=.18$, *ns*]; Day 3: 71% case-marking in production, not significantly different from the input [$\chi^2(1)=.89$, $p=.3$, *ns*]).

There was a significant tendency to condition case-marking on sentence word order throughout the experiment (see Figure 4): Learners overtly marked objects with case significantly more often if sentence word order was OSV ($\beta=1.11$, $z=17$, $p<.001$). A significantly higher proportion of

object case-marking in OSV sentences compared to SOV sentences was observed on every individual day of training (Day 1: ($\beta=1.53$, $z=12.6$, $p<.001$); Day 2 ($\beta=.9$, $z=8.6$, $p<.001$); Day 3 ($\beta=.91$, $z=8.23$, $p<.001$)).

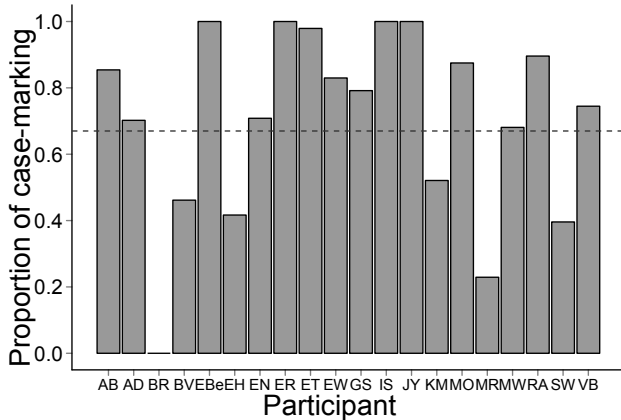


Figure 3: Individual preferences in case-marker use in the flexible word order language on the final day of training. The dashed line indicates the input proportion of case-marking.

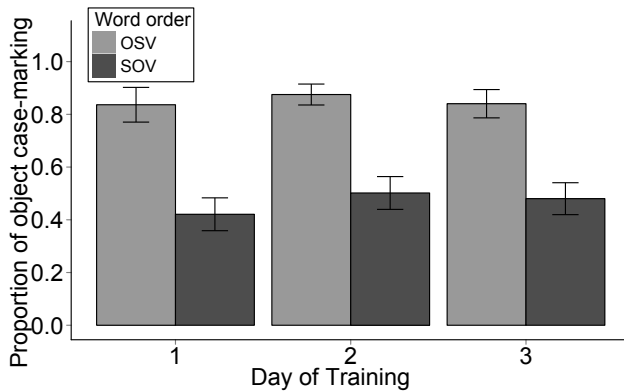


Figure 4: Case-marker use by sentence word order in production (flexible word order condition)

Learning outcomes in each language condition thus support our hypothesis. As expected, learners of the fixed word order language tend to gradually drop case-marking as they acquire the novel language. Learners of the flexible word order language maintain case-marking in the language they acquire and make efficient use of it by conditioning case-marking on sentence word order.

Discussion

Our results provide experimental evidence that some of the typological properties observed cross-linguistically can stem from learners' biases towards communicatively efficient systems. We found that language learners have a bias to balance the amount of information provided by two cues to sentence meaning (case and word order) and tend to use an

additional cue significantly more often in those cases when the other cue does not provide sufficient information to successfully parse the sentence.

Importantly, our results parallel synchronic and diachronic typological data from natural languages. We presented learners with languages that contained the same amount of case-marking, but the learning outcomes consistently differed depending on the amount of word order flexibility allowed in the input language. While learners of the flexible word order language retain case-marking in their own productions, as do Russian, Latin and other free word order languages, learners of the fixed word order language tend to lose case-marking as they acquire the new language, mimicking fixed word order languages such as English and French.

The learning outcomes in our experiment also parallel diachronic phenomena such as the evolution of English from Old English to Modern English. Our data, however, does not speak to whether word order fixing was a result of the loss of case-marking or whether case-marking became a redundant cue and was lost after English word order became fixed for independent reasons. Under our hypothesis, both processes will yield the same outcome.

We found that learners of the flexible word order language did not regularize case-marking uniformly across the two possible word orders. Instead they restructured the language to make efficient use of case-marking by conditioning it on sentence word order, using significantly more case-markers when sentence word order was OSV.

Why do learners preferentially case-mark objects in sentences with OSV word order and not the other way around? This behavior could be reflective of a cognitive bias to mark the atypical, somewhat resembling the cross-linguistically common phenomenon known as 'word order freezing' (Lee, 2001). Many flexible word order languages (Turkish, Russian, German, Hindi, Japanese, etc.) enforce default word order for sentences in which case-marking is uninformative (e.g., in the absence of case-marking or in the presence of case syncretism) and require direct objects to be overtly case-marked if sentence word order deviates from the dominant one. For instance, the Russian example in (1) with case-syncretism can only be interpreted as SVO, although Russian generally allows both SVO (dominant) and OVS orders (non-dominant).

- (1) Mat' l'jubit doch'.
- Mother-NOM/ACC loves daughter-NOM/ACC
- 'Mother loves daughter'

Even though SOV and OSV are equally likely both in the input and in participants' productions in our experiment, OSV word order is typologically rare and is almost always a non-default word order cross-linguistically and thus might attract a higher proportion of case-marking.

The second possibility is that learners prefer to put more informative cues earlier in the sentence to allow for faster processing, as has been hypothesized by Hawkins (2004)

and Nichols (1986). Preliminary support for this hypothesis comes from Fedzechkina et al. (2012), who found that learners of a language with object case-marking preferred to use more case-marking in OSV sentences, whereas learners of a language with subject case-marking used significantly more case-marking in SOV sentences. That is, in both cases, learners preferred case-marking on the sentence-initial argument, thereby providing the disambiguating cue as early as possible in the sentence.

Conclusions

We used an artificial language learning paradigm to ask whether language structures are shaped by communicative biases. We find that the cross-linguistically common correlation between word order flexibility and the presence of case-marking can be at least partially explained by domain-general learning biases stemming from a preference to balance the amount of information provided by the cues to sentence meaning.

Acknowledgments

We thank T. Stanley and S. Krishnan for their help with data collection and data coding; C. Anzalone and G. Smith for their help with data coding; C. Donatelli, I. Minkina and A. Wood for their help with video stimuli. This work was supported in part by NSF grants BCS-0845059 and IIS-1150028, as well as an Alfred P. Sloan Fellowship to T.F.J., and by NIH grant DC00167 and HD037082 to ELN.

References

- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, 3–73.
- Blake, B. J. (2001). *Case*. Cambridge: CUP.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 119): The MIT press.
- Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals *The Evolution of Language: 3rd International Conference*.
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122, 306-329.
- Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345), 79-82.
- Fedzechkina, M., Jaeger, T. F., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proc Natl Acad Sci USA*.
- Finley, S., & Badecker, W. (2008). *Substantive biases for vowel harmony languages*. Paper presented at the Proceedings of WCCFL 27.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of human language* (pp. 73-113): MIT Press.
- Hawkins, J. A. (2004). *Efficiency and Complexity in Grammar*. Oxford: Oxford University Press.
- Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychol*, 59(1), 30-66.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychol*, 61, 23-62.
- Lee, H. (2001). Markedness and word order freezing. In P. Sells (Ed.), *Formal and empirical issues in optimality theoretic syntax* (pp. 63-128). Stanford: CSLI.
- Lewis, M. (Ed.). (2009). *Ethnologue: Languages of the world* (16th edition ed.). Dallas, TX: SIL International.
- Newport, E. (1981). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Lang Scie*, 10, 147-172.
- Newport, E., & Aslin, R. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychol*, 48(2), 127-162.
- Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 56-119.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition*, 122, 280-291.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA*, 108(9), 3526.
- Qian, T., & Jaeger, T. F. (2012). Cue Effectiveness in Communicatively Efficient Discourse Production. *Cognitive Sci*, 36, 1312-1336.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. *First Language Acquisition: The Essential Readings*.
- Slobin, D. I. (1977). Language change in childhood and in history. In J. Macnamara (Ed.), *Language Learning and Thought* (pp. 185-214). New York: Acad. Press.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.