Communicative Efficiency, Language Learning, and Language Universals

by

Maryia Fedzechkina

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor T. Florian Jaeger

Professor Elissa L. Newport

Department of Brain and Cognitive Sciences

Arts, Sciences and Engineering

School of Arts and Sciences

University of Rochester

Rochester, New York

2014

Biographical Sketch

The author was born in Minsk, Belarus, in January 1983. She attended Minsk State Linguistic University in Minsk, Belarus, and graduated with a Bachelor of Arts in English (with a minor in German) in 2007. She received her Master of Arts degree in Linguistic Data Processing (with minors in Phonetics and English) from the University of Cologne, Germany in 2009. In the fall of 2009, she began doctoral studies in the Department of Brain and Cognitive Sciences at the University of Rochester. She received a Master of Arts degree in Brain and Cognitive Sciences in the Fall of 2012. She pursued her research at the University of Rochester under the direction of Professors T. Florian Jaeger (Brain and Cognitive Sciences) and Elissa L. Newport (Neurology, Georgetown University).

The following publications are a result of research the author conducted while pursuing doctoral studies at the University of Rochester:

Fedzechkina, M., Newport, E., Jaeger, T. F. (under review). Balancing effort and information transmission during language acquisition: Evidence from word order and case-marking.

Fedzechkina, M., Jaeger, T. F., Newport, E. (2013). Communicative biases shape structures of newly acquired languages. *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

Fedzechkina, M., Jaeger, T. F., Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109(44): 17897-17902*.

Fedzechkina, M., Jaeger, T. F., Newport, E. (2011). Functional biases in language learning: Evidence from word order and case-marking interaction. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

Acknowledgements

My first big, huge, ginormous 'thank you' goes to my adviser Florian Jaeger. I don't think I would have survived in graduate school without his support, insight and guidance. And definitely the absolutely amazing experience I had in Rochester would not be the same without him. Thank you, Florian, for intellectually challenging me to get out of my comfort zone all these years! I hated it every time I had to do it, but I loved everything I've accomplished because of it.

My second big, huge 'thank you' goes to my other adviser Elissa Newport, whose insights and elegant thinking have challenged me to be more precise and coherent in the way I think, write, and evaluate my work.

I thank Dick Aslin for his advice and encouragement over the years, whose wisdom and enthusiasm have been a great source of inspiration. I regret starting talking to Davy Temperley so late in the process. Thank you, Davy for always asking great questions! They have prompted me to think about my ideas and assumptions underlying my work more critically. I thank Greg Carlson for his support and encouragement.

I've never had so much fun both intellectually and personally than during the time I spent in Brain & Cognitive Sciences at Rochester. I'm very lucky to have been surrounded by an amazing group of incredibly smart colleagues who were weird enough to also be my friends. I'm especially happy that I had a chance to bounce some of my ideas off Alex Fine, Judith Degen, Patty Reeder, David Ruskin, Esteban Buz, Dave Kleinschmidt, Ting Qian, Ilker Yildirim, Amanda Pogue, Chigusa Kurumada, Scott

Fraundorf, Klinton Bicknell, Bozena Pajak, Catie Hilliard, Thomas Farmer, Lauren Emberson, Maureen Gillespie (even though she's not technically a BCS member, I first met her in Rochester, so I'm adding her here).

A special thanks goes to Lizz Karuza, who was my office mate for all five years of graduate school and a close friend for the last two of them. Weird math… But I'm beyond happy it eventually worked out this way. Another special thanks goes to Dan Pontillo – yelling at each other about science has never been that fun… well, maybe except for yelling about science with Lucia Pozzan. I thank her for that too.

Although I'm looking to collaborate more closely with John Trueswell in the near future, he deserves a 'thank you' already in the present. The ideas and opinions he has generally offered have influenced the way I think about my work. My work has also benefitted from critical discussions and feedback from Jim Blevins, Roger Levy, and Scott Paauw.

I was fortunate to have worked with a team of outstanding research assistants, who taught me a great deal about what kind of mentor I eventually want to be: Rebecca Glass, Colleen Dolan, Keryn Bernstein, Vivian Choi, Tara Stanley, Greta Smith, Theo Lincoln, Cynthia Orr, Rachel Aronovitz, Elisa Hoyos, Chris Anzalone, Meghan Hennessy, Kate Heffernan, Erica Williams, Shreya Krishnan, Suzanne Blackley, and Becky Chu. Thank you, guys! This work would not have been possible without you.

I would also like to thank Jen Gillis, Chris Dambra, Chris Freemesser, Andrew Watts, and Kathy Corser for all their help with administrative/ technical/ lab-related stuff.

I thank Patrick Rault for keeping me (reasonably) sane this past year.

I thank my mom for letting me do whatever I wanted since I was a kid. And specifically for letting me take up random careers in random countries without any good reason. Thanks, mom! It took me a while to find what I wanted, and I wouldn't be here without you.

Another 'thank you' goes to Flugit and George-Euclid for providing purrs and cuddles when I really needed them (and when I didn't). I thank Lucia Pozzan and Steve Marchette for pointing out that the term *crazy cat lady* applies only when there are three or more cats per capita in a household.

I would like to thank my former adviser Juergen Rolshoven from the University of Cologne for encouraging me to pursue a PhD. I also thank my former lab mates from the University of Cologne, especially Juergen Hermes, Stephan Schwiebert, Sonja Subicin, Frauke Schmidt, and Daniel Alberts for their kindness and support.

I thank Valeri Nikolaev for pointing out a possibility of pursuing a PhD degree in the US. Without your encouragement, I might have ended up again in some other random country for no good reason.

I thank Silvia Lipski, who took her valuable postdoc time, and let me hang out at the MPI for Neurological Research in Cologne and introduced me to experimental work in phonetics. Silvia, I might not be here if I didn't have this chance!

Abstract

Languages around the world share a number of commonalities often referred to as *language universals*. The nature of these universals has been a matter of debate in linguistics and cognitive science. Some researchers have suggested that these commonalities arise from learning biases that are exclusive to language acquisition (Chomsky, 1965; Fodor, 2001; Pinker, 1984). Others have argued that constraints on other cognitive systems as well as pressures associated with language use and communication can shape language structures over time (Bever, 1970; Givón, 1991; Hawkins, 2004).

This dissertation examines the hypothesis that the cross-linguistic distribution of grammars can be accounted for, at least in part, in terms of their communicative utility. In a series of miniature artificial language learning experiments, I investigate the hypothesis that communicative pressures (specifically, a trade-off between effort and robust information transfer) operate already during language acquisition, biasing learners towards more efficient linguistic systems.

The findings support this hypothesis: when presented with relatively inefficient input languages, learners deviate subtly but systematically from the input, restructuring the languages to make them more suitable for efficient information transfer. Learning outcomes in these experiments closely mirror typological synchronic and diachronic phenomena such as patterns in cue trade-offs in morpho-syntax (Chapters 2 and 3) as well as properties of differential and optional case-marking systems (Chapter 4), thereby

suggesting that some language structures are shaped by learners' biases for efficient information transfer.

Contributors and Funding Sources

I am the primary author of the entire text of this dissertation, supervised by a committee consisting of Professors T. Florian Jaeger, Elissa L. Newport, Richard Aslin, David Temperley, and Greg Carlson.

All stages of experiment design, data analysis, and interpretation of results were conducted in collaboration with T. Florian Jaeger and Elissa L. Newport, who have also co-authored manuscripts for publication based on this work.

Experiment 1 presented in Chapter 2 was conducted with assistance of Tara Stanley, Shreya Krishnan, Chris Anzalone, Greta Smith, Suzanne Blackley, and Erica Williams. Parts of the data reported in Chapter 2, and earlier versions of the analyses were previously published in a proceedings paper in 2013, as well as in a journal article currently under review, both listed in the Biographical Sketch. Experiment 2 was carried out with assistance of Elisa Hoyos, Kate Heffernan, Meghan Hennessy, and Rachel Aronowitz. This experiment has not been previously published. Experiment 3 was conducted with assistance of Colleen Dolan, Keryn Bernstein, and Chris Anzalone. Previous analyses on a subset of the data from this experiment were previously published in a proceedings paper in 2011, as listed in the Biographical Sketch. Experiments 4 and 5 were conducted with assistance of Colleen Dolan and Vivian Choi. These data were published in the *Proceedings of the National Academy of Sciences* in 2012, listed in the Biographical Sketch.

All work presented here has been supported in part by NSF grants IIS-1150028 and BCS-0845059, as well as an Alfred P. Sloan Fellowship BR2011-092 awarded to T. Florian Jaeger, and by NIH grants DC00167 and HD037082 awarded to Elissa Newport.

Table of Contents

BIOGRAPHICAL SKETCH .... **ii**
ACKNOWLEDGEMENTS .... **iv**
ABSTRACT .... **vii**
CONTRIBUTORS AND FUNDING SOURCES .... **ix**

CHAPTER 1: INTRODUCTION **1**
1.1. TYPOLOGICAL APPROACH TO LINGUISTIC UNIVERSALS 3
1.2. LANGUAGE LEARNING APPROACHES TO LINGUISTIC UNIVERSALS 6
1.3. CONVERGING EVIDENCE FOR LINGUISTIC UNIVERSALS 17
1.4. THE CURRENT PROPOSAL 22
1.5. COGNITIVELY-ARBITRARY CONSTRAINTS 29
1.6. COGNITIVELY-MOTIVATED CONSTRAINTS 31
1.7. FUNCTIONAL CONSTRAINTS 35
1.8. OUTLINE OF THIS DISSERTATION 44

CHAPTER 2: BALANCING ROBUST INFORMATION TRANSFER AND EFFORT: DIFFERENTIAL REGULARIZATION OF CASE-MARKING DEPENDING ON CONSTITUENT ORDER FREEDOM **48**
2.1. DESIGN AND MATERIALS 50
2.2. PROCEDURE 54
2.3. RESULTS 58
2.4. DISCUSSION 77

CHAPTER 3: BALANCING ROBUST INFORMATION TRANSFER AND EFFORT: DIFFERENTIAL REGULARIZATION OF CONSTITUENT ORDER DEPENDING ON THE PRESENCE OF CASE-MARKING **80**
3.1. EXPERIMENT 2 81
3.2. EXPERIMENT 3 93
3.3. GENERAL DISCUSSION 104

CHAPTER 4: USING CUES TO SENTENCE MEANING EFFICIENTLY: EVIDENCE FROM OPTIONAL CASE-MARKING **109**
4.1. THE PHENOMENON: DIFFERENTIAL CASE-MARKING SYSTEMS 110
4.2. THE APPROACH 113
4.3. EXPERIMENT 4 120
4.4. EXPERIMENT 5 124
4.5. DISCUSSION 127

CHAPTER 5: DISCUSSION **132**
5.1. OUTSTANDING QUESTIONS 134
5.2. FUTURE DIRECTIONS 144
5.3. CONCLUSIONS 152
BIBLIOGRAPHY **153**

List of Tables

TABLE 2.1: THE ARTIFICIAL LEXICON. 51
TABLE 3.1: THE ARTIFICIAL LEXICON USED IN EXPERIMENT 3 94
TABLE 3.2: VERB CLASSES USED IN THE LANGUAGES IN EXPERIMENT 3 95

List of Figures

FIGURE 1.1: EXAMPLE OF A COMPREHENSION TEST USED IN MINIATURE ARTIFICIAL LANGUAGE EXPERIMENTS. 11
FIGURE 1.2: TWO PARADIGMS TYPICALLY USED IN ARTIFICIAL LANGUAGE LEARNING RESEARCH. 15
FIGURE 1.3: HYPOTHESIZED RELATIONSHIP BETWEEN INDIVIDUAL-LEVEL BIASES AND LANGUAGE STRUCTURES. 26
FIGURE 2.1: ILLUSTRATION OF FORM TO GRAMMATICAL FUNCTION MAPPINGS IN TWO EXTREME SCENARIOS (FIXED AND RANDOM CONSTITUENT ORDER LANGUAGES) IN EXPERIMENT 1. PICTURES ARE STILL IMAGES OF SAMPLE VIDEOS WITH THEIR ENGLISH GLOSSES (NOT SHOWN TO PARTICIPANTS) AND TRANSLATIONS (NOT SHOWN TO PARTICIPANTS). ARROWS INDICATE FORM TO MEANING MAPPINGS IN THE TWO LANGUAGES. SOLID ARROWS INDICATE ONE-TO-ONE FORM-MEANING MAPPINGS SUCH AS IN THE ABSENCE OF CONSTITUENT ORDER VARIATION OR IN THE PRESENCE OF CASE-MARKING (UNDERLINED). DASHED ARROWS INDICATE ONE-TO-MANY FORM-MEANING MAPPINGS, SUCH AS IN THE ABSENCE OF CASE-MARKING FOR VARIABLE CONSTITUENT ORDER. THE STRIKETHROUGH FORM REFERS TO THE UNGRAMMATICAL (OSV) SENTENCES IN THE FIXED (SOV-ONLY) LANGUAGE. 52
FIGURE 2.2: EXPERIMENTAL PROCEDURE. PICTURES ARE STILL IMAGES OF SAMPLE VIDEOS USED IN EXPERIMENT 1. 57
FIGURE 2.3: CONSTITUENT ORDER USE BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. 61
FIGURE 2.4: INDIVIDUAL PREFERENCES IN CONSTITUENT ORDER USE IN THE FIXED (TOP PANEL), FLEXIBLE (MIDDLE PANEL), AND RANDOM (BOTTOM PANEL) CONSTITUENT ORDER LANGUAGES ON THE FINAL (3RD) DAY OF TRAINING. THE DASHED LINES INDICATE THE INPUT PROPORTION. 62
FIGURE 2.5: CASE-MARKER USE BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE INPUT PROPORTION (EQUAL ACROSS LANGUAGES). 65
FIGURE 2.6: INDIVIDUAL PREFERENCES IN CASE-MARKER USE IN THE FIXED (TOP PANEL), FLEXIBLE (MIDDLE PANEL), AND RANDOM (BOTTOM PANEL) CONSTITUENT ORDER LANGUAGES ON THE FINAL (3RD) DAY OF TRAINING. THE DASHED LINES INDICATE THE INPUT PROPORTION. 66
FIGURE 2.7: CASE-MARKER USE BY SENTENCE CONSTITUENT ORDER IN FLEXIBLE (TOP PANEL) AND RANDOM (BOTTOM PANEL) CONSTITUENT ORDER LANGUAGES. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINES REPRESENT THE INPUT PROPORTION (EQUAL ACROSS CONSTITUENT ORDERS). 70
FIGURE 2.8: UNCERTAINTY VS. EFFORT TRADE-OFF ON THE FINAL (3RD) DAY OF TRAINING. DIAMONDS REPRESENT INPUT LANGUAGES. SOLID CIRCLES REPRESENT MEAN OUTPUT LANGUAGES. EMPTY CIRCLES REPRESENT HYPOTHETICAL OUTPUT LANGUAGES THAT USE THE SAME OVERALL PROPORTIONS OF CONSTITUENT ORDER AND CASE-MARKING

AS THE ACTUAL OUTPUT, BUT DO NOT CONDITION CASE-MARKING ON CONSTITUENT ORDER. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS.74
FIGURE 2.9: UNCERTAINTY VS. EFFORT TRADE-OFF ON THE FINAL (3RD) DAY OF TRAINING. DIAMONDS REPRESENT INPUT LANGUAGES. CIRCLES REPRESENT MEAN OUTPUT LANGUAGES. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. SMALL SEMI-TRANSPARENT DOTS REPRESENT LANGUAGES PRODUCED BY INDIVIDUAL LEARNERS. 75
FIGURE 2.10: ENTROPY REDUCTION STRATEGIES CHOSEN BY INDIVIDUAL PARTICIPANTS IN THE FLEXIBLE (LEFT) AND RANDOM (RIGHT) CONSTITUENT ORDER LANGUAGES. THE SECTIONS REPRESENT THE NUMBER OF PARTICIPANTS WHO ADOPTED EACH OF THE STRATEGIES. 76
FIGURE 3.1: STILL IMAGES OF SAMPLE VIDEOS USED IN EXPERIMENTS 2 AND 3. 84
FIGURE 3.2: CONSTITUENT ORDER USE IN COMPREHENSION BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE INPUT PROPORTION (EQUAL ACROSS LANGUAGES). 86
FIGURE 3.3: CONSTITUENT ORDER USE IN PRODUCTION BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE INPUT PROPORTION (EQUAL ACROSS LANGUAGES). 88
FIGURE 3.4: CONDITIONAL ENTROPY OF GRAMMATICAL FUNCTION ASSIGNMENT BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS CONDITIONAL ENTROPY OF GRAMMATICAL FUNCTION ASSIGNMENT IN THE INPUT (EQUAL ACROSS LANGUAGES). 91
FIGURE 3.5: CONSTITUENT ORDER USE IN COMPREHENSION BY LANGUAGE ON THE FINAL (4TH) DAY OF TRAINING. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE INPUT PROPORTION (EQUAL ACROSS LANGUAGES). 98
FIGURE 3.6: CONSTITUENT ORDER PRODUCTION BY VERB CLASS IN THE TWO LANGUAGES. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE NUMBERS ON THE PANELS REPRESENT DAYS OF TRAINING. 100
FIGURE 3.7: CONSTITUENT ORDER PRODUCTION BY LANGUAGE. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE INPUT PROPORTION (EQUAL ACROSS LANGUAGES). 102
FIGURE 3.8: CONSTITUENT ORDER PRODUCTION FOR PREVIOUSLY UNSEEN VERBS IN THE TWO LANGUAGES. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. THE DASHED LINE REPRESENTS THE OVERALL SOV FREQUENCY IN TRAINING (EQUAL ACROSS LANGUAGES). 103
FIGURE 5.1: CONSTITUENT ORDER PREFERENCES IN EXPERIMENT 1 REPORTED IN CHAPTER 2 IN PRODUCTION (UPPER PANEL) AND COMPREHENSION (LOWER PANEL) ON CASE-MARKED (LEFT PANELS) AND NON-CASE-MARKED (RIGHT PANELS) TRIALS. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. 137
FIGURE 5.2: DEVIATIONS FROM THE INPUT PROPORTION OF THE DOMINANT CONSTITUENT ORDER IN EXPERIMENT 1 REPORTED IN CHAPTER 2 IN THE FIXED (UPPER PANEL), FLEXIBLE (MIDDLE PANEL), AND RANDOM (BOTTOM PANEL) CONSTITUENT ORDER IN

COMPREHENSION AND PRODUCTION. THE ERROR BARS REPRESENT BOOTSTRAPPED 95% CONFIDENCE INTERVALS. 138

Chapter 1

Introduction

As first observed by linguist Joseph Greenberg (Greenberg, 1963), languages across the world seem to share properties at all levels of linguistic organization. Some of these patterns capture regularities in cross-linguistic distributions of a single element (the so called *non-implicational universals*). For example, sentential subjects almost always precede objects in declarative sentences (Greenberg, 1963). The majority of patterns that have been documented by linguists, however, take the form of implicational statements that describe correlations between several elements in a language: If a language has property A, then it most likely has property B. An example of such an implicational universal is the well-documented correlation between constituent order freedom and the presence of case-marking (Blake, 2001; Sapir, 1921): Languages with flexible constituent order often use morphological means, such as case, to mark grammatical function assignment (e.g., German, Japanese, and Russian), whereas languages with fixed constituent order typically lack case morphology (e.g., English, Mandarin, and Italian).

The discovery of these typological universals was initially based on small cross-linguistic corpora (e.g., 30 languages in Greenberg's survey). As empirical work on language universals continued, initial observations about cross-linguistically frequent patterns were reevaluated via large-scale language databases. While this research corroborated many of the language universals observed by Greenberg (Bickel, 2011; Cysouw, 2011; Dryer, 1992), it became apparent that most if not all language universals are not absolute (i.e., occurring in all languages without exception) but rather are statistical tendencies that occur with higher than chance frequency cross-linguistically (*statistical language universals*). Throughout this dissertation, I will assume the notion of language universals in the broadest sense ranging from broad features of natural languages such as recursion and compositionality (cf. Chomsky, 1986; Hockett, 1960) to more specific statistical universals (cf. Greenberg, 1963).

The existence of implicational universals points towards constraints on the space of structures possible or preferred in natural language since some of the theoretically possible feature combinations are cross-linguistically observed more frequently than others. The origins of such recurring patterns have been the subject of long-standing debate in linguistics and cognitive science. Most theories have argued that language universals originate at the level of an individual and thus suggest that language structures are shaped by biases and limitations on human cognitive systems (Chomsky, 1965; Fodor, 2001). Views differ, however, as to whether these biases are specific to language (Chomsky, 1965; Fodor, 2001) or shared by other human cognitive mechanisms (Bates & MacWhinney, 1982; Bever, 1970; Christiansen & Chater, 2008; Deacon, 1997; Givón, 1991; Hawkins, 1994; Newport, 1981b; Slobin, 1973).

Capturing language universals and understanding their causes has been one of the central objectives in modern linguistics. If language universals are indeed not due to chance and originate in human cognition, understanding their causes would give us insights into the nature of the constraints underlying language processing and representation in the human brain.

In the remainder of this chapter, I briefly summarize the typological approach traditionally used to study language universals and outline two major limitations of this approach, one of which has to do with the structure of typological data itself and the other with the inability to address the causes of language universals. I then introduce a complementary approach that draws conclusions about cross-linguistically frequent patterns by studying patterns in first language acquisition. This approach has its own limitations that are primarily due to the lack of control over the input learners receive. This leads me to suggest that an alternative method – miniature artificial language learning – is a suitable complement to typological data since it studies patterns in acquisition but does not have the limitations of natural language learning. Finally, I highlight some converging themes emerging from all three approaches and outline the questions this dissertation sets out to address.

1.1. Typological approach to linguistic universals

Traditionally, the study of linguistic universals has drawn evidence primarily from cross-linguistic and historical typological data. This approach has been productive in identifying a large number of language universals (Croft, 2003; Dryer, 1992; Greenberg, 1963) as well as linking typological data to processing preferences (Hawkins, 2004). The development of large data sets such as *The World Atlas of Linguistic Structures* (Dryer & Haspelmath, 2011), substantial increase of newly documented languages, and developments in linguistic theory (Croft, 2003; Dryer, 1998) that originated within this approach have prompted a shift away from categorical notions (e.g., absolute linguistic universals, narrowly defined language types, etc.) towards a probabilistic view of language structures, where language diversity is viewed as typological distributions closely interrelated with distributions of other factors such as gradient universals of human cognition, cultural underpinnings, and population movements.

However, an approach relying solely on typological data has two serious limitations. The first is the sparsity of independent data points. Most, if not all, languages are directly genetically related (in the sense of evolving from common ancestors) and thus would be expected to share properties. Language contact can further diminish the independence of languages since languages that remain in contact (sometimes extending over centuries) tend to share lexical and structural properties. A famous example of areal influences is the Balkan Sprachbund, where geographically contiguous but genealogically unrelated languages (Macedonian, Bulgarian, Greek, Albanian, and Balkan Romance) have acquired a number of common features often referred to as *Balkanisms* in phonology, morpho-syntax, and lexicon (Friedman, 2006).

Both genetic and areal (language contact) influences drastically reduce the effective sample size available for statistical tests of hypothesized universals –a challenge to typological approaches that has long been recognized (Dryer, 1989), but has only recently begun to be addressed (Cysouw, 2010; Dryer, 1989; Dunn, Greenhill, Levinson, & Gray, 2011; Jaeger, Graff, Croft, & Pontillo, 2011; Rafferty, Griffiths, & Klein, in press). Using advanced statistical methods, some of this recent work has called into question typological generalizations that were long assumed to hold (e.g., Dunn et al., 2011).

Second, while the typological approach has been instrumental in documenting a large number of language universals and identifying their fine-grained structure, it is in principle limited in that it cannot directly answer questions concerning the *origin* of linguistic universals. For example, if language universals are indeed shaped by constraints on the human cognitive system, how do these constraints enter the linguistic system and come to shape it over time? Do they originate during language acquisition (i.e., before a mature language system is in place) or through language use after the users' language has matured?

As discussed next, both of these shortcomings of typological approaches can be addressed by complementing typological work on language universals with behavioral evidence from language learning experiments (for a similar perspective, see also Tily and Jaeger, 2011) and computational approaches that use multi-agent simulations to study how hypothesized language universals emerge and evolve over time (e.g., Niyogi, 2006; Steels, 1995, 2006).

Computational approaches span a variety of methodologies and theoretical standpoints, but typically fall into two broad types. In one class of simulations referred to as *language games*, language structures emerge through communication in a population of agents, all of whom have certain aspects of language (e.g., a lexicon or biases towards certain constructions) hard-coded in their 'genome'. Agents, who are more successful at communicating, typically have an advantage in propagating their 'genes' across generations (Cangelosi & Parisi, 1998; Nyogi & Berwick, 1997). The other class of models (*iterated learning models*) creates a chain of 'parent'- 'child' agents, where the language output produced by the 'parent' becomes the input to the 'child' (Kirby, 1999; Smith, Kirby, & Brighton, 2003). These models set out to show that if a population of agents has certain pre-defined weak biases, these biases would be amplified through the number of generations, causing a shift towards a language that expresses these biases in its grammar. The behavior of simulated agents often mirrors human performance. For example, one of the major results of the iterated learning models – the emergence of compositional structure (Smith, Brighton, & Kirby, 2003) – has been replicated in a series of studies with human participants (Kirby, Cornish, & Smith, 2008; Smith & Wonnacott, 2010). The remainder of this chapter will focus on biases that human learners bring to language acquisition and will not discuss computational approaches in further detail (but see Steels (2011) for a recent comprehensive review of computational approaches to language universals).

1.2. Language learning approaches to linguistic universals

Hypotheses about underlying causes of language universals are often tested by directly studying learning outcomes during language acquisition. While views differ on whether language learning is constrained by strong biases imposed during acquisition (Chomsky, 1965; Fodor, 2001; Pinker, 1984) or by weaker constraints arising from acquisition and use (Bever, 1970; Christiansen & Chater, 2008; Deacon, 1997; Hawkins, 2007; Newport, 1981b, 1990), researchers in both traditions are in agreement that language acquisition can provide a vehicle for the existence of language universals. As a result, both approaches have generated a substantial body of behavioral evidence, providing insights into how typological generalizations emerge. In particular, this work has uncovered striking parallels between language development in child learners and phenomena observed in typology.

1.2.1. First language acquisition

The major challenge in the study of learning biases in natural languages is that language development, cognitive development, and linguistic experience are typically intertwined and the impact of each of these factors on the learning outcomes is hard to isolate in a naturalistic setting.

One approach, often associated with the functionalist tradition (e.g., MacWhinney & Bates, 1989; Slobin, 1973, 1977; Slobin & Bever, 1982), addresses this problem by tracking the order of acquisition of linguistic devices that express the same semantic notions (e.g., locative expressions or direct objects) in several unrelated languages. As argued here, there is little reason to assume that the rate of development of semantic notions should differ across languages, and therefore cross-linguistic differences in language acquisition could shed light on the origin of some typological generalizations. Another approach, often associated with the generative tradition, tries to draw inferences about language universals by analyzing 'mistakes' infants and children make during first language acquisition (Crain, Goro, & Thornton, 2006; Hyams, 1983).

Shared by both traditions is the assumption that, if learners have some prior biases about natural language structures, then the ungrammatical constructions they might use before fully mastering the target language should gravitate towards constructions that are common across languages of the world.

1.2.2. Language acquisition in atypical populations

While evidence from first language acquisition approaches suggests strong parallels between patterns observed in typology and learning, it does not directly support the hypothesis that language universals originate during language acquisition since it is typically unclear whether learning biases are strong enough to bring about a change in the linguistic system and thus shape language structures over time.

A small number of studies have addressed this question by exploring language development in atypical populations. Instead of studying language acquisition in normal circumstances where learners are exposed to very rich input, this tradition has sought to identify exceptions to this scenario and traced language acquisition in learners who received impoverished or highly inconsistent language input such as deaf children developing gestural communication when their parents do not sign (Goldin-Meadow & Mylander, 1983, 1998; Singleton & Newport, 2004) or learners exposed only to pidgin-like languages (Senghas & Coppola, 2001; Senghas, Coppola, Newport, & Supalla, 1997).

These studies provide direct evidence for the changes introduced by learners who are exposed to languages that do not conform to typologically widespread patterns and thus provide a potential insight into the direction of language change. For example, evidence from Nicaraguan Sign Language (NSL) that developed autonomously in a community of signers without effective contact with other languages suggests that changes introduced by a few generations of learners can gradually cause a language to develop properties shared by many fully-fledged linguistic systems. Senghas and colleagues (Senghas & Coppola, 2001; Senghas et al., 1997) studied the emergence of spatial grammatical devices (typically used as agreement systems in developed sign languages) in successive cohorts of NSL signers. They found that spatial modulations became more frequent and increasingly more consistent as they passed through several generations of learners. Similarly, studies investigating language development in deaf children (discussed in more detail in Section 1.3.) find that young learners who receive atypical or reduced language input restructure the languages they are learning (Singleton & Newport, 2004) or create their own communicative systems (Goldin-Meadow & Mylander, 1983, 1998; Goldin-Meadow, Mylander, de Villiers, Bates, & Volterra, 1984), whose properties closely resemble phenomena frequently observed cross-linguistically. These findings suggest that at least some patterns commonly observed in typology may originate during language learning

The data obtained in these studies is, however, highly limited by the availability of suitable populations, which often prevents large-scale cross-linguistic investigations and constrains the range of phenomena that can be addressed in the given paradigm.

Additionally, all first language acquisition approaches to language universals face the challenge of a general lack of control over the input: It is almost impossible to have a complete picture of the amount of language exposure and the frequency of grammatical structures of interest that learners receive prior to testing, which creates a potential confound that needs to be taken into consideration while interpreting the results. Recent work in language development has begun studying the input learners receive by using head-mounted eye-tracking and video recording of child interactions with care-givers and the environment in naturalistic settings (M. Frank, Simmons, Yurovsky, & Pusiol, 2013; Pusiol, Soriano, Fei-Fei, & Frank, 2014; Roy et al., 2006, and work by Elika Bergelson). These promising new developments would allow researchers to test hypotheses about the precise role of language input in learning outcomes, although they pose serious practical challenges concerning analyzing large-scale video datasets.

1.2.2. Miniature artificial language learning approaches to linguistic universals

An alternative to first language acquisition approaches in studying linguistic universals is presented by a miniature artificial language learning paradigm where participants (infants, children or adults) are exposed to languages that are experimentally designed in the laboratory to have certain properties of interest, which allows researchers to obtain precise control over the stimuli and isolate the input dimensions of interest. These miniature languages are small enough to be acquired in the lab within a short period of time, with exposure typically ranging from several minutes for simpler languages (Saffran, Aslin, & Newport, 1996) to several one-hour sessions distributed over several days for more complex languages that generally involve a meaning component (Amato & MacDonald, 2010; Fedzechkina, Jaeger, & Newport, 2012; Hudson Kam & Newport, 2005, 2009; Wonnacott, Newport, & Tanenhaus, 2008). In such experiments participants are typically asked to listen to a sound stream often accompanied by visual stimuli and respond to questions about them (e.g., rate how familiar the stimuli are, answer comprehension questions, etc., see Figure 1.1 for an example).

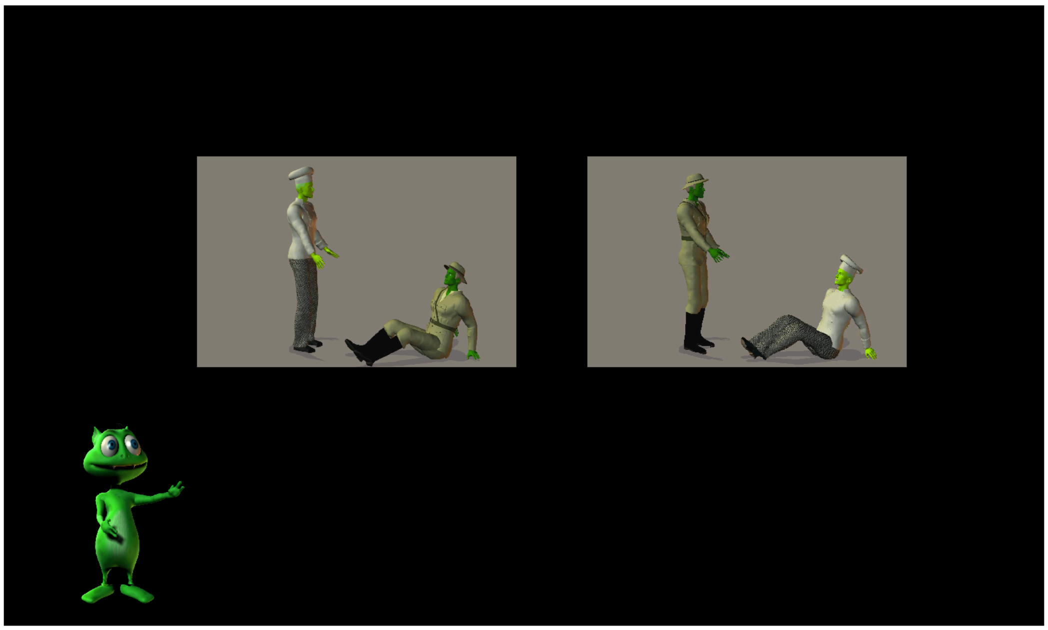

Figure 1.1: Example of a comprehension test used in miniature artificial language experiments.

Although the first miniature artificial language learning experiments were conducted over 30 years ago (Cook, 1988; MacWhinney, 1983; Morgan, Meier, & Newport, 1987; Reber, 1967), it was work in the late 1990s that established miniature artificial language learning as a standard method in language acquisition research. Most of the pioneering work using this paradigm has focused on studying the information learners can extract from the input to acquire the underlying grammar and prior biases that learners bring into the acquisition process. For example, Saffran et al. (1996)

exposed 8-month old infants to a stream of nonsense words (pibadu-budaka-….) that contained no information about word identity except transitional probabilities between the syllables (high within the word and low at word boundaries). After as little as 2 minutes of passive exposure to this sound stream, infants could successfully discriminate nonsense words from part words suggesting that learners were exquisitely sensitive to the distributional information present in the input. Later work has established that infants can successfully extract higher order regularities between items in the input that have no common surface features and generalize them to previously unseen examples (Gerken, 2006; Marcus, Vijayan, Bandi Rao, & Vishton, 1999).

Over the years researchers have expanded the paradigm to study increasingly more complex phenomena such as acquisition of constituent order (Christiansen, 2000; Culbertson, Smolensky, & Legendre, 2012; Thompson & Newport, 2007; Tily, Frank, & Jaeger, 2011; Wonnacott et al., 2008) and morphology (Culbertson & Legendre, 2010; Hudson Kam & Newport, 2005, 2009). This paradigm has also been combined with current sentence processing methodologies such as eye-tracking (Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Wonnacott et al., 2008) and self-paced reading (Amato & MacDonald, 2010; Karuza, Farmer, Fine, Smith, & Jaeger, 2014) to study how newly acquired distributional information is used in real time during incremental processing of novel languages.

These developments made it apparent that a miniature artificial language learning paradigm is well-suited to study syntactic and morphological universals, which have traditionally been of great interest in linguistic typology (Bickel, 2011; Cysouw, 2011;

Dryer, 1992). This has caused a rapid revival of interest in applying this paradigm to cross-linguistic generalizations (e.g., to phonology: Finley & Badecker, 2008, Wilson, 2006; word formation: Newport & Aslin, 2004; morphology: Fedzechkina, Jaeger, & Newport, 2012; Hudson Kam & Newport, 2005, 2009; Hupp, Sloutsky, & Culicover, 2009; St Clair, Monaghan, & Ramscar, 2009; syntax: Morgan, Meier, & Newport, 1987; Tily, Frank, & Jaeger, 2011; Thompson & Newport, 2007).

Conclusions about the relationship between language learning and language structure in miniature artificial language research are typically drawn from data obtained using two somewhat different paradigms (see Figure 1.2). The most widely used approach compares the learnability of two or more artificial languages that differ in certain crucial properties. In these experiments, participants are typically trained on a subset of items generated by the underlying artificial grammar. Participants' performance is subsequently assessed based on how well they generalize the newly acquired artificial grammar to previously unseen data (Finley & Badecker, 2008; Tily, Frank, & Jaeger, 2011) or discriminate between grammatical and ungrammatical items in the novel language (Reeder, Newport, & Aslin, 2009, 2010). The logic behind this approach is as follows: If grammatical structures that are acquired more easily in the miniature artificial languages mirror typologically frequent patterns, then these cross-linguistic patterns could have originated in human cognitive mechanisms. Indeed, the recurrent finding from this paradigm is that typologically common phenomena are acquired faster and more easily than less common ones at all levels of linguistic organization (e.g., phonology: Finley & Badecker, 2008, Wilson, 2006; word formation: Newport & Aslin, 2004;

morphology: Hupp, Sloutsky, & Culicover, 2009; St Clair, Monaghan, & Ramscar, 2009; syntax: Morgan, Meier, & Newport, 1987; Tily, Frank, & Jaeger, 2011). Some of these studies, however, test phenomena that are present in the learners' native language, which needs to be taken into account when interpreting the obtained results. For example, Hupp et al. (2009) and St Clair et al. (2009) find that miniature artificial languages that are exclusively suffixing are acquired more accurately by native English speakers than artificial languages that have prefixing only. Typologically, however, English is a strongly suffixing language (Dryer, 2013), and thus, it is unclear whether the observed learning outcomes are due to a cognitive bias favoring suffixing over prefixing as the authors claim or are simply caused by a transfer from participants' native language. I return to the role of native language background in miniature artificial language learning in more detail in Section 1.4.

The second, more recent and arguably more powerful, miniature artificial language learning paradigm has sought to address questions about linguistic universals by investigating whether learners who are exposed to artificial languages that do not conform to cross-linguistically common patterns would introduce changes into the languages they learn, making them more closely aligned with naturally observed types (Culbertson et al., 2012; Fedzechkina et al., 2012; Hudson Kam & Newport, 2005, 2009). The common finding from this paradigm is that children are generally willing to deviate from the input, while adult learners typically match the statistics of the input very closely (Hudson Kam & Newport, 2005, 2009). However, in the rare cases when adults do deviate from the input they receive, their productions tend to reflect typologically more common patterns (Culbertson et al., 2012; Fedzechkina, Jaeger, & Newport, 2011; Fedzechkina et al., 2012; Fedzechkina, Jaeger, & Newport, 2013, which form part of the studies reported in this dissertation).

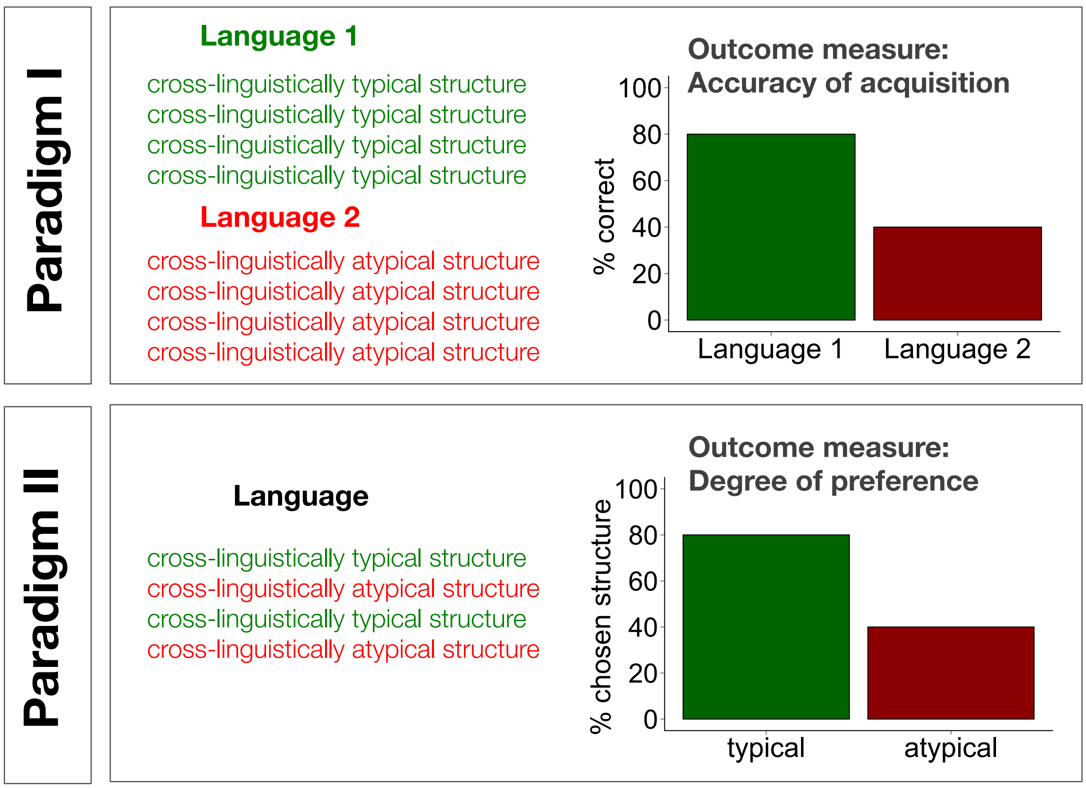

Figure 1.2: Two paradigms typically used in artificial language learning research.

The miniature artificial language learning approach is not without its limitations. Miniature languages need to be simple enough to be learnable within at most several short visits to the lab, and thus lack the complexity of natural languages on virtually every dimension. They usually contain very few words (typically between 4 and 15, but see M. Frank, Tenenbaum, & Gibson (2013) for a 'miniature language' with 1000 words) which have little variation in length and syllabic complexity and are often stripped of many acoustic cues natural languages have such as pitch changes and relevant rhythmic patterns. Sentences or phrases in artificial language learning experiments generally express fairly simple meanings such as describing simple transitive events or object locations. This raises the question of whether the learning outcomes observed in miniature artificial language learning are representative of natural language acquisition.

While this question cannot be definitively answered without extensive further research, the parallels between patterns in typology and learning outcomes in miniature artificial language learning are numerous and striking. This suggests that a miniature artificial language learning paradigm can be used to complement typological data and provide independent behavioral evidence in arguments for or against language universals. The similarity in patterns observed in natural languages and miniature language learning further suggests that this paradigm is powerful enough to uncover the biases learners have during language acquisition and thus to directly probe the origin of language universals. Furthermore, the miniature language learning paradigm has a unique ability to test whether learners' biases are strong enough to introduce change into a linguistic system. Recent developments in iterated language learning (Kirby et al., 2008; Smith & Wonnacott, 2010), where the output of one learner becomes the input to another generation, have allowed researchers to study how biases observed in one generation of learners spread diachronically in the linguistic system, which further underscores the suitability of miniature language learning applications to typological research.

As mentioned above, the parallels between learning performance in miniature artificial language learning, first language acquisition, and typological data are numerous.

To highlight some converging themes that emerge from these approaches, in the next session I will review evidence for two linguistic universals that have received a lot of attention both in typology and language learning: the cross-linguistic preferences for conditioned variation and for consistent headedness.

1.3. Converging evidence for linguistic universals

Evidence from a large number of typological studies (Dryer, 1992; Greenberg, 1963; Hawkins, 1983) and work in theoretical syntax (Chomsky, 1988; Jackendoff, 1977) suggests that languages tend to *consistently* order dependents with respect to their heads: Constituent orders where heads either all precede or all follow their dependents tend to be preferred cross-linguistically. A well-attested example of this preference for consistent headedness is the implicational universals describing a correlation between the order of elements in the verb phrase, adpositional phrase, genitive phrase, and the position of the relative clause with respect to its head noun. Cross-linguistically, languages that have verb-final constituent order typically also have noun-postposition, genitive-noun, and relative clause-noun orders, while verb-initial languages tend to have preposition-noun, noun-genitive and noun-relative clause orders (Dryer, 1992; Greenberg, 1963). Recent work, however, has failed to find evidence for these universals after the genetic dependencies between languages are appropriately accounted for (Dunn et al., 2011), thus calling into question the reality of these cross-linguistic patterns.

One of the early miniature artificial language learning studies conducted by Cook (1988) investigated whether learners of miniature artificial languages exposed to one grammatical construction of interest would generalize it to another construction of interest in line with the cross-linguistic preference for consistent headedness. For example, some learners were exposed exclusively to verb phrases and then tested on adpositional phrases. The results provide only partial support for the preference for consistent headedness. Cook found, in particular, that learners of both verb-final and verb-initial orders generalized to noun-postposition order, which goes against typological data for the second group of learners. However, there was a significant difference in the *degree of preference* for the noun-postposition order: In line with typological data, learners exposed to the verb-final constituent order chose the noun-postposition order significantly more often. It is possible, however, that the somewhat mixed pattern of results is due to the methodological shortcomings of the study. For instance, learners were given an English translation for every artificial sentence during training and test, so it is possible that they experienced interference from their native language, which made the results less clear-cut.

More recently, Christiansen (2000) investigated the same bias for consistent head ordering using a somewhat different paradigm. In the experiment, two groups of participants were exposed to nonsense letter strings (e.g., VVQXQXS or VQQVXXS) that were generated by either a consistent head ordering grammar (where all phrases were head-final) or an inconsistent head ordering grammar (where some phrases were head-final and some head-initial). At test, participants were asked to classify previously unseen strings as grammatical or ungrammatical. In line with typological data, participants exposed to artificial languages that contained consistent head orderings performed significantly better than participants trained on languages with inconsistent headedness.

Culbertson et al. (2012) further suggested that learners of miniature artificial languages that have probabilistic inconsistencies in head ordering are likely to restructure the input to make headedness more consistent in the newly acquired languages. Culbertson and colleagues explored whether learners' biases favor Greenberg's Universal #18 (Greenberg, 1963) that concerns the order of numeral, adjective and noun – prenominal adjectives and postnominal numerals tend to be dispreferred cross-linguistically, while other orders of adjective and numeral with respect to the noun are well-attested. Culbertson and colleagues found that learners acquiring a miniature artificial language with variable order of numeral and adjective with respect to the noun tend to regularize typologically frequent patterns: They strongly prefer the so called harmonic patterns (where the adjective and numeral occur either before or after the noun) and a less common but still widely attested pattern where the numeral precedes the noun and the adjective follows it. No regularization was, however, observed for the opposite typologically infrequent order.

Another linguistic universal that is widely supported by typological and language acquisition data is a preference for consistent rule-based linguistic systems that have *predictable* variation. It is highly unlikely for a language to have several forms that carry the same meaning and are in free variation in the same context. Instead, natural languages tend to have predictable variation where the use of competing forms is conditioned by semantic, pragmatic, phonological, and other factors (Givon, 1985; Labov, 1963).

Support for this cross-linguistic property of human language comes from developmental studies showing that children acquiring consistent systems in their native language typically find them easier to learn compared to systems with a lot of idiosyncrasy. For example, Slobin (1973) found that Serbo-Croatian and Hungarian bilingual children first mastered the locative expressions in Hungarian which has consistent locative inflections, while the full mastery of Serbo-Croatian locatives that involve an inconsistent combination of often ambiguous prepositions and case inflections appeared later in their speech. Monolingual children acquiring Turkish (a language with free constituent order and rich morphology) have been shown to acquire the nominative/accusative case inflections before the basic constituent order (Slobin & Bever, 1982). A possible explanation for this asymmetry is that the case system in Turkish is consistent and displays little case syncretism. In contrast, Turkish constituent order is highly variable and allows scrambling (Kornfilt, 2003). The learners' preference for consistency has also been shown in the early productions of children learning free constituent order languages such as Russian: They typically start out using one constituent order variant while other possible orders appear later (Slobin, 1973, 1977).

A similar bias against linguistic systems with unpredictable variation is evident in learners receiving atypical input. Singleton and Newport (2004) followed language development in a deaf child named Simon. Simon's ASL input came exclusively from his non-native signing parents, whose speech contained a lot of inconsistencies as typical for late learners. However, Simon's performance on most tests was comparable to his peers who acquired ASL from native signers, suggesting that Simon restructured the input he received and introduced a more consistent system in his own productions.

Similar observations were made by Goldin-Meadow and colleagues (Goldin-Meadow & Mylander, 1983, 1998; Goldin-Meadow et al., 1984), who studied the emergence of simple gestural systems used by deaf children to communicate with their hearing parents (so called *home sign* systems). In a series of studies, they found that profoundly deaf children who received no sign language input tended to develop *structured* gestural communication systems that had clear characteristics of natural languages (e.g., relatively consistent constituent order and recursion). Importantly, the gesture order used by these children was not modeled after the caregivers whose gesture use was highly inconsistent and often less complex than the children's gestural systems, suggesting that word order consistency was created 'from scratch' by the learners.

These results are supported by laboratory studies conducted by Hudson Kam and Newport (Hudson Kam & Newport, 2005, 2009) with hearing participants. Adult and child learners in their experiments were exposed to miniature artificial languages that contained unpredictable variation: All nouns in the experiment were followed by nonsense determiners that varied probabilistically in the input. Young learners readily regularized the inconsistent input they received and used one of the determiners in almost all of their productions. The behavior of adult learners was less clear-cut: They mostly reproduced unpredictable variation present in the input but showed regularization behavior when the complexity of the system was very high.

Laboratory studies also suggest that biases that are too weak to be reliably detected in one generation of learners (e.g., the lack of regularization in adult learners observed by Hudson Kam and Newport) can result in a more regular system after it is passed down from generation to generation. Such accumulation of biases is typically experimentally studied using the iterative learning paradigm (Kirby et al., 2008; Smith & Wonnacott, 2010), which simulates cultural transmission of language by using the output of one learner as the input for the next learner. For example, Smith and Wonnacott (2010) used this approach to study regularization of unpredictable variation in noun plural marking similar to the type of probabilistic variation in the experiments by Hudson Kam and Newport (Hudson Kam & Newport, 2005, 2009). Indeed even if the productions of the first generation of learners exhibited significant unpredictable variation, this variation became conditioned on the noun, thus resulting in a regular linguistic system (characterized by zero conditional entropy of plural marking) only after five generations of learners.

1.4. The current proposal

As outlined above, the parallels between typological data and learning outcomes during language acquisition of natural and miniature artificial languages are ample. None of these approaches are without their limitations: Typological data are sparse and have difficulty answering questions about the origin of cross-linguistically frequent patterns; fist language acquisition approaches have little control over the input learners receive, which may introduce potential confounds; and miniature artificial language learning is limited by the complexity of languages that can be used since they need to be acquired within a short time period. Crucially, however, all these approaches are complementary since they have different limitations.

Using language learning approaches to complement typological data was advocated by Dan Slobin over 30 years ago (Slobin, 1977). He suggested in particular that the processes underlying language change and ultimately leading to the observed synchronic distribution of grammars are subject to the same set of constraints as language acquisition. Thus, Slobin argued, the study of cross-linguistic universals would necessarily benefit from understanding biases that are at work during language acquisition.

This interdisciplinary research program, however, did not gain much traction in language universals research, which continued in two separate traditions in linguistics and psychology. Linguistics was traditionally interested in implicational universals and primarily relied on typological data, while psychology typically employed learning approaches to investigate broader non-implicational universals such as the presence of compositional structure or conditioned variation in the language (but see Christiansen, 2000; Cook, 1988; Culbertson & Newport, in prep; Culbertson et al., 2012 for notable exceptions).

One goal of this dissertation is to bring the two disciplines closer together and to behaviorally test several implicational universals that have been widely discussed in the typological literature, but have not received much attention in language acquisition. An interdisciplinary approach to linguistic universals is of special importance in light of the recent increase in interest in the purely historical account of language universals, described next.

On the historical view, properties that appear to be universal are better explained as a product of historical evolution that originates through historic accidents such as common ancestry, geographic proximity, macro-areal contact, etc. (Dunn et al., 2011; Putnam, 1971). This hypothesis has been recently revisited in biology and linguistics. For example, the suggestion that language universals are the vestiges of the ancient Ur-language, from which all human languages ultimately originated (Putnam, 1971) has received some support from studies arguing that human populations and languages have originated in a small group that migrated from Africa (Atkinson, 2011; Cavalli-Sforza, 1997; but see Bowern, Hunley, & Healy, 2012 and Cysouw, Dediu, & Moran, 2012 for a critique of this position). Other support for these accounts comes from recent work applying statistical models borrowed from evolutionary genetics to typological diachronic data sets that failed to find evidence for well-established implicational universals in word order and morpho-syntax (Bickel, Witzlack-Makarevich, & Zakharko, in press; Dunn et al., 2011). Thus, complementing typological research on implicational universals with behavioral data from learning approaches is of special interest since it is this type of language universal that has been called into question.

In this thesis I will use the miniature artificial language learning paradigm to ask in particular to what extent biases observed during language acquisition can impact the structure of natural languages, thus creating a seed for language universals. In all experiments reported here, learners will be exposed to languages that contain several probabilistically occurring forms, which express the same meaning (conveyed by accompanying short videos). At test, learners will be asked to produce their own utterances in novel languages and their systematic deviations from the input frequency of each competing form (if observed) will be taken as evidence for language change introduced by learners as a result of a learning bias.

The reason to incorporate variability into input languages is two-fold. First a large body of work in artificial language learning suggests that learners are not likely to introduce innovations into perfectly consistent languages, at least within the short amount of time spent in the laboratory (e.g., Christiansen, 2000; St Clair et al., 2009; Tily et al., 2011). The input containing several grammatical structures that express the same meaning creates a situation similar to language change or a pidgin. When there is a lot of variability present in the linguistic system such as during creolization, language change can happen as fast as within one generation of speakers. Thus, incorporating variability into artificial languages is likely to make learning biases manifest within a short period of time (e.g., several 30-minute visits to the lab). Second, assessing learners' preferences for one of the two competing forms will allow us to not only gain insight into synchronic preferences learners might have, but also into how these preferences may change the linguistic system over time. These deviations from the input introduced by learners can create a seed for language universals. If these deviations are picked up across the population and spread over multiple generations of speakers, they can eventually cause language change towards a system that explicitly reflects the biases learners have (see Figure 1.3 for a schematic representation of this scenario). In this way, the paradigm employed in this dissertation contributes to the question of how biases operating during language learning change languages over time, thereby potentially causing the observed cross-linguistic patterns.

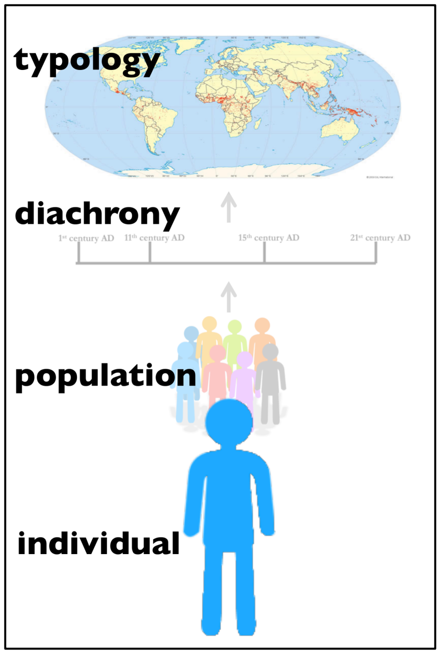

Figure 1.3: Hypothesized relationship between individual-level biases and language structures.

The logic behind this approach is as follows: If the input distributions of competing forms do not a priori bias learners in a certain direction, then the observed deviations from the input *that are not reducible to native language biases* would be indicative of more general biases about natural language structure learners may have.

I highlight the issue of controlling learners' native language biases, as it has not yet received due attention in artificial language research (but see Goldberg, 2013). As this methodology is used increasingly to study the impact of learning biases on linguistic structures, understanding and controlling the scope of native language influence on the acquisition of artificial languages becomes more pressing. While many studies have raised the issue of native language influences (Culbertson et al., 2012; Fedzechkina et al., 2011; Tily et al., 2011; Wonnacott et al., 2008), systematic investigations of these questions have only recently begun in artificial language research. There is, however, a long tradition of using artificial language learning to investigate the role of native language background in second language learning (see, for instance, Pajak, 2010). It remains to be determined how much transfer from the native language takes place in these experiments and the exact circumstances leading to it, which will inevitably involve comparing the performance of speakers with different language backgrounds on the same artificial language learning tasks.

So far studies linking learning biases to language structures have employed different types of controls. Some studies (e.g., Christiansen, 2000; Finley, 2011) tried to assess the degree of structural preference stemming from native language biases by having a control group of learners who were tested on artificial grammar without receiving prior training in the novel language. Any preference observed in such group would stem from prior native language experience. Others (e.g., Culbertson et al., 2012), exposed a control group of learners to a random mixture of patterns to achieve the same goal. In some sense, however, these controls might be ineffective. It is possible, for example, that learners' biases stemming from their prior experience are too weak to be captured by the random or zero-exposure conditions while still affecting other conditions of interest (see Goldberg, 2013 for similar and additional arguments).

To mitigate native language biases as a potential confound, all experiments test the acquisition of phenomena not present in the learners' native language. I will focus, in particular, on several phenomena involving the acquisition of case-marking including the trade-off between constituent order flexibility and case as well as optional case-marking systems. All participants in these studies are monolingual native speakers of English (a language with no productive nominal morphology), who are not fluent in other languages and have not learned other languages before the age of 12. Therefore, any deviations from the input languages (designed in such a way as not to bias learners towards any particular structure) in the direction of typologically frequent patterns could be taken as evidence for more general learning biases that a) cannot be reduced to transfer from the native language and b) are unlikely to be due to transfer from non-native languages.

The ultimate goal of this dissertation is to go beyond establishing a parallel between biases observed during acquisition and typological distributions. Rather the goal is to gain insight into the nature of these biases. Before I introduce the experiments, I will briefly review current accounts that link individual-level biases operating during language learning or language use to structures observed in natural languages. Despite the recent increase in interest in purely historical accounts, they are by no means the dominant view in the field. Most influential accounts, in fact, assume that language universals in one way or another can be traced back to a set of constraints on human cognitive mechanisms. The exact nature of these constraints remains a matter of debate. Some argue that language universals are due to innate arbitrary constraints specific to language acquisition and are not shared by other aspects of human cognitive systems

(Chomsky, 1965; Fodor, 2001; Pinker, 1984). Others suggest that cross-linguistically frequent patterns are more likely to be shaped by multiple soft constraints as well as biases associated with language use and communication (Bever, 1970; Christiansen & Chater, 2008; Deacon, 1997; Hawkins, 2007; Newport, 1981b, 1990). Next, I discuss these different types of accounts in turn.

1.5. Cognitively-arbitrary constraints

On this view, innate linguistic-specific knowledge predisposes learners to acquire grammatical structures that conform to a set of hard-wired universal structural principles. Under this hypothesis, grammars that conform to these innate properties are easily learnable and are thus more likely to persist cross-linguistically, while grammars that deviate from these constraints are hard or even impossible to learn and are unlikely to arise (Chomsky, 1965; Fodor, 2001; Pinker, 1984). For example, one formulation (Chomsky, 1981) postulates a set of universal linguistic principles that hold in all natural languages and a number of binary parameters capturing cross-linguistic variation, whose values are set by the learner based on the evidence received from the input.

The fundamental assumption underlying this approach is that the capacity for language is independent of other human cognitive capacities (e.g., memory, perceptual-motor system, processing, etc.). Thus, many constraints on language that are unique to this system are fundamentally arbitrary, and cannot be due to functional considerations.

These accounts provide a plausible explanation of how cross-linguistic generalizations emerge and also offer an elegant solution to some open questions in language learning. Language acquisition is an inductive inference problem – to successfully acquire a language, learners must infer the grammar from available data by testing hypotheses about the underlying structure. The problem, known as the *poverty of stimulus,* as argued by some (e.g., Chomsky, 1965) is that the data available to learners might not provide enough evidence to reject incorrect hypotheses since a) learners are only exposed to a small finite sample of structures possible in a language, b) the input does not contain the evidence needed to reject incorrect hypotheses (i.e., negative evidence) since caregivers typically correct the content but not the form of children's utterances and c) the input is full of noise. And yet, children do converge on the grammar of the target language fairly quickly and thus, as argued by the supporters of the nativist view, must have innate constraints that guide their exploration of the hypothesis space.

Without necessarily calling into question the fundamental tenets of this perspective, recent work has, however, uncovered that the input learners receive is richer than originally assumed. In particular, learners are highly sensitive to statistical patterns in the input and are able to extract these regularities from noisy data (Gerken, 2006; Marcus et al., 1999; Reeder, Newport, & Aslin, 2013; Saffran et al., 1996). While explicit negative evidence is lacking in the input, findings from computational modeling (Perfors, Tenenbaum, & Regier, 2011) and statistical learning suggest that learners are able to exploit this distributional information present in the input in lieu of negative evidence to inform their hypotheses about possible grammars (Brooks & Tomasello, 1999; Goldberg, 1995; Reeder et al., 2013).

Another problem with the nativist account – at least in its original categorical formulations – is directly pertinent to language universals: If humans are indeed born with innate machinery hard-wired for language, why are absolute language universals nearly non-existent? This problem was also recognized by proponents of nativist accounts and later formulations of this hypothesis have tried to move away from postulating a large system of arbitrary constraints. For instance, the minimalist program suggests that only the computational system underlying language is universal. On this more recent view, only recursion is argued to occur in all human languages (Hauser, Chomsky, & Fitch, 2002). However, even the absolute nature of recursion has been questioned by some (Everett, 2005, 2009).

1.6. Cognitively-motivated constraints

Many researchers have argued that languages are shaped by multiple weak cognitively-motivated constraints that originate at the level of the individual and ultimately give rise to observable cross-linguistically frequent patterns (Bever, 1970; Christiansen & Chater, 2008; Deacon, 1997; Hawkins, 1994, 2004; Newport, 1981b, 1990). On this view, cognitive factors (such as memory, perception, etc.) influence but do not determine how languages develop over time. That is grammatical structures that are more easily perceived, learned, or processed are more likely to be used more frequently by a single speaker and to be picked up at the population level. As these preferred structures get passed down across several generations of learners, they become conventionalized (cf. Bybee, 2003; Bybee & Hopper, 2001; Newport, 1981b) and become part of the grammar.

Unlike the nativist approach, these accounts can easily explain the statistical nature of the majority of linguistic universals (in fact, on this view language universals are *expected* to be inherently probabilistic, cf. Hawkins [2004]). All else equal, speakers should slightly prefer patterns that are more advantageous according to some metric (e.g., more easily perceived, processed, etc.). However, the situation when everything else is equal rarely (if ever) arises in natural languages. First, the population level and diachronic processes are critical for this view. Synchronic and diachronic language structures are not only influenced by cognitive constraints, but also by social factors and historical accidents. Certain constructions can become more popular not only because they are more advantageous for cognitive processing, but also because they have higher prestige among certain groups or due to language contact. For example, Heine and Kuteva report that perfect tense forms using the auxiliary *have* ('I have finished' in English, 'ho finito' in Italian, or 'habe geendet' in German) arose as it became fashionable after the fall of the Roman Empire to coin perfect tense based on the meaning of the verb *have* (Heine & Kuteva, 2006). Less frequent word order patterns can often become dominant as a result of language contact -- if they minimize the distance between the two contact languages (Heine, 2008). Second, there are multiple cognitive constraints that compete with each other to begin with so it is possible that the same structures would be more advantageous on one dimension while being less advantageous on the other. Thus, different languages may find different solutions for the same problem.

What are some of these constraints? Language learnability has received much attention in the literature (Kirby, 1999; Kirby et al., 2008; Newport, 1981a, 1982). As discussed above, language learners, who are exposed to a finite amount of input data, ultimately acquire a system for producing and comprehending an infinite set of meaningful utterances in a given language. Thus, as learnability accounts argue, natural languages appear to have certain properties such as a high degree of internal structure that make them more easily generalizable to novel instances. Views differ, however, on the nature of mechanisms that cause languages to become increasingly more learnable and more structured.

One position, prominently advocated by Simon Kirby, Morton Christiansen, and colleagues (Christiansen & Chater, 2008; Griffiths & Kalish, 2007; Kirby, 1999; Kirby et al., 2008; Reali & Griffiths, 2009), suggests that language is inherently a product of cultural evolution and as such biases at the individual level alone cannot explain cross-linguistically common structural properties. On this view, changes to the linguistic systems resulting from individual-level biases are amplified as they percolate across multiple generations of learners, gradually causing the linguistic system to become increasingly more structured and more learnable (the process termed *cumulative cultural evolution*). Evidence for this approach comes from studies using the iterative learning paradigm, which simulates cultural transmission of language by using the output of one learner as the input to another learner either in laboratory settings (Kirby et al., 2008; Smith & Wonnacott, 2010) or via computational models (Kirby, 1999; Reali & Griffiths, 2009). For example, Kirby et al. (2008) found that after passing through 10 generations of learners a system of initially random strings used to describe object color, shape, and motion resulted in compositional structure characteristic of natural languages where the random strings were reanalyzed as 3 morphemes consistently expressing object color, shape, and motion.

One caveat about the findings emerging from this paradigm is that learners' behavior could be due to their native language biases: Since natural languages have compositional structure, learners might be introducing a system already familiar to them from their native language into the newly acquired language (see also the discussion of L1 influences above). A more general problem with Kirby's account is that it implicitly assumes that individual learning biases cause a shift in the linguistic system (if there were no a priori preferences, nothing would be amplified during cultural transmission) but remains agnostic as to what these biases are. Thus, while it can explain *how* cross-linguistic generalizations spread diachronically, it cannot explain *where* they ultimately originate. Other proponents of cultural transmission, notably Christiansen and Chater, have explicitly identified several types of domain-general constraints that are expected to play a role in shaping language structures: pragmatic, perceptual-motor constraints, constraints from thought, and constraints on processing and learning (Christiansen & Chater, 2008).

Another prominent account that explicitly links cross-linguistic structural properties to first language acquisition was proposed by Newport (Newport, 1981a, 1981b, 1982, 1990) who argued that a high degree of internal structure in natural languages might arise through the process of language acquisition. Newport argued that linguistic systems become increasingly more consistent and structured as they are passed through several generations of learners. Unlike the cultural evolution accounts, Newport attributes this property to the process of first language acquisition (i.e., acquisition of native language by child learners). In particular, more limited cognitive abilities of child learners (such as memory limitations) may allow children to perceive and store only smaller components of complex stimuli. Thus, child learners may be in a better position to perceive the internal structure of linguistic stimuli (e.g., analyze words into morphemes) than adults, who are more readily able to remember entire complex stimuli and thus fail to analyze them into components. This hypothesis (known as *Less is More*) is supported by the analysis of errors made by child and adult learners: Typical late learner errors have to do with producing whole unanalyzed forms, while typical early learner errors involve errors of componential structure such as omitting required components (Newport, 1981a, 1981b, 1982, 1990).

1.7. Functional constraints

A third type of account has argued that pressures associated with language use and communication shape the historical development of languages over time. Such functional constraints can be seen as a specific type of cognitively-motivated constraints, discussed in the previous section. A recurrent idea in this tradition is that cross-linguistic properties of language structures represent efficient trade-offs between two competing motivations: forces of unification and diversification (Zipf, 1949), iconicity and economy (Aissen, 2003; Bates & MacWhinney, 1982; Croft, 2003; Givón, 1985; Hockett, 1958), clarity and ease (Slobin, 1977), efficiency and complexity (Hawkins, 2004), robust information transfer and production effort (Aylett & Turk, 2004; Genzel & Charniak, 2002; Jaeger, 2010; Levy, 2005).

The idea that languages have properties that make them particularly suited for communication was popularized by Zipf in the early 20th century (Zipf, 1932, 1949). He suggested in particular that human behaviors (including language) are organized according to a single principle of least effort that seeks to minimize the average work expenditure in the long-term (Zipf, 1949). According to Zipf, this principle can explain for example why cross-linguistically more frequent words tend to be shorter than less frequent ones: More frequent words have become shorter so that speakers spend on average less effort producing their utterances.

If speaker's effort expenditure was the only factor at play, however, linguistic forms would most likely undergo extreme reduction and might eventually all reduce to the least effortful sound (e.g., a schwa) or no sound at all. Thus, efficient communicative systems need to consider two competing pressures – the preferences of the speaker (or the *force of unification* in Zipf's terminology) and the preferences of the listener (or the *force of diversification*). From the speaker's point of view, prior to taking into consideration the speakers' goal to communicate, using one form for all possible meanings would be most economical since it would minimize the effort needed to acquire and maintain a large vocabulary. Using one form for all possible meanings, however, would maximize listener's effort since it would be impossible to determine the meaning intended by the speaker. Instead, a listener would prefer one-to-one form-meaning mappings since this would ensure most efficient meaning recovery (but see more recent work by Ferrer i

Cancho and Díaz-Guilera (2007) and Piantadosi, Tily, and Gibson (2011a) for a suggestion that some out-of-context ambiguity is expected in efficient information systems). Thus, an efficient linguistic system adopted by a community of speakers should strike a balance between these two pressures.

The ideas proposed by Zipf, which were developed prior to the advent of information theory (Shannon, 1948), were subsequently recast in information and probability theoretic terms. For example, recent work has provided evidence that word lengths are better predicted by the word's in-context predictability than frequency as originally assumed by Zipf (Manin, 2006; Piantadosi, Tily, & Gibson, 2011b), suggesting that this cross-linguistic property of lexicon strikes an efficient balance between the speaker's effort and successful message decoding (or the listener's preferences).

In later years, the focus in the functionalist tradition has shifted to the idea that language use and acquisition involve domain-general mechanisms that are shared between language and other cognitive systems such as memory, perception, motor planning, etc., that are also responsible for interactions with the non-linguistic world. Thus, linguistic forms that are maximally isomorphic to the underlying non-linguistic experience with the outside world should be more easily stored, retrieved or communicated using these shared domain-general resources (e.g., Givón, 1985). In other words, the structure of the linguistic forms should be iconic of the structure of some aspects of cognitive *non-linguistic* experience associated with this from. Broadly construed, iconicity is often invoked as an antonym to arbitrariness, and in some sense captures different non-arbitrary aspects of language including but not limited to the following:

- **meaning to form equivalence** (e.g., Croft, 2003; Slobin, 1977): if there is a meaning to express, it should be expressed by a non-zero form associated with it
- **complexity** (e.g., Aissen, 2003; Givón, 1991; Langacker, 2000): more complex meanings should be associated with more complex forms
- **conceptual distance** (e.g., Haiman, 1983): meanings that are conceptually closer should be expressed by more compact linguistic units

Indeed, iconicity correctly captures many cross-linguistic properties: meanings that are conceptually close like inalienable possession or direct causation receive more compact forms cross-linguistically than their conceptually more distant counterparts such as alienable possession or indirect causation (Haiman, 1983); more marked (i.e., more complex) categories such as object (vs. subject), plural (vs. singular), causative (vs. non-causative) are typically overtly expressed while less complex categories are often zero-marked. There are of course many aspects of language that are non-iconic. For example, zero-marking or case syncretism are abundant cross-linguistically, which clearly goes against the iconic motivation. Following Zipf (1949), these phenomena are attributed to reduction of frequent or predictable material due to a competing motivation of speakers' economy (Aissen, 2003; Bybee & Hopper, 2001; Haiman, 1983).

Functional accounts of this type are intuitively appealing since they recognize a striking parallel between typological universals and universals of human cognition and communication. Claims about functional motivations are, however, generally made at a very abstract level and typically do not go beyond establishing an intuitive link to psychology. Furthermore, the notions of complexity, conceptual distance, etc., are not precisely defined in functional literature, which has made it hard to derive empirically testable predictions based on these theories and ultimately has resulted in the lack of popularity of these accounts outside of linguistics.

One attempt to bridge the gap between the study of typological universals and research in psychology is Hawkins' performance-grammar correspondence hypothesis (Hawkins, 1994, 2004, 2007) that examines cross-linguistic patterns in grammar with reference to incremental processing (see Bever, 1970; Nichols, 1986; Slobin, 1973 for earlier processing accounts of typological universals). On this view, grammars are not just accessed in performance, but performance preferences actively influence what types of structures are going to be conventionalized as part of grammar. In other words, the cross-linguistic distribution of grammars is expected to mirror gradient processing preferences of an individual speaker, where grammatical structures that are processed more easily are preferred.

The source of processing difficulty is attributed primarily to memory resources: During incremental sentence processing words are integrated one by one into the structural representation, which requires retrieval of previous material from memory. A large body of research in psycholinguistics has shown that processing difficulty correlates with the number of dependencies (words that depend on each other for interpretation) that need to be retrieved since longer dependencies presumably tax the memory capacity (Fedorenko, Gibson, & Rohde, 2006; Gibson, 1998, 2000; Grodner & Gibson, 2005).

In Hawkins's account, grammatical structures that permit more efficient integration of words into the structural representation should be preferred cross-linguistically. Efficiency is achieved through three broad principles that powerfully capture a variety of typological phenomena:

- **Minimize Domains:** This principle links efficiency in grammars to complexity and argues that word orders that minimize dependency lengths should be preferred cross-linguistically. Consistent headedness and cross-linguistic asymmetry in heavy constituent ordering depending on language headedness can be plausibly accounted for by principle. This principle is also in line with recent work suggesting that dependency length plays a role in language change (Tily, 2010) and that at least languages that do not have a case system seem to have average dependency length close to the theoretical minimum (Gildea & Temperley, 2010).
- **Minimize Forms:** This principle relates efficiency to the selection of grammatical structures and lexical items with respect to their frequency and availability. Grammars should minimize production effort by reducing contextually predictable forms (cf. Zipf's principle of least effort and Haiman's economy principle) and heavily rely on contextual information to achieve successful communication. This principle correctly captures a

variety of reduction phenomena and such properties as avoidance of complete synonymy, vagueness, and zero specification.

- **Maximize Online Processing:** This principle ties efficiency to the speed of processing and argues that informative cues should be provided early during parsing to avoid delayed or incorrect recognition of a constituent. The cross-linguistic correlation between headedness and marking type in a language (head-final languages typically have dependent marking while head-initial languages tend to have head-marking) can be explained by this principle.

Further support for the performance grammar correspondence hypothesis comes from the observation that phenomena categorically required in some languages tend to correlate with speakers' gradient preferences in languages where the grammar allows choices (Bresnan, Dingare, & Manning, 2001). For instance, constituent orders in fixed constituent order languages typically correspond to preferred orders in flexible constituent order languages (Hawkins, 2004). Another example comes from the effects of animacy on word order. The animacy of the grammatical object, which is an obligatory determinant of word order in ditransitives in Sesotho (Morolong & Hyman, 1977) and Mayali (Evans, 1997), influences speakers' gradient preference between the two permissible orders in ditransitive alternations in English (Bresnan, Cueni, Nikitina, & Baayen, 2007).

Unlike many of the previously discussed functionalist accounts, this approach formally operationalizes complexity and efficiency and makes direct quantitative predictions about the structures that should be preferred in online production and cross-linguistically.

A further attempt to more tightly link the cross-linguistic distribution of grammars to current research in psycholinguistics is more closely related to the studies in this thesis. This approach is presented by recent work applying mathematical theories of communication (Shannon, 1948) to linguistic research. This work has argued in particular that language structures are at least partially shaped by a trade-off between speaker's production effort and communicative goals (Aylett & Turk, 2004; A. Frank & Jaeger, 2008; Genzel & Charniak, 2002; Jaeger, 2010; Kurumada & Jaeger, 2013; Levy & Jaeger, 2007): The speaker tries to achieve robust information transmission (i.e., successful interpretation of the intended message by the listener) on the one hand, while weighing this preference against the resources necessary to encode the intended message on the other. Crucially, since communication takes place in the presence of environmental and neural noise, the speaker's message may get corrupted during the transmission process and thus miscommunicated. Such miscommunications are more likely to occur when the message is less expected (i.e., more surprising or more informative) given the linguistic signal and non-linguistic cues such as world knowledge and context. Thus, a production system that strikes an efficient balance between the probability of communicative success and effort is expected to provide more redundancy in the linguistic signal for less expected messages compared to more expected messages.

Indeed, this inverse relation between expectedness and the amount or quality of the linguistic signal is widely attested during online production in the phonetic realization of words (Aylett & Turk, 2006; van Son & van Santen, 2005), contraction (A. Frank & Jaeger, 2008), case-marker omission (Kurumada & Jaeger, 2013, submitted), optional use of function words (Jaeger, 2010), and the choice of the number of clauses (Gomez Gallo, Jaeger, & Smyth, 2008).

Recent work has also shown that at least some cross-lexical and grammatical properties of languages are beneficial for efficient information transmission, though most of this work has been limited to the properties of the lexicon (Manin, 2006; Piantadosi et al., 2011a, 2011b) rather than the grammar (but see Maurits, Perfors, & Navarro, 2010; Qian & Jaeger, 2012).

Unlike many other functional accounts, information-theoretic approaches are based on prior principles (but see Ferrer i Cancho, Debowski, & Moscoso del Prado Martin, 2013 for critical discussion) and provide quantifiable predictions derived from mathematical theories of communication about what properties of grammar and lexicon should persist cross-linguistically if they were shaped by considerations of efficient communication. These accounts originate in current psycholinguistics research and directly link patterns observed in typology to speakers' online production. The mechanism that gives rise to preference for efficient information transmission is still not well understood. On some accounts, speakers implicitly learn to balance robust information transfer and effort through implicit feedback about previous communicative successes (Jaeger, 2013; Jaeger & Ferreira, 2013). I postpone a more detailed discussion of these and other accounts to Chapter 5, when I examine them in relation to the learning outcomes observed in this dissertation.

These approaches, however, leave open the question of how preferences for efficient information transmission enter the linguistic system and come to shape grammar over time. Two mutually compatible scenarios are possible (cf. Bates & MacWhinney, 1982; Jaeger, 2010). Communicative pressures throughout life can cause adult speakers to subtly change the input provided to the next generation. Alternatively, communicative pressures can operate during language acquisition, biasing learners to slightly deviate from the input they receive. Despite the long history of these claims within the functional literature, empirical tests of these two hypotheses are lacking.

1.8. Outline of this dissertation

In this dissertation, I focus on the cause of cross-linguistic universals and ask whether some typologically frequent phenomena can be explained by domain-general biases associated with considerations about human communication. I explore in particular whether language acquisition can provide a potential mechanism through which these biases come to shape language structures.

The experiments in this dissertation approach this question by studying the acquisition and production of case-marking and constituent order as well as the trade-off between these two means of conveying sentence meaning. I present five miniature artificial language learning experiments designed to address the question of whether learners are biased towards efficient linguistic systems (in an information-theoretic sense) even during the process of language acquisition (i.e., before a mature language system is in place). In all experiments, I present learners with input languages that are somewhat inefficient from the point of view of mathematical theories of communication (and deviate from naturally occurring types). I explore whether learners restructure these ‘inefficient’ input languages as they acquire them to increase their communicative efficiency and whether by doing so they introduce typologically frequent phenomena into the newly acquired grammars. These learning outcomes, if observed, would provide the first direct test of the hypothesis that functionally-motivated learning biases can create the seed for at least some language universals.

Throughout this dissertation, I employ an information-theoretic perspective that views efficient communication as a trade-off between robust information transfer, or the speaker’s goal to be understood, and a goal to conserve effort (an assumption shared with Zipf). In this framework, robust information transmission refers specifically to the amount of uncertainty about the intended sentence meaning that a listener would experience if they had a perfect knowledge of the grammar. Specifically, I focus on one aspect of sentence meaning, namely the assignment of grammatical functions (subject and object) or thematic roles (agent, theme; the two notions are not distinguished by the experiments reported in this dissertation).

For accessibility reasons, I make a simplifying assumption in framing these hypotheses and assume noise-free recognition of word sequences. However, the perspective assumed here is readily extendable to a plausible assumption of noisy acoustic input and noisy recognition (for the relevant work in comprehension, see Kleinschmidt & Jaeger, under review; Levy, 2011; Levy, Bicknell, Slattery, & Rayner, 2009; Norris & McQueen, 2008). The logic of the arguments presented in this dissertation also holds for this more plausible scenario. This means that with regard to all experiments presented here robust information transfer can also be conceptualized as ambiguity avoidance (e.g., Bolinger, 1972; Temperley, 2003). There are, however, several reasons why I adopt the information-theoretic perspective. First, it is based on prior principles and is broadly supported beyond cases of ambiguity (e.g., in phonetic and morphological reduction (Aylett & Turk, 2006; A. Frank & Jaeger, 2008; Jaeger, 2010; Kurumada & Jaeger, 2013, submitted; van Son & van Santen, 2005) and comprehension (Hale, 2001; Kleinschmidt & Jaeger, under review; Levy, 2011; Levy et al., 2009; Norris & McQueen, 2008). Second, research in psycholinguistics has not found conclusive evidence for ambiguity avoidance, in particular with regard to syntactic ambiguity, as studied here (Arnold, Wasow, Asudeh, & Alrenga, 2004; Ferreira & Dell, 2000). The observed effects of ambiguity avoidance are typically quite weak and often do not hold across different data sets for the same phenomenon (compare, for example, Jaeger, 2006; Jaeger, 2010; Roland, Elman, & Ferreira, 2006; Temperley, 2003; Wasow & Arnold, 2003). Most work on ambiguity avoidance has focused, however, on temporary ambiguity, not global ambiguity as in the experiments presented here. I postpone a more detailed discussion of how the experiments reported in this dissertation bear on the work on ambiguity avoidance to Chapter 5.

The remainder of this dissertation is structured as follows. In Chapters 2 and 3, I discuss how the well-described correlation between constituent order freedom and the presence of case-marking in a language can be derived from biases to trade off robust information transmission and effort during language acquisition. In Chapter 2, I test whether the amount of constituent order flexibility in a language can influence acquisition of case-marking. In Chapter 3, I explore the corollary of this prediction: whether the presence of case-marking in a language impacts constituent order regularization during language acquisition. Chapter 4 investigates whether the omission of morphological case in differential and optional case-marking systems can stem from pressures associated with efficient information transfer. Behavioral evidence presented in this chapter is especially timely in light of recent claims that patterns in case omission in differential case-marking systems are an artifact of areal diffusion and are thus unlikely to be functionally motivated (Bickel et al., in press). In Chapter 5, I summarize the major findings and discuss the conclusions that can be drawn from this work as well as outline potential avenues for further research.

Chapter 2

Balancing robust information transfer and effort: Differential regularization of case-marking depending on constituent order freedom

In Chapter 2, we are interested in the implicational universal that describes the correlation between constituent order freedom and the presence of case-marking (Blake, 2001; Sapir, 1921): Languages with flexible constituent order often use morphological means, such as case, to mark grammatical function assignment (e.g., German, Japanese, and Russian), whereas languages with fixed constituent order typically lack case morphology (e.g., English, Mandarin, and Italian). We argue that this typological correlation can be understood as a trade-off between effort and robust information transmission (cf., Jaeger, 2013; Piantadosi, Tily, & Gibson, 2011b; for further references see Jaeger and Tily, 2011). In languages with fixed (or relatively fixed) constituent order, the ordering of arguments is highly informative about their grammatical function (i.e., which is the subject and which is the object of the clause). Hence there is little uncertainty about the intended meaning after hearing a sentence in such a language (formally, the conditional entropy over grammatical function assignments), and case-marking would provide little additional information beyond that conveyed by constituent order. In languages with flexible constituent order, however, the ordering of arguments alone leaves comparatively high uncertainty about the intended grammatical function assignment. For such languages, case-marking is an informative cue to grammatical function. While in both types of languages the effort for the production of case is identical, the utility of case-marking is higher in flexible constituent order languages (for related arguments, see also Haspelmath, 1999; Hawkins, 2004; Tily, 2010). Thus, if communication is a sufficiently important function of language use and if robust information transfer is weighed against production or processing difficulty, cues to grammatical function will be expected to trade off at different levels of linguistic organization (e.g., syntax and morphology). Consider the aforementioned implicational universal regarding the correlation between constituent order flexibility and the presence of case-marking in a language (Blake, 2001; Sapir, 1921). This universal allows efficient linguistic systems and prohibits linguistic systems that either do not guarantee successful communication or have high levels of cue redundancy.

We use a miniature artificial language learning paradigm to ask in particular whether this preference originates during language acquisition. If biases towards efficient communication operate during acquisition, we would expect learners to restructure their input languages, making them more efficient as learning unfolds. The experiment reported below tests this prediction by investigating the trade-off between constituent order (a syntactic cue) and case-marking (a morphological cue). We further investigate *how* case is used in languages that have constituent order variation. We ask, in particular, whether learners increase communicative success by favoring robust information transfer over effort and regularize case-marking in the language overall, or whether they favor an efficient balance between these two goals, conditioning case-marking on constituent order.

Participants

Monolingual native speakers of English were recruited from the University of Rochester. Following our previous work (Fedzechkina et al., 2012), recruitment continued until the number of participants who successfully learned the miniature languages reached 20 in each condition. A total of 75 participants were recruited for the experiment, most of whom received $25 for their time (participants in the flexible constituent order condition received $30 for their participation since it was added later after some changes to the study protocol were implemented). 15 participants were excluded from the analysis for the following reasons: failure to achieve 65% accuracy on the comprehension test (13 participants); computer error (1 participant); being bilingual (1 participant). This left the data from 60 participants for the analysis.

2.1. Design and Materials

Each of the three miniature artificial languages contained 10 novel content words (4 verbs and 6 nouns) and a case-marker 'kah' (see Table 2.1). All words were phonotactically legal non-words of English. Individual words were synthesized using the AT&T speech synthesizer (voice 'Crystal') and concatenated into sentences with 35ms silence between the words using Praat (Boersma, 2001). This procedure ensured that the stimuli did not contain prosodic cues to sentence meaning. All sentences described short videos created using Poser Pro software that depicted simple transitive events such as 'hug' or 'poke' performed by two male actors such as 'chef' or 'referee' (see Figure 2.1 and Figure 2.2 for example stimuli).

Nouns	Verbs	Case-marker
glim flugit spad bliffen norg melnawg	geed kleidum zamper shen	kah

Table 2.1: The artificial lexicon.

All verbs occurred equally frequently within each language overall and with each constituent order allowed by the language. All nouns occurred equally often in the subject and object position with each verb.

Participants were pseudo-randomly assigned to one of the three language conditions. The three languages contained optional case-marking: 67% of all objects were marked with a case-maker 'kah' and 33% of objects had no overt marking. Subjects were never case-marked in any of the languages. All languages had head-final constituent order (i.e., the verb followed both the subject and the object). This constituent order was chosen since it is cross-linguistically more common in languages with a case system (Dryer & Haspelmath, 2011; Greenberg, 1963).

The languages differed in their constituent order consistency (and therefore in the amount of information that constituent order conveyed about sentence meaning). In the *random constituent order language,* subject-object-verb (SOV) and object-subject-verb (OSV) orders occurred equally frequently in the input. Thus, in this language, constituent order was uninformative about grammatical function assignment, and case-marking added important information to decode sentence meaning.

The *fixed constituent order language* did not contain constituent order variation: SOV constituent order occurred in 100% of the input sentences. In this language, constituent order was highly informative and always disambiguated grammatical function assignment; case-marking was a redundant cue to grammatical function.

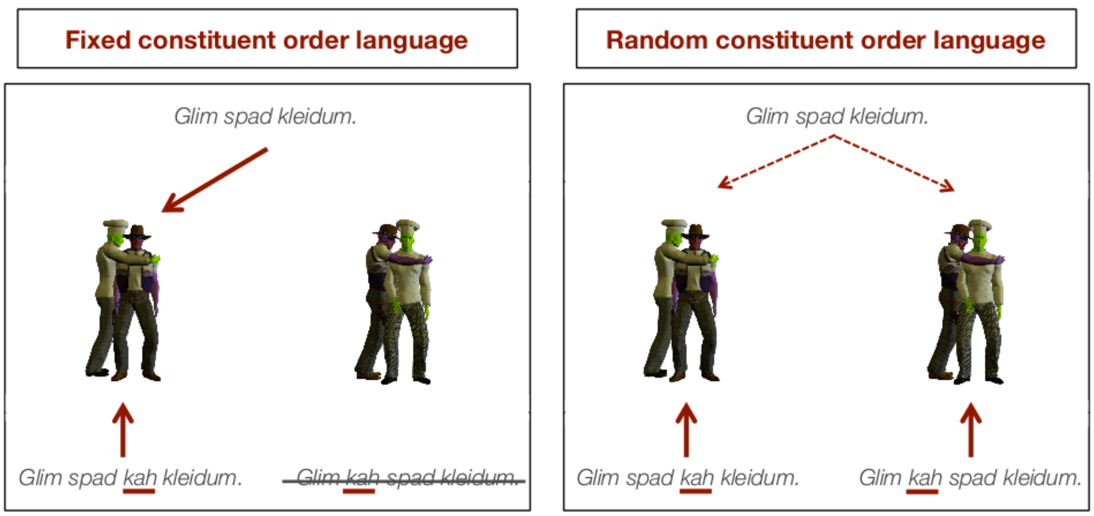

Figure 2.1: Illustration of form to grammatical function mappings in two extreme scenarios (fixed and random constituent order languages) in Experiment 1. Pictures are still images of sample videos with their English glosses (not shown to participants) and translations (not shown to participants). Arrows indicate form to meaning mappings in the two languages. Solid arrows indicate one-to-one form-meaning mappings such as in the absence of constituent order variation or in the presence of case-marking (underlined). Dashed arrows indicate one-to-many form-meaning mappings, such as in the absence of case-marking for variable constituent order. The strikethrough form refers to the ungrammatical (OSV) sentences in the fixed (SOV-only) language.

While these two extreme cases of constituent order freedom are theoretically possible cross-linguistically, they do not reflect the typologically common types.

Languages with completely fixed or completely flexible constituent orders are almost non-existent cross-linguistically. Even languages that have very rigid word order (e.g., English), typically allow some constituent order variability, while languages with extremely variable constituent order, like Russian or Latin, tend to have a dominant constituent order (i.e., constituent order used most frequently across contexts). To investigate whether our predictions are supported for typologically more common types as well, we included a cross-linguistically more plausible language in our experiment.

The *flexible constituent order language* contained some constituent order variation: SOV constituent order occurred in 75% of the input sentences and OSV constituent order occurred in 25% of the input sentences. This language thus represented the intermediate step between the two extreme cases of constituent order consistency outlined above.

If language acquisition is indeed biased towards efficient linguistic systems, we expect this to be reflected in the languages learners acquire. This makes three predictions. First and most importantly, we predicted that learners would use more case-marking when it is highly informative of grammatical function assignment (random constituent order language) than when it is a redundant cue (fixed constituent order language). Second, we expected learners to show at least a numerical gradient increase in case-marker use as the informativity of constituent order decreased across the three languages as shown in (1):

(1)

fixed constituent order < flexible constituent order < random constituent order

Finally, if learners strongly disprefer to spend effort when it is not required for successful communication, we might see that the redundant case-marking is frequently omitted by learners of the fixed constituent order language, compared to the input language.

2.2. Procedure

The procedure followed Hudson Kam and Newport (2005, 2009). At the beginning of the experiment, participants were informed that they were learning a novel 'alien' language by watching short videos and hearing their descriptions in this language, but they received no explicit instructions about language structure.

Participants completed 3 experimental sessions (each lasting 30-35min) spread over 3 days, with at most 1 day between the sessions. During each visit, participants completed a series of exposure and test blocks that focused on noun and sentence learning. All sessions followed the same overall procedure; the number of blocks, however, differed from session to session (see Figure 2.2).

Noun exposure and tests

<u>Noun exposure.</u> Each experimental session began with a *noun exposure block*, where participants were presented with pictures of characters accompanied by their names in the novel language and instructed to repeat the names to facilitate learning.

<u>Noun comprehension.</u> The 2-alternative forced choice *noun comprehension block* followed. Participants heard a name of a character in the novel language accompanied by

two pictures and were asked to choose the correct picture. Feedback was provided after each trial.

Noun production. Participants were shown pictures of characters one at a time and asked to provide a label for them in the novel language. Feedback on correctness was provided on each trial.

Noun exposure and comprehension blocks included 12 trials each on Day 1 and 6 trials each on Days 2-3. The noun production block included 6 trials on all days of training. On Day 1, the three blocks were repeated immediately after completion of the noun production test. The noun exposure and comprehension blocks were also presented immediately before the sentence production test. On Days 2-3, participants completed only the noun production block before the sentence production block.

Sentence exposure and tests

Sentence exposure. The *sentence exposure block* followed noun exposure and tests on all days of training. Participants were shown short computer-generated videos accompanied by their descriptions in the novel language and asked to repeat the descriptions out loud to facilitate learning. Participants were allowed to listen to the novel description as many times as they liked during the first sentence exposure block on Day 1; replay was disabled for all other blocks during the experiment.

Sentence comprehension. On all days of training, sentence exposure was followed by a *sentence comprehension block*. On each trial, participants were shown two previously unseen videos accompanied by a sentence in the novel language. Both videos depicted the same action performed by the same characters, but the order of the actor and patient was reversed in the two videos. Participants were asked to choose the video that best matched the sentence they heard. No feedback on correctness was provided during the test.

On all days of training, participants were presented with two sets of two sentence exposure blocks and one sentence comprehension block (24 trials each).

Sentence production. Each experimental session ended with a *sentence production block* (48 trials total). Participants viewed previously unseen videos and were asked to describe them in the novel language they learned during the experiment. To facilitate production, participants were auditorily presented with a novel verb prompt. No feedback was provided during this test.

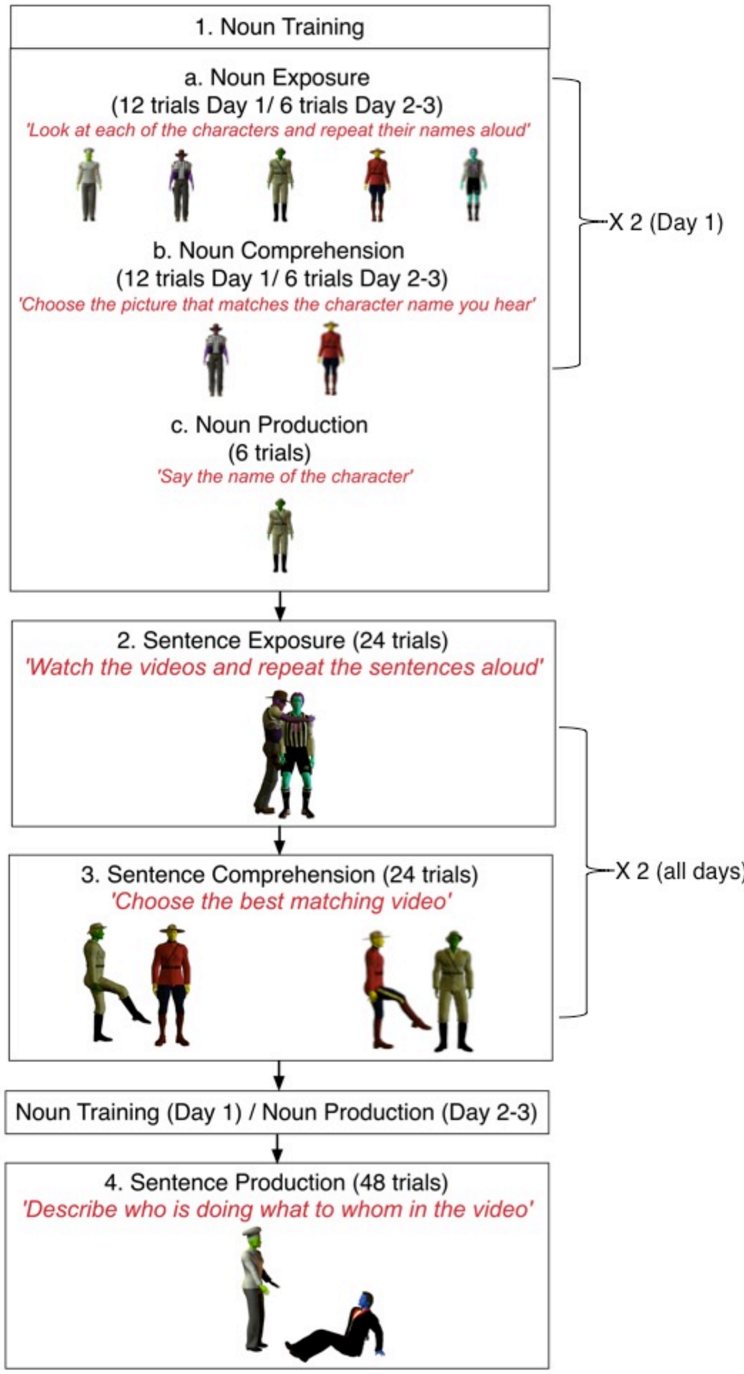

Figure 2.2: Experimental procedure. Pictures are still images of sample videos used in Experiment 1.

2.3. Results

2.3.1. Scoring

Comprehension accuracy

Participants who did not achieve 65% accuracy on the comprehension test on the final day of training were removed from the analysis. For this purpose we only analyzed responses on case-marked (i.e., unambiguous) trials. This excluded 13 participants, 10 in the random and 3 in the flexible order language. This is not surprising since participants were monolingual native speakers of English, a language that has no constituent order variation or case-marking. This makes languages that contain constituent order variation considerably harder to learn.

For the remaining 60 participants, mean comprehension accuracy was 96% across languages (99% for the fixed constituent order language, 93% for the flexible constituent order language, and 96% for the random constituent order language) on the final day of training. The results reported below did not change when participants who failed to pass the 65% criterion were included in the analysis.

Thus, even though we cannot know precisely how participants have interpreted the case-marker in our experiment (e.g., as a postposition following the object-noun or as a bound morpheme attached to the object-noun, etc.), the comprehension performance suggests that participants have learned the function of case-marking and were using case as a cue to grammatical function assignment.

Production accuracy

For each trial, we scored constituent order used in the utterance, the presence of case-marking on the object, lexical (using incorrect vocabulary) and grammatical mistakes (using a constituent order not allowed by the language or using a case-marker on a constituent other than the object). On a small number of utterances, participants mispronounced the name of a referent or an action. If it was impossible to determine the constituent order used in the utterance (e.g., when both referent names were mispronounced), the production was scored as both lexically and grammatically incorrect If it was still possible to determine the constituent order used in the utterance (e.g., the name of only one referent was incorrect), the production was coded as a lexical mistake and was scored for grammatical mistakes.

Both languages were acquired with a high degree of accuracy. On the final day of training, participants made 2.3% lexical mistakes across languages (1% in the fixed, 4.5% in the flexible, and 1.6% in the random constituent order language [$\chi^2(1)=1.02$, $p=0.31$]) and 0.8% grammatical mistakes (0.6% in the fixed, 0.4% in the flexible, and 1.3% in the random constituent order language), suggesting that the task was feasible for our participants. All analyses reported below are based on grammatically correct productions only. The same results were obtained when productions containing lexical mistakes were also removed from the analysis.

The data presented in this section suggest that learners in our experiment have acquired the miniature artificial languages very well. On the final day of training, participants made very few lexical and grammatical mistakes and learned the function of case-marking in the language. Thus, the deviations in participants' productions from the input distribution of case-marking (if observed) are unlikely to be due to arbitrary mistakes since learners appear to have acquired all other aspects of the miniature artificial languages very well.

2.3.2. Constituent order in production

Since learners in our experiment received no instruction as to which structures to use in their own productions, they could have made languages more efficient by changing the amount of constituent order freedom allowed in the languages (e.g., by making the random or flexible constituent order languages more fixed or the other way around) or by differentially using case-marking depending on the amount of constituent order flexibility. For example, learners of the non-fixed constituent order languages could have regularized the dominant SOV constituent order and thus reduced the uncertainty about grammatical function assignment to zero. If learners preferentially regularize constituent order in the non-fixed order languages, differential use of case-marking would not be expected in our experiment.

Learners did not vary constituent order properties of the input languages (see Figure 2.3). Learners of the random and fixed constituent order languages maintained the input constituent order proportions on all days of training. Participants in the flexible constituent order condition showed an initial preference to regularize SOV order on Days 1 and 2, possibly due to the English bias (Day 1: 86% SOV in production, significantly higher than the 75% input proportion [χ^2 (1)=5.12, $p<0.05$]; Day 2: 83% SOV in production, significantly higher than the input [$\chi^2(1)=4$, $p<0.05$]). However, learners of the flexible constituent order language gradually converged on the input by the final day of training (80.5% SOV in production, not significantly different from the input, $\chi^2(1)=2$, $p=0.1$; see Figure 2.3). We discuss the absence of constituent order regularization in more detain in Chapter 3.

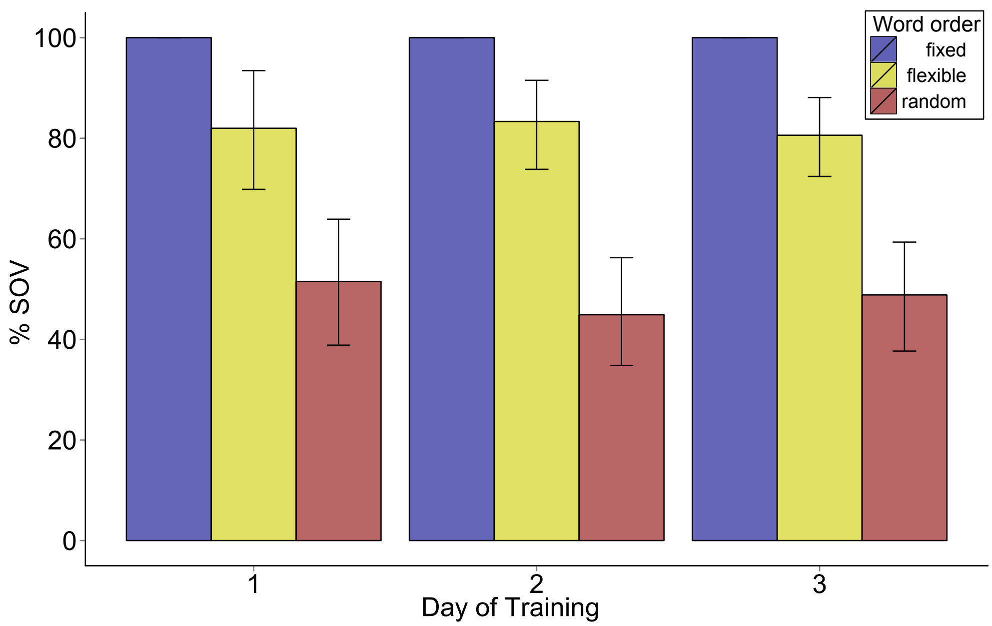

Figure 2.3: Constituent order use by language. The error bars represent bootstrapped 95% confidence intervals.

Although learners on average tended to match the input frequencies of constituent order variants, there was substantial between-participant variability in the use of the dominant constituent order in the non-fixed constituent order languages (see Figure 2.4). While the majority of participants tended to stay close to the input proportion, quite a few learners (especially in the flexible constituent order language) substantially deviated from the input either in the direction of SOV or OSV. There was no between-participant variability in the fixed constituent order language: All learners used exclusively SOV constituent order in their productions. This learning outcome is in line with prior work within the miniature language learning paradigm suggesting that learners acquire consistent input very well but are not likely to introduce innovations into a perfectly consistent linguistic system (e.g., Christiansen, 2000; St Clair et al., 2009; Tily et al., 2011).

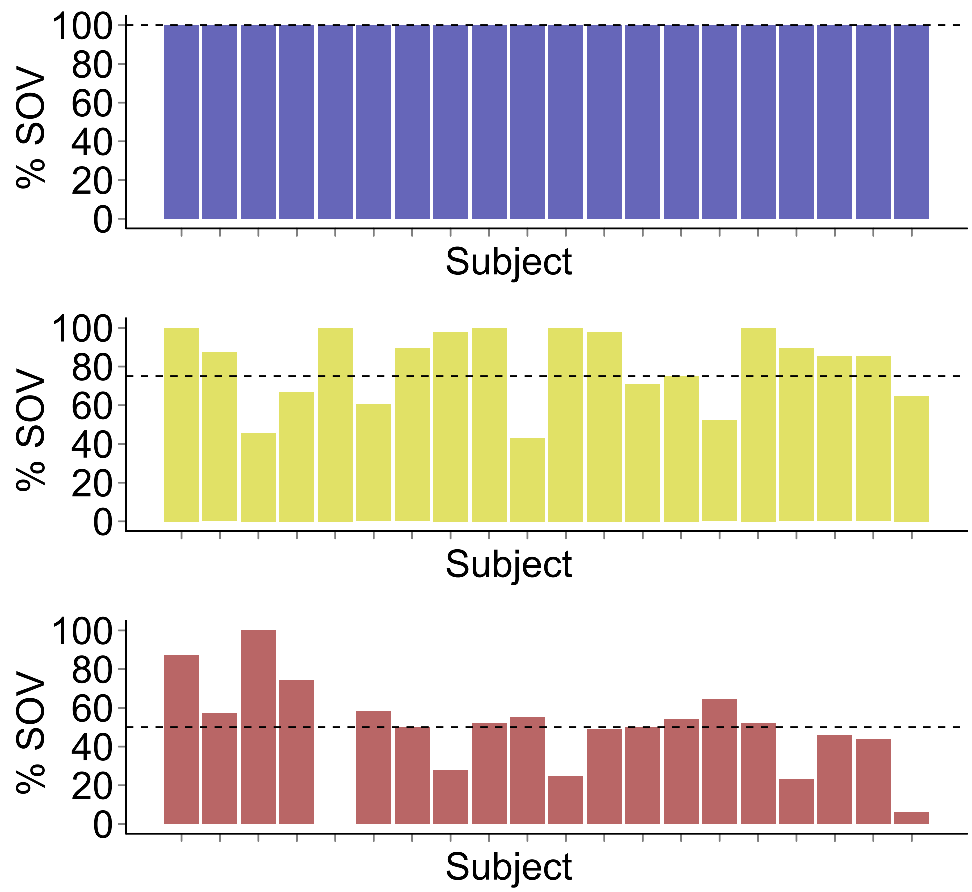

Figure 2.4: Individual preferences in constituent order use in the fixed (top panel), flexible (middle panel), and random (bottom panel) constituent order languages on the final (3rd) day of training. The dashed lines indicate the input proportion.

2.3.3. Case-marker use in production

The first and central prediction of the current study is that language learners are biased against excessive redundancy in linguistic systems and use additional cues to grammatical function only if the existing cues do not provide sufficient information for successful recovery of sentence meaning. That is, we expect learners of the random constituent order language to use more case-markers than learners of the fixed constituent order language. We also predict that learners should show gradient sensitivity to the informativity of constituent order and thus use case-marking proportionally to the amount of information provided by constituent order in the three languages.

To test these predictions, we used mixed logit regression (Jaeger, 2008) to regress the presence of case-marking on the object onto full factorial design (all main effects and interactions) of language condition (*fixed vs. flexible vs. random* constituent order) and day of training (*1-3*). All analyses reported below contained the maximal random effects structure justified by the data based on backward model comparison. All results also hold when the fullest converging random effects structure is used.

There was a significant difference in case-marker use between the two extreme language conditions (see Figure 2.5): Learners of the random constituent order language used significantly more case-marking in their productions than learners of the fixed constituent order language ($\hat{ß}$ =1.42, z=2.27, p<0.05). The random language condition interacted with Day 2 ($\hat{ß}$=0.32, z=1.7, p=0.09) and Day 3 of training ($\hat{ß}$=0.21, z=2.02, p<0.05). Simple effects test revealed that the difference between the two language conditions (*fixed* vs. *random* constituent order) was significant on Day 2 ($\hat{ß}$=1.53, z=2.6, p<0.01) and Day 3 ($\widehat{ß}$=1.83, z=2.8, p<0.01) of training. Thus, as predicted, learners of the random constituent order language were more likely to use case-marking.

There was no statistically significant difference in case-marker use between the flexible constituent order language and either of the two extreme cases (i.e., case-marker use did not differ between the fixed and flexible constituent order language ($\widehat{ß}$=0.94, z=1.5, p=0.13) or between the flexible and random constituent order languages ($\widehat{ß}$=0.48, z=0.77, p=0.44). There was, however, a significant linear trend in the data ($\widehat{ß}$=2.01, z=2.27, p<0.05): Learners showed a reliable increase in case-marker use as constituent order flexibility increased across languages (see Figure 2.5), suggesting that learners are indeed sensitive to the gradient changes in cue informativity across the three languages. This tendency was more pronounced on the final day of training, as suggested by its significant interaction with Day 3 of training ($\widehat{ß}$=0.3, z=2.02, p<0.05). There was no quadratic trend in the data ($\widehat{ß}$=-0.37, z=-0.43, p=0.67).

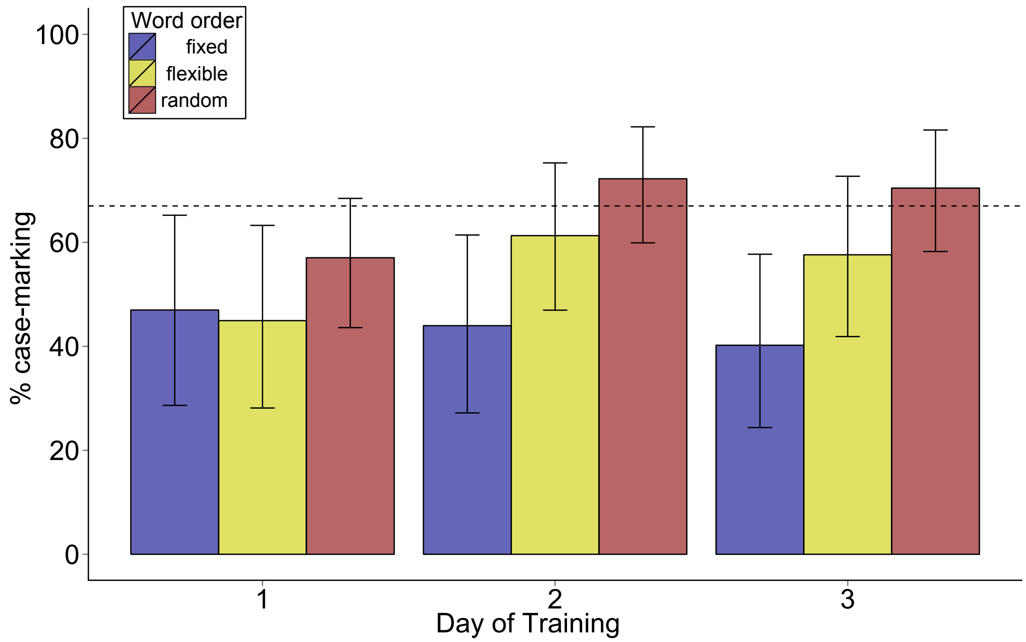

Figure 2.5: Case-marker use by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across languages).

Second, we also predicted that learners of the fixed constituent order language would reduce the amount of case-marking compared to the input since the production of case consumes effort but adds no information above that already conveyed by constituent order. Indeed, learners of the fixed constituent order language used case-marking in their own productions significantly below the input on all days of training, supporting our hypothesis (Day 1: 50% case-marking in production, significantly lower than 67% input proportion [χ^2 (1)=23.51, p<0.001]; Day 2: 45% case-marking in production, significantly lower than the input [χ^2 (1)=40.6, p<0.001]; Day 3: 41% case-marking in production, significantly lower than the input [χ^2 (1)=61.87, p<0.001]). The analysis of individual learner preferences further supports our prediction: Out of 20 learners, only 4 participants produced more case-marking than the input proportion and 8 used no case-marking in all their own productions on the final day of training (see Figure 2.6).

Overall, there was considerable variation in case-marker preferences among individual participants in the three languages: While some learners regularized case-marking and used it in all their productions, roughly the same number of learners never used case in their productions; and many learners used case variably (see Figure 2.6).

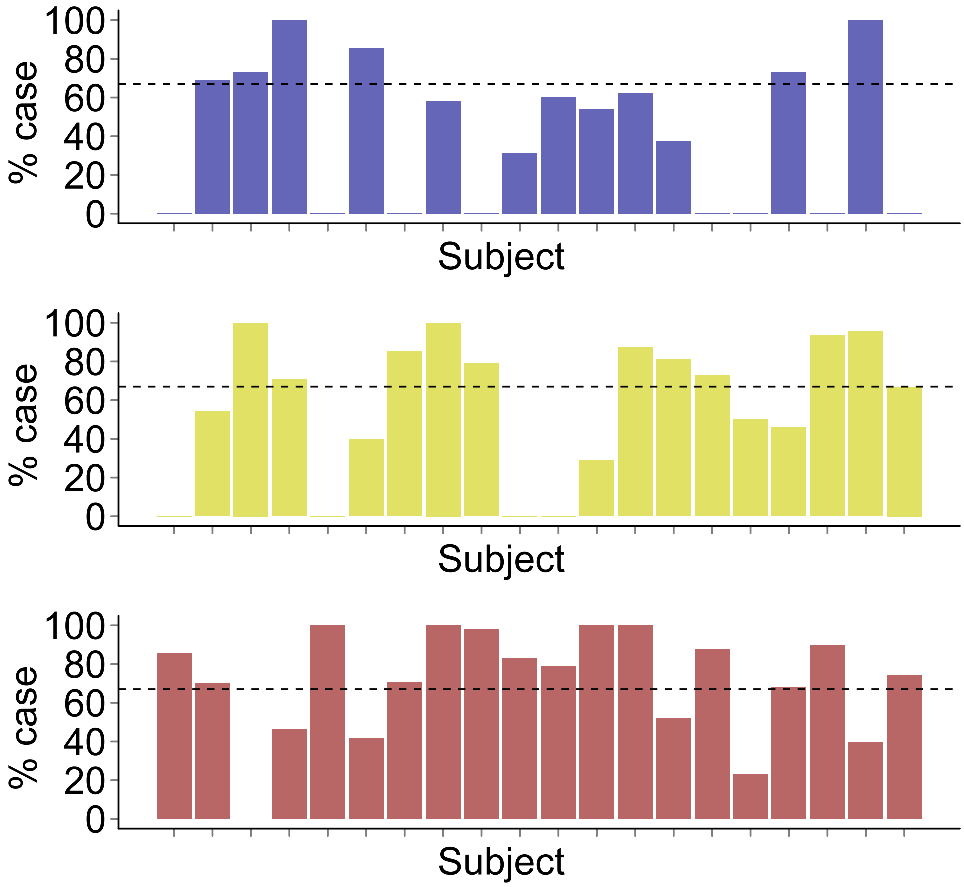

Figure 2.6: Individual preferences in case-marker use in the fixed (top panel), flexible (middle panel), and random (bottom panel) constituent order languages on the final (3rd) day of training. The dashed lines indicate the input proportion.

2.3.4. How is case-marking used in the non-fixed order languages?

With the three basic predictions confirmed, we next investigated in more detail *how* speakers of the non-fixed order languages (i.e., random and flexible order languages) used case-marking. There are at least two ways in which learners could increase the robustness of information transmission in a non-fixed order language. Let us consider a hypothetical language with 50-50% SOV-OSV constituent order and 50% case-marking that is independent of constituent order (i.e., a language that has maximal uncertainty about grammatical function assignment). In this language, there are two ways learners can achieve zero uncertainty about grammatical function assignment while keeping the same overall proportions of constituent order. First, they can regularize case-marking in the languages overall (i.e., produce it more frequently than in the input overall). This system would increase production effort compared to the input since all objects will need to be case-marked. Alternatively, learners could condition case-marking on constituent order and always use it with one variant but never with the other (i.e., have perfectly asymmetric case-marking). This would also result in zero uncertainty about grammatical function assignment since the absence of case-marking would be informative of grammatical function assignment. Crucially, this latter strategy is more efficient than the former since it avoids overtly marking all sentential objects, thereby resulting in a more robust system without a concomitant increase in effort compared to the input. More generally, asymmetric case-marking – even if not perfect (e.g., if one constituent order is more often case-marked than the other but case-marking is not perfectly correlated with constituent order) – is always at least as efficient as symmetric case-marking and often more efficient.

The same logic applies to our input languages. Below, we explore which of these strategies learners used in our experiment.

Do learners regularize case-marking in the languages overall?

As a first step, we investigated whether learners of the non-fixed order languages favored robust information transmission over effort and regularized case-marking in the languages overall. Learners of the non-fixed order languages did not increase the frequency of case-marking in their productions. Overall, case use by learners of the random constituent order language was the same as in the input on Days 2-3 (Day 1: 55% case-marking in production, significantly below 67% input proportion [χ^2 (1)=9.5, p<0.01]; Day 2: 72% case-marking in production, not significantly different from the input [χ^2 (1)=1.78, p=0.18]; Day 3: 71% case-marking in production, not significantly different from the input [χ^2 (1)=0.89, p=0.3]). Learners of the flexible constituent order language produced significantly fewer case-markers than in the input on the first day of training (44% case-marking, significantly lower than the input [χ^2 (1)=38.9, p<0.001]) and on the final day of training (58% case-marking, significantly lower than the input [χ^2 (1)=6.41, p<0.05]), but they did not deviate from the input on Day 2 of training (61.2% case-marking, [χ^2 (1)=1.93, p=0.16]). This behavior suggests that learners of neither flexible nor random constituent order languages regularized case-marking in the respective language overall and thus did not reduce uncertainty about the intended meaning by choosing a strategy that favors robust information transfer over production effort.

Do learners condition case-marking on constituent order?

Next, we explore whether learners trade off robust information transmission and effort efficiently by conditioning case-marking on constituent order. As a first step, we regressed the presence of case-marking onto constituent order (*SOV/OSV*), day of training (*1-3*), language condition (*random vs. flexible constituent order*) and their interactions. Learners of both languages did not use case-marking uniformly across the two orders, but instead produced case significantly more often in OSV sentences ($\widehat{\beta}$=1.63, z=5.78, $p<0.001$; see Figure 2.7). This preference did not interact with day of training (Day 2 x Constituent order interaction: [$\widehat{\beta}$=0.11, z=9.5, p=0.34]; Day 3 x Constituent order interaction: [$\widehat{\beta}$=-0.07, z=-1.2, p=0.22]) or with language condition ($\widehat{\beta}$=0.004, z=0.02, p=0.98), suggesting that the preference to use more case-marking in OSV sentences was equally strong on all days of training across the two languages. There was a significant Constituent order x Day 2 x Flexible constituent order language interaction ($\widehat{\beta}$=0.61, z=5.2, $p<0.001$), suggesting that the preference to use more case-marking in OSV sentences was stronger on Day 2 compared to Day 1 in the flexible order condition.

Thus, learners of both non-fixed order languages did not use case-marking uniformly across the two constituent orders, but instead introduced asymmetrical case-marking into the acquired languages. This strategy efficiently trades off robust information transmission and effort.

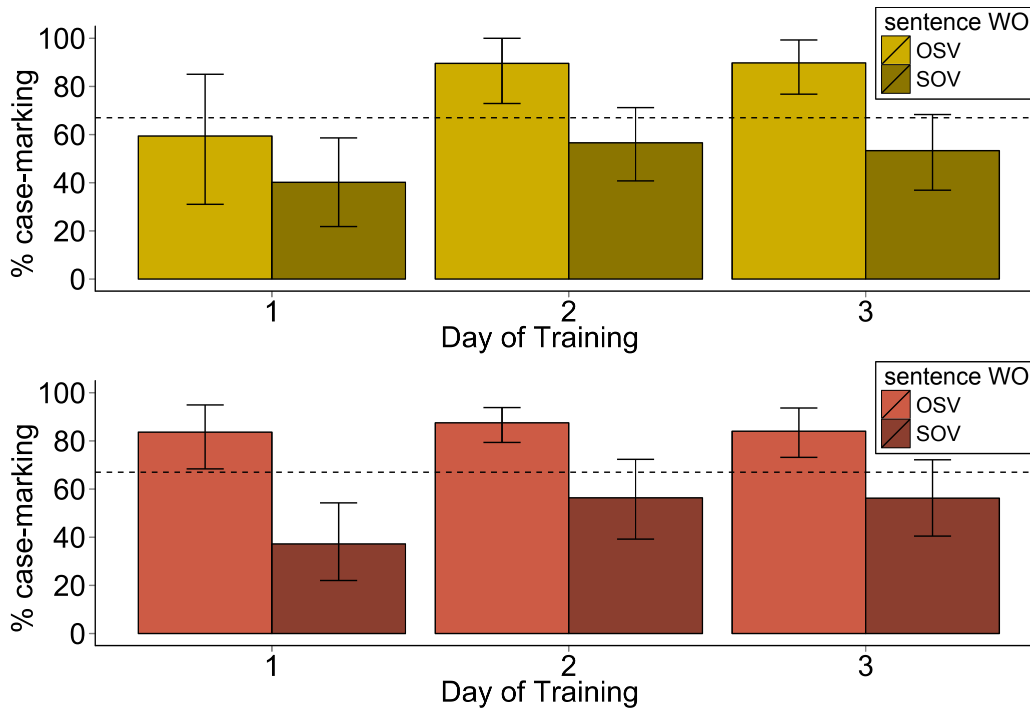

Figure 2.7: Case-marker use by sentence constituent order in flexible (top panel) and random (bottom panel) constituent order languages. The error bars represent bootstrapped 95% confidence intervals. The dashed lines represent the input proportion (equal across constituent orders).

What can account for learners' preference to use more case-marking in OSV sentences? One possibility is that this behavior could be indicative of a bias to overtly mark the atypical since OSV constituent order is typologically rare and is uncommon in English and thus could have attracted a higher proportion of case-marking. Alternatively, the preference to use more case-marking in OSV sentences could be due to pressures associated with incremental processing (Hawkins, 1994, 2004; Nichols, 1986). Since our input languages have object case-marking only, case would provide the earliest point of disambiguation in OSV sentences and thus allow comprehenders to commit to the intended parse early on in the sentence. We will further discuss this finding in Chapters 4 and 5.

2.3.5. Conditional entropy of grammatical function assignment

While the above analyses provide evidence in support of our predictions, they raise several questions. First, learners in the non-fixed order languages do not converge on a categorical system where OSV constituent order is always case-marked and SOV constituent order is never case-marked. Thus, it is unclear *how much* uncertainty reduction is actually achieved by conditioning case-marking on constituent order in our experiment.

Second, significant individual variability in Figures 2.4 and 2.6 highlights the fact that there are several different ways in which learners can restructure the input languages. None of the analyses reported so far captures this fully. For example, some learners can introduce asymmetric case-marking, but choose to use more case with SOV constituent order instead of OSV. Or, to give another example, some participant might fix constituent order. These strategies would not be readily apparent in the above analyses. Despite the high degree of variability between individual learners, it is, however, also possible that all learners follow the same general principle of balancing effort and the goal to be understood.

Next, we introduce a new approach to this question that lets us directly assess whether learners strive for such a balance, independent of the means by which they achieve it (i.e., independent of whether they, for example, fix constituent order, fix case-marking, or condition case-marking on constituent order). For this purpose, we estimated production effort and the amount of uncertainty about grammatical function assignment for each language acquired by individual participants. Production effort can be formalized in a variety of ways, but in practice these measure are typically correlated (Szmrecsanyi, 2004). We chose to formalize effort as the average number of syllables per sentence. As shown in Figure 2.8, the three input languages in our experiment have the same amount of effort (ranging from 4.5 to 5.5 syllables per sentence on average) since they use the same artificial lexicon and have the same amount of case-marking in the input.

The average uncertainty about grammatical function assignment experienced by the listener who has a perfect knowledge of the grammar used by the speaker was captured as weighted conditional entropy over grammatical function assignments:

$$H\ (GF|sent.form) = p(sent.form) * -\sum_{GFs} \sum_{sent.forms} p(sent.form, GF) * log_2\, p(GF|sent.form)$$

where the sum is over grammatical function assignments (subject-object, object-subject) and three possible sentence forms in the languages (NP1-no case NP2-case, NP1-case NP2-no case, NP1-no case NP2-no case) weighted by the probability of each of the three sentence forms in participant's production.

In our languages, the conditional entropy of grammatical function assignment is 0 bits for sentences with case-marked objects or for all sentences if there is no constituent order variation in the language. Thus, the fixed constituent order language had the lowest weighted conditional entropy in the experiment – that of 0 bits since this language contained no constituent order variability. The random constituent order (input) language had the highest weighted conditional entropy. 67% of the sentences were case-marked and thus resulted in a conditional entropy of 0 bits. 33% of the sentences were not case-marked and thus resulted in a conditional entropy of 1 bit since constituent order was 50% SOV and 50% OSV independent of case-marking. Thus, the average weighted conditional entropy of the random constituent order language was: 1*0.33+0*0.67=0.33 bits. The weighted conditional entropy of grammatical function assignment in the flexible constituent order language was 0.26 bits. 67% percent of all sentences resulted in conditional entropy of 0 bits since they were case-marked. The non-case marked sentences had conditional entropy of 0.8 bits since constituent order was somewhat informative in this language (i.e., 0.8*0.33+0*.67=0.26 bits).

As expected under our hypothesis, learners systematically deviated from the input towards languages that traded off effort and uncertainty about sentence meaning more efficiently (see Figure 2.8). As predicted, by dropping redundant case-marking in their productions, learners of the fixed constituent order language converged on a language that had lower average effort compared to the input. In line with our hypothesis, the output languages produced by learners in the non-fixed constituent order conditions tended to have lower weighted conditional entropy of grammatical function assignment compared to the input without a concomitant increase in effort. In fact, learners of the flexible constituent order language tended to lower effort compared to the input. This behavior highlights the benefit of asymmetric case-marking: For the same overall effort, entropy reduction is larger for systems that have asymmetric case-marking compared to systems with symmetric case-marking (i.e., case-marking independent of constituent order) for the same overall constituent order proportion (see Figure 2.8)

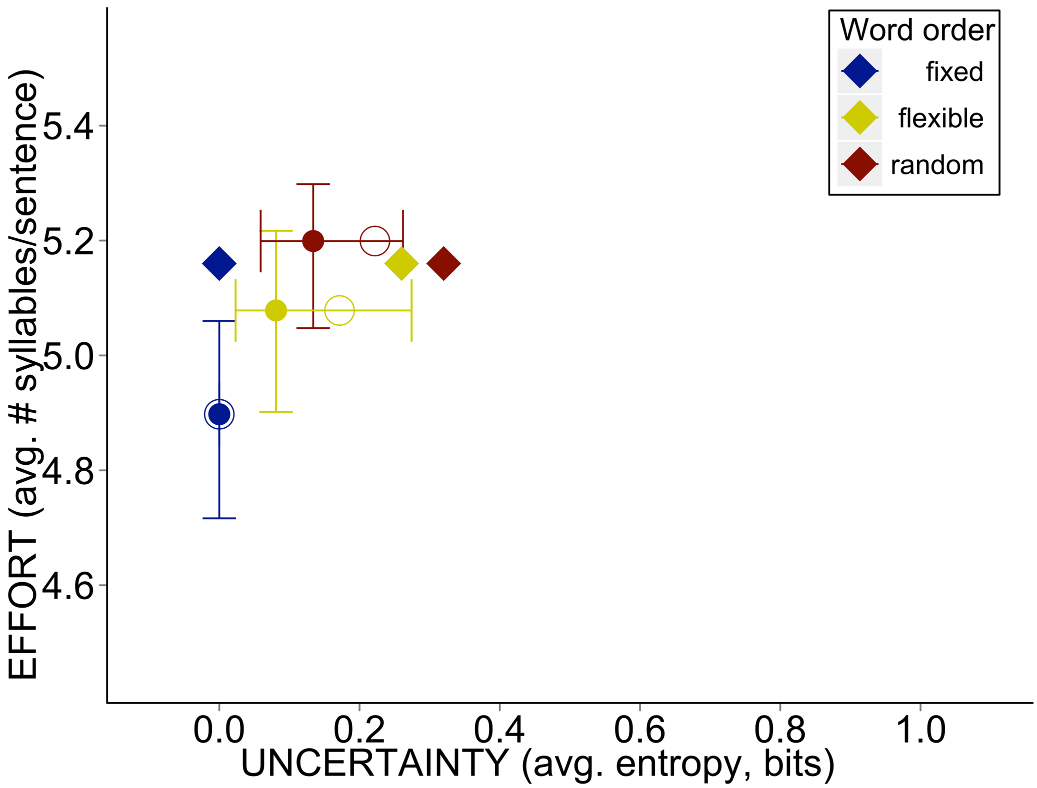

Figure 2.8: Uncertainty vs. effort trade-off on the final (3rd) day of training. Diamonds represent input languages. Solid circles represent mean output languages. Empty circles represent hypothetical output languages that use the same overall proportions of constituent order and case-marking as the actual output, but do not condition case-marking on constituent order. The error bars represent bootstrapped 95% confidence intervals.

Analysis of individual learners' preferences

The analysis of individual participants' behavior further supports our hypothesis. First, as Figure 2.9 shows, the majority of learners of the fixed constituent order language followed our prediction and reduced effort without increasing uncertainty. All learners of this language produced languages that had zero conditional entropy of grammatical function assignment, and 14 (out of 20) learners reduced effort compared to the input.

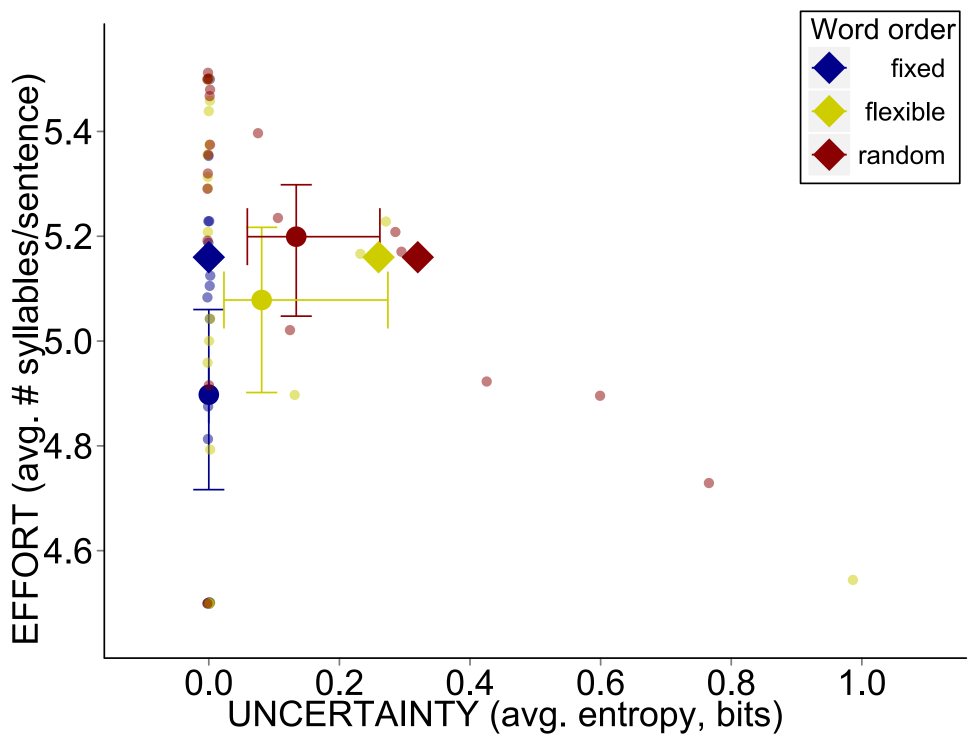

Figure 2.9: Uncertainty vs. effort trade-off on the final (3rd) day of training. Diamonds represent input languages. Circles represent mean output languages. The error bars represent bootstrapped 95% confidence intervals. Small semi-transparent dots represent languages produced by individual learners.

In the non-fixed constituent order languages, there was overall a strong bias to reduce uncertainty about grammatical function assignment: 28 out of 40 participants produced languages that had zero conditional entropy of grammatical function assignment (see Figure 2.9) and 6 out of the remaining 12 learners produced languages with lower uncertainty than in the input. In other words, 70% of participants restructured the input languages to have consistent rule-based grammars, which is highly atypical for adult learners' productions (Hudson Kam & Newport, 2005, 2009). Figure 2.10 illustrates the distribution of different strategies used by learners of non-fixed constituent order languages to achieve the same goal of reducing uncertainty about grammatical function assignment. While the majority of participants who reduced uncertainty did so by introducing asymmetric case-marking, some people (especially in the flexible constituent order language) chose to fix constituent and yet other learners regularized case-marking in the language overall. Only 5 participants (2 in the flexible and 3 in the random constituent order language) increased conditional entropy of grammatical function assignment. However, most of them did so while decreasing effort. Thus, 59 out of 60 participants in Experiment 1 followed the principle of trading off robust information and effort.

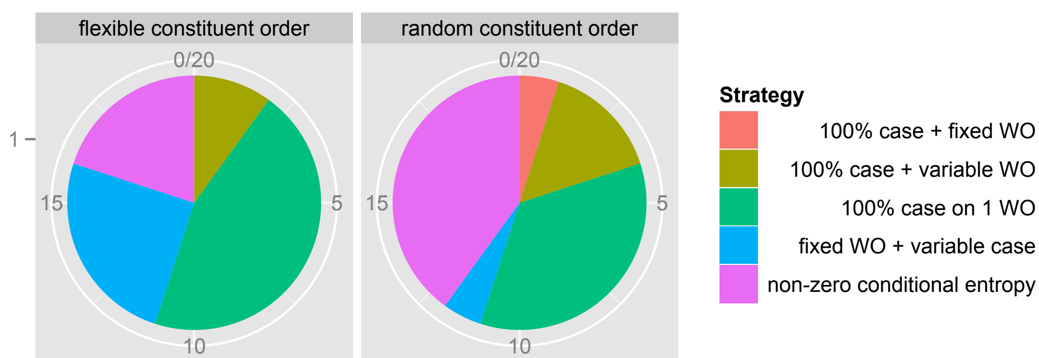

Figure 2.10: Entropy reduction strategies chosen by individual participants in the flexible (left) and random (right) constituent order languages. The sections represent the number of participants who adopted each of the strategies.

The overall picture emerging from individual learners' performance suggests that they indeed have a preference to balance robust information transmission and effort. Overall, the preference to reduce uncertainty in grammatical function assignment was stronger than a preference to reduce effort: In all languages, effort increases were acceptable as long as conditional entropy of grammatical function assignment was reduced. However, if uncertainty about grammatical function assignment was increased, there was a clear preference to reduce effort. These learning outcomes suggest that despite the fact that learners employ vastly differing strategies, learning behavior in our experiment is guided by a deeper abstract principle of a trade-off between robust information transmission and effort.

Our findings provide an insight into linguistic diversity: For any grammatical system, the general principle of trading off robust information transmission and effort can manifest in a variety of innovations created by learners. Which of them would survive at the population level and eventually become part of the grammar depends on a variety of factors, including the inherent efficiency of the innovation, other properties of the linguistic system, historical factors such as language contact (cf. Heine, 2008), and social factors such as social influences of a person creating the innovation (cf. Nettle, 1999b).

2.4. Discussion

Our findings add to the growing body of research showing that learners preferentially acquire and regularize typologically frequent patterns (Christiansen, 2000; Culbertson et al., 2012; Finley & Badecker, 2008; Morgan et al., 1987; Newport & Aslin, 2004). Learning outcomes in our experiment parallel synchronic and diachronic cross-linguistic phenomena, thereby providing converging evidence that these patters are not due to chance.

In our experiment, learners of the non-fixed constituent order languages chose to condition case-marking on constituent order rather than to regularize case-marking in the language to 100%. This result corroborates prior work showing that adult learners prefer linguistic systems that contain conditioned variability over systems with no variability (Smith & Wonnacott, 2010; but see Hudson Kam & Newport, 2005, 2009 for the opposite findings in children) and points to learners' preference for another cross-linguistically frequent property of human language – a tendency to have predictable variation by conditioning the use of competing forms on semantic, pragmatic, phonological, and other factors (Givon, 1985; Labov, 1963).

Conditioned variability in our experiment is also likely to increase processing speed since learners preferentially restructure input languages in such a way as to put more informative cues earlier in the sentence. These results also bear on the role of cue order during language acquisition and use. The recurrent finding in the developmental literature is that young learners are highly sensitive to the order of cues to sentence meaning and disproportionally rely on early arriving cues when making parsing decisions (Choi & Trueswell, 2010; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999). The same general preference has been found in natural language production (Hawkins, 1994, 2004) and typological distributions (Hawkins, 1994, 2004; Nichols, 1986). This behavior might reflect a preference to maximize the number of linguistic dependencies processed at each point in time (cf. 'Maximize On-line Processing' principle (Hawkins, 1994, 2004). Even though the two constituent orders were case-marked equally often in the input, learners produced significantly more case-markers in OSV sentences, which provides an earlier disambiguation point. These results are in line with Hawkins' Maximize Online Processing principle and tentatively suggest that processing biases influence the development of languages over time.

Learning outcomes in our experiment also parallel diachronic typological patterns, for example, the change from Old English (a language with flexible constituent order and a rich case system) to Modern English (a language with fixed constituent order and no case system). Whether the historical loss of case-marking precipitated (Marchand, 1951; Sapir, 1921) or followed constituent order fixing (Lehnert, 1957) is a matter of debate, and our current results are unable to tease apart these two alternatives. Under our hypothesis, however, both processes should result in the same outcome. Learners of a flexible order language should maintain constituent order flexibility if this language has a case system (cf. Old English) and gradually lose it if this language has no case system (cf. Modern English). We address this possibility in Chapter 3.

Chapter 3

Balancing robust information transfer and effort: Differential regularization of constituent order depending on the presence of case-marking

In Chapter 2, we found that learners exposed to miniature artificial languages differentially acquired optional case-marking depending on the amount of constituent order freedom in the language. We observed, in particular, that case-marker use in learners' productions gradually increased with the increasing amount of constituent order flexibility (i.e., with decreasing informativity of constituent order) in a language.

These learning outcomes parallel a variety of natural typological phenomena: the presence of case systems in languages with flexible constituent order (e.g., Russian or Latin) and their absence in languages with fixed constituent order (e.g., English or French) and patterns of diachronic change such as that from Old English (a language with flexible constituent order and rich case-marking) to Modern English (a language with fixed constituent order and only rudimentary case-marking).

Several hypotheses have been proposed to explain the patterns in historical change from Old English to Modern English. Some have argued that constituent order fixing was a result of case-marker loss (Marchand, 1951; Sapir, 1921), while others have suggested that case-marking became a redundant cue and was lost after English constituent order became fixed for independent reasons (Lehnert, 1957). Under our hypothesis, both processes should yield the same outcome.

In Chapter 2, we found support for the second view – we observed differential case-marker regularization depending on the amount of constituent order flexibility in a language. We did not, however, find much support for the first view: While on average, participants did not regularize constituent order in the non-fixed order languages, some individual learners regularized constituent order (see Figure 2.4). These learning outcomes suggest a possibility that constituent order is less susceptible to change than case-marking (at least in our experimental setup). In this chapter, we further explored the cross-linguistic correlation between case and constituent order flexibility. We probed in particular whether learners of a flexible order language maintain constituent order flexibility if this language has a case system (cf. Old English) and gradually lose it if other cues to sentence meaning (specifically case) are unavailable in the language (cf. Modern English).

3.1. Experiment 2

In Experiment 2, we exposed two different groups of learners to two miniature artificial languages that had flexible constituent order but differed in whether case-marking was consistently present or absent in a language. If learners indeed have a bias to trade off effort and robust information transmission, we expect them to be biased against excessive uncertainty about grammatical function assignment and predict that learners of the no-case language would be more likely to regularize constituent order, making it a more informative cue to grammatical function assignment.

3.1.1. Participants

21 monolingual native speakers of English were recruited from the University of Rochester to participate in Experiment 2. Recruitment continued until the number of participants who successfully learned the miniature languages reached 20 in each condition (1 participant was excluded from the study for failing to acquire basic verb-final order by the final day of training). All participants received $25 for their participation.

3.1.2. Design and Materials

Participants in Experiment 2 were pseudo-randomly assigned to one of the two miniature artificial languages. Both languages contained the same amount of constituent order variation – subject-object-verb (SOV) was the dominant constituent order and occurred in 67% of input sentences; object-subject-verb (OSV) was the minority constituent order and occurred in 33% of input sentences. There were no verb-specific restrictions on constituent order: All verbs occurred equally frequently in the input overall and with each constituent order variant in each of the two languages. The input languages differed in the presence of case-marking on the object. In the *no-case language* objects were never overtly case-marked (i.e., case-marking was absent in all sentences). In the *case language*, objects were always overtly case-marked. Subjects were never overtly marked in either of the two languages.

Thus, in the case language, grammatical function assignment was disambiguated by case-marking that was always present in the input. In the no-case language, however, there was a lot of uncertainty about grammatical function assignment since constituent order alone did not provide enough information to successfully decode sentence meaning. For example, in the no-case language, the two scenes shown in Figure 3.1 could be described as either *Flugit glim daf* (SOV) or *Glim flugit daf* (OSV). In the case language, however, the patient was always unambiguously identified by the accusative case-marker '*kah*', as in *Flugit glim kah daf* (SOV) or *Glim kah flugit daf* (OSV).

A crucial difference between the design of Experiment 1 presented in Chapter 2 and the experiments in the current chapter is in the case-marker distribution. In Chapter 2, we presented learners with variable constituent order languages that had optional case-marking. We found that even though uncertainty in grammatical function assignment could be reduced by introducing changes into either constituent order or case-marking, learners preferentially chose to change case-marking. In this chapter, we present learners with languages that have flexible constituent order and highly consistent case-marking that is either always present or always absent in the language. Thus, in the no-case language, constituent order is the only cue to sentence meaning. Learners are very unlikely to spontaneously introduce new forms (such as case or new constituent orders) within the short training regimes of studies like ours.[1] Thus, the design of Experiment 2 is more likely to reveal learners' biases towards changes in constituent order as a result of a bias a balance between robust information transfer and effort.

Figure 3.1: Still images of sample videos used in Experiments 2 and 3.

We predicted that if language learners are indeed biased to trade off cues to sentence meaning, then in the absence of case-marking, they should be more likely to regularize constituent order in the language overall (i.e., reduce uncertainty about grammatical function assignment associated with constituent order, making it a more informative cue to sentence meaning). In contrast, the presence of (highly informative) case-marking in a language should limit constituent order regularization. We, thus, expected learners of the case language to acquire greater constituent order variation.

[1] See for instance, Experiment 1 in Chapter 2, where learners almost never introduce new constituent orders into the fixed order input language.

The artificial lexicon and video stimuli for this experiment were adopted from Experiment 1 (see Table 2.1 and Figure 2.1 in Chapter 2). This experiment also followed the procedure used in Experiment 1 (see Figure 2.2 in Chapter 2).

3.1.3. Experiment 2: Results

3.1.3.1. Scoring

The overall scoring procedure described in Chapter 2 was followed in this experiment with one exception. Since both constituent orders were acceptable in comprehension in the no-case language, we scored participants' deviations from the input proportion of the dominant SOV constituent order (see Figure 3.2).

In the case language, constituent order in participants' comprehension responses closely mirrored the input on all days of training (Day 1: 66% SOV, not significantly different from the 67% input proportion [χ^2 (1)=0.09, p=0.75]; Day 2: 67% SOV [χ^2 (1)=0.01, p=0.9]; Day 3: 66% SOV [χ^2 (1)=0.02, p=0.88]), suggesting that learners acquired the meaning of case-marking and relied on it as a cue to sentence meaning.

In the absence of case-marking, there was a slight tendency to regularize the dominant constituent order in comprehension (Day 1: 67% SOV, not significantly different from the 67% input proportion [χ^2 (1)=0.01, p=0.94]; Day 2: 75% SOV, significantly higher than the input [χ^2 (1)=4.11, $p<0.05$]; Day 3: 74% SOV, marginally higher than the input [χ^2 (1)=3.39, p=0.07]).

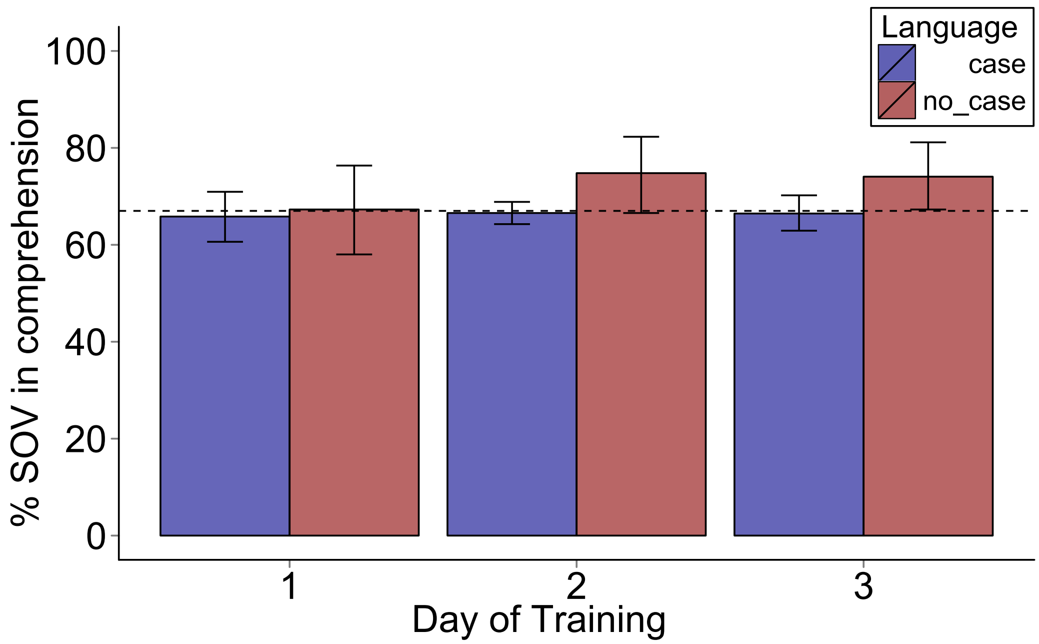

Figure 3.2: Constituent order use in comprehension by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across languages).

Production accuracy

Participants acquired the miniature languages very accurately, providing evidence that the task was feasible. On the final day of training, the proportion of lexical mistakes (such as mispronouncing the name of a character or verb) was 1.4% in the case language and 2.6% in the no-case language. Grammatical mistakes (i.e., using constituent order variants not allowed in the languages or incorrectly using the case-marker in the case language) were also rare: 0.4% (all in the case language). All analyses reported below are based on grammatically correct productions only.

3.1.3.2. Constituent order in production

The main prediction of the current experiment is that if language learners are biased against excessive uncertainty in grammatical function assignment, learners of the no-case language should regularize constituent order (i.e., use the dominant constituent order variant more often in their own productions than in the input). Since case-marking is a highly informative cue to grammatical function assignment, we expected it to limit constituent order regularization in the case language. That is, we predicted that learners of the no-case language should use the dominant SOV constituent order significantly more frequently than learners of the case language.

We tested this prediction using a mixed logit model to predict the proportion of SOV constituent order in participants' productions based on language condition (*case* vs. *no-case* language) and day of training (*1, 2, 3*) as well as the interactions between these two factors. The model included the maximal still converging random effects structure, which included by-subject and by-item random intercepts as well as by-subject and by-item random slopes for day.

There was no significant main effect of language condition ($\widehat{ß} = 0.16$, z=0.68, p=0.49) and no significant interactions between language condition and Day 2 ($\widehat{ß} = 0.25$, z=1.64, p=0.1) or Day 3 of training ($\widehat{ß} = 0.09$, z=1.12, p=0.26). Thus, counter to our prediction, learners did not differentially regularize constituent order depending on the presence of case-marking in the language on any day of training. There was, however, a numerical trend in the expected direction on Days 2 and 3 (see Figure 3.3).

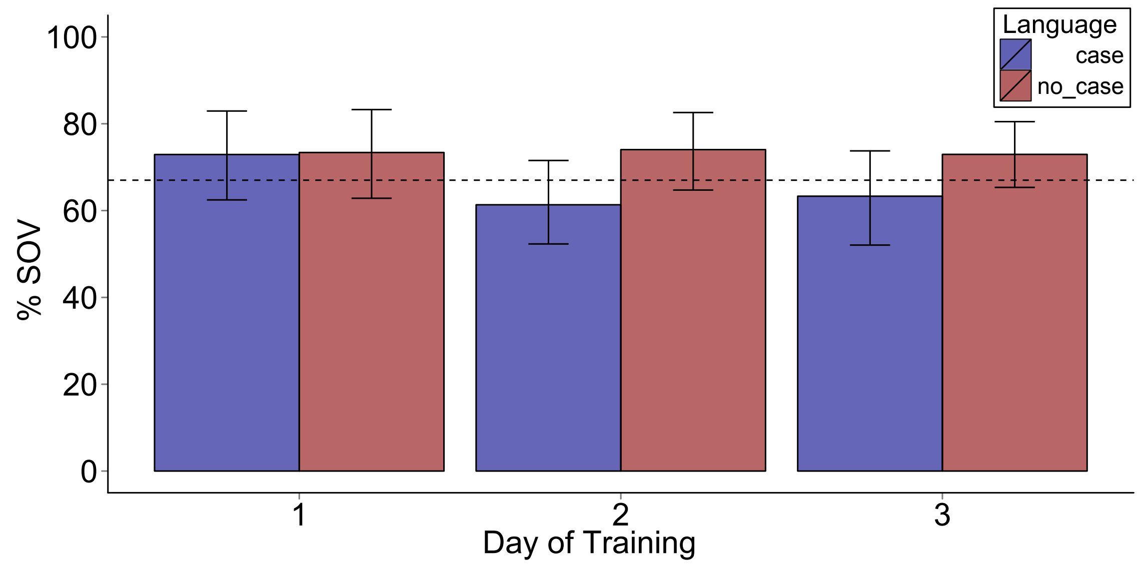

Figure 3.3: Constituent order use in production by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across languages).

The learning outcomes in this experiment did not follow our prediction: Learners of the no-case language did not show a preference to regularize the dominant constituent order more strongly than learners of the case language. One thing to keep in mind is that our participants do not receive explicit instructions as to which grammatical structures to use in their own productions, and there is typically considerable variation in the strategies individual learners use in experiments of this type (cf. Figure 2.4 and Figure 2.6 in Chapter 2). The above analysis, however, might obscure some of the possible strategies participants may use to reduce uncertainty in grammatical function assignment. First, while it is more likely that learners would regularize the dominant SOV order, some learners might regularize the minority (OSV) constituent order. Second, the overall constituent order use that we scored in the analysis above does not speak to the predictability of grammatical function assignment: Within the overall proportions, systems with predictable and unpredictable grammatical function assignments are possible. For instance, in the input 67-33% SOV-OSV language, the use of constituent order variants was not conditioned by individual verbs and grammatical function assignment was probabilistically inconsistent. One can imagine another language, however, with the same overall constituent order proportions, where SOV order is used consistently with 2 out of 4 verbs, OSV constituent order is always used with 1 out of 4 verbs, and SOV and OSV vary for 1 out of 4 verbs. In this system, the overall proportions of constituent order variants will remain the same, but grammatical function assignment will become more predictable if one takes into account verb-specific constituent order preferences. The strategy of conditioning constituent order on the verb in the newly acquired language is quite plausible given the learners' preference for conditioned variability observed in Chapter 2.

To address the possibility that learners in our experiment conditioned their constituent order preferences on the verbs, we conducted a follow-up analysis. For each participant, we calculated conditional entropy of grammatical function assignment based on verb-specific constituent order preferences for the two languages in Experiment 2:

(2)

$$H\ (GF|verb) = -\sum_{GFs}\sum_{verbs} p(verb, GF) * log_2\,(GF|verb)$$

where we sum over both possible grammatical function assignments (subject-object, object-subject) and all four verbs present in the language. This analysis does not take into account the information carried by case-marking in the case language and thus allows us to directly compare the uncertainty in grammatical function assignment carried by constituent order alone in the two languages.

The conditional entropy of grammatical function assignment is lower the more biased the verb-specific constituent order preferences are. For example, a verb contributes 0 bits to the condition entropy if it always occurs with the same constituent order. A verb contributes 1 bit if it occurs half of the time with one and half of the time with the other constituent order.

If language learners are indeed biased against excessive uncertainty in grammatical function assignment, then we would expect learners of the no-case language to use constituent order more consistently making it a more informative cue to sentence meaning (i.e., to have lower conditional entropy of grammatical function assignment) than learners of the case language since constituent order is a redundant cue to sentence meaning in this language. Learners' preferences, however, did not support our hypothesis: ANOVA analysis revealed no significant effect of language on conditional entropy of grammatical function assignment across the three days of the experiment ($F(1,117)=1.93$, $p=0.17$). As in the above analysis of overall constituent order preferences, there was a numerical trend in the expected direction: At least numerically, learners of the no-case language showed a stronger preference to reduce uncertainty in grammatical function assignment associated with constituent order (see Figure 3.4).

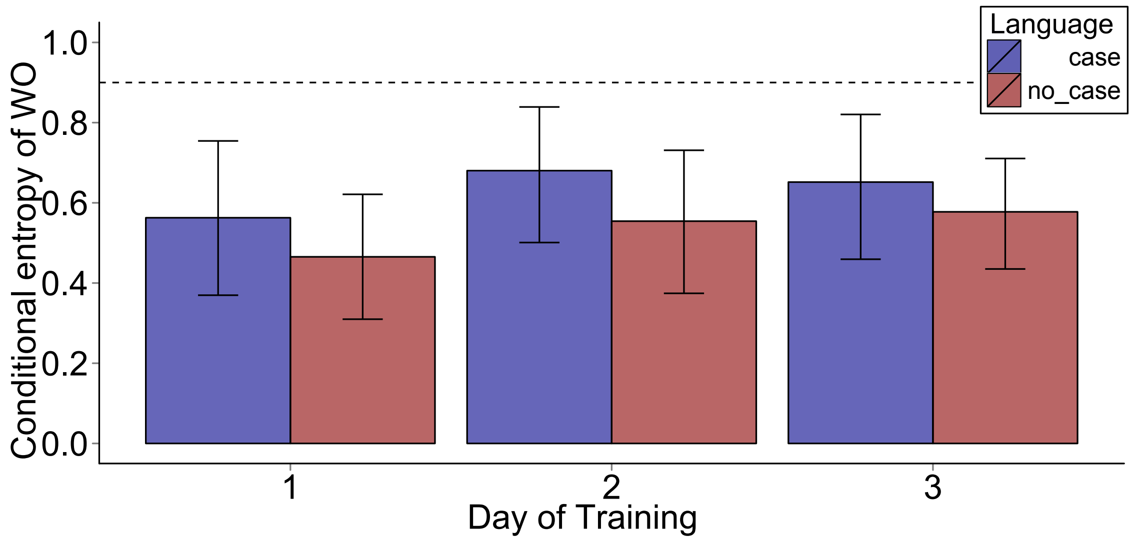

Figure 3.4: Conditional entropy of grammatical function assignment by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents conditional entropy of grammatical function assignment in the input (equal across languages).

3.1.4. Experiment 2: Discussion

In Experiment 2, we further investigated the cross-linguistic correlation between constituent order freedom and the presence of case-marking in a language. If this correlation is indeed due to the trade-off between robust information transmission and effort, then we expect learners to maintain constituent order flexibility if a language has a case system (cf. Old English) and to gradually lose it if a language has no case system (cf. Modern English). The learning outcomes in Experiment 2, however, did not support this hypothesis: Learners did not show differential constituent order regularization depending on the presence of case-marking in the language.

This result is surprising for several reasons. First, both we (see Chapter 2 and 4) and others (Culbertson & Newport, 2012) have found case-marker regularization using the miniature artificial language learning paradigm, which provides evidence that this method is powerful enough to detect learners' deviations from the input as a result of a functionally-motivated bias. Second, Culbertson et al. (2012), who used a miniature language learning paradigm similar to ours, observed word order regularization for 2-word phrases (numeral-noun and adjective-noun combinations) that mirrored the typological distribution. This suggests that our failure to detect constituent order regularization is unlikely to be due to the limitations of the artificial language learning *paradigm* used in the experiment.

The null effect observed in Experiment 2 is more likely due to the design of the *input languages* used in the experiment. Adult learners are strongly biased to match the statistics of the miniature artificial grammar in experiments of this type. This finding is widely supported both in linguistic (Hudson Kam & Newport, 2005, 2009) and non-linguistic (Ferdinand, Thompson, Kirby, & Smith, 2013) domains. Research in miniature artificial language learning has provided evidence, however, that this bias against deviating from the input can be overcome in input languages that are fairly complex (e.g., languages in which learners need to keep track of multiple frequencies simultaneously), while being still learnable within a short period of time. It is possible that our input languages were not complex enough and thus was acquired veridically. This could have potentially obscured a bias to trade off robust information transfer and effort. If this was indeed the case, then we should observe the expected differential constituent order regularization depending on the presence of case for more complex languages, where learners are required to track conditional constituent order distributions. This prediction was tested in Experiment 3.

3.2. Experiment 3

In Experiment 3, we exposed learners to two miniature artificial languages that were similar to the languages used in Experiment 2 in that they had flexible constituent order but differed in the presence of case-marking. Unlike Experiment 2, where constituent order variants occurred equally frequently with each of the verbs, in this experiment constituent order was conditioned on a verb class. Based on previous work (Ferdinand et al., 2013; Hudson Kam & Newport, 2009), we hypothesized that a more complex and irregular system would limit the learners' default preference to match the input statistics and would reveal their biases stemming from considerations for efficient information transmission.

3.2.1. Participants

41 monolingual native speakers of English were recruited from the University of Rochester to participate in Experiment 3. The procedure for recruitment termination was the same as in Experiment 2 (1 participant was excluded from the study due to a computer error). All participants received $40 for their participation.

3.2.2. Design and Materials

The two miniature artificial languages used in Experiment 3 contained 10 novel verbs and 6 novel nouns and (in the case language only) a case-marker ‘kah’ (see Table 3.1). Individual words were synthesized using the AT&T speech synthesizer (voice ‘Crystal’) and concatenated into sentences with 35ms silence between the words using Praat (Boersma, 2001) as in previous experiments.

Nouns	Verbs	Case-marker
glim	geed	kah
flugit	kleidum	
zub	zamper	
bliffen	shen	
norg	daf	
melnawg	jentif	
	blerfee	
	prog	
	mawg	
	slergin	

Table 3.1: The artificial lexicon used in Experiment 3

Constituent order was flexible in both languages: 65% of input sentences had SOV constituent order and 35% of input sentences had OSV constituent order. Within these overall percentages, a system of verb-specific biases was introduced into the language to increase the complexity of the system. The 10 verbs in the language were divided into four classes as shown in Table 3.2.

Number of verbs associated with verb bias groups			
OSV-only group (0% SOV)	Equibiased group (50 % SOV)	SOV-biased group (83% SOV)	SOV-only group (100% SOV)
2 verbs	2 verbs	3 verbs	3 verbs

Table 3.2: Verb classes used in the languages in Experiment 3.

3.2.3. Procedure

Participants were trained and tested on one of the languages over four consecutive days. The same procedure was followed on days 1-3, and a slightly different procedure was adopted for the final day of training.

Days 1-2

Noun Training. Participants viewed static pictures of the characters and heard their names in the novel language. The initial exposure was followed by a series of short tests where participants were asked to choose the character whose name they heard from a group of characters and to name the character shown on the screen. Feedback on performance was provided after each trial.

Sentence exposure. Participants viewed 60 short videos depicting transitive actions, one at a time, and heard an accompanying sentence describing the event in the novel language. They were instructed to repeat each sentence aloud to facilitate learning. On Day 1, participants could replay the first 12 scenes as many times as they wished to familiarize themselves with the language; no repetitions were allowed at any other time.

Noun Test. Participants viewed static pictures of each of the characters once and named them in the novel language. They received feedback from the experimenter on their performance.

Comprehension Test. In each trial, participants heard a novel sentence in the language and were shown two scenes in which the actor and patient were reversed. They were asked to choose the scene that matched the sentence. All scenes (60 total) contained novel combinations of familiar nouns and verbs.

Production Test. Participants were shown a transitive scene and were instructed to describe it in the language learned during the experiment, using the auditorily provided verb prompt. All scenes (60 total) contained novel combinations of familiar nouns and verbs, different from the ones used in the comprehension test.

Day 4

On the final day of training the comprehension and production tests included 12 additional scenes depicting familiar characters performing previously unseen actions. These scenes were introduced to assess participants' constituent order generalization to novel items. To avoid possible effects of fast implicit learning of verb-specific biases for these novel verbs through exposure during the comprehension test, the order of the tests was reversed: Participants performed the sentence production test first and then were presented with the comprehension test. Apart from these changes the procedure remained the same.

3.2.4. Experiment 2: Results

We predicted that in the absence of case-marking, learners would be more likely to make constituent order more informative of grammatical function assignment by regularizing it.

In contrast, if case-marking is available and language users use it as a cue to sentence meaning, it should limit constituent order regularization and learners should acquire greater constituent order variation.

Given the design of our languages, learners of the no-case language have two ways to increase the informativity of constituent order: They can regularize constituent order in the language overall (i.e., across verb classes) or they can take advantage of verb-specific information and regularize constituent order within a singular verb or a verb class.

3.2.4.1. Scoring

The scoring procedure used in Experiment 2 was adopted for this experiment. Since both constituent orders were acceptable in comprehension in the no-case language, we scored learners' deviations from verb-specific biases. As Figure 3.5 shows, on the final day of training participants' constituent order preferences in comprehension closely corresponded to constituent order distribution for each of the four verb classes in the case language (OSV-only verbs: 3.8% SOV in comprehension, not significantly different from the 0% input proportion [χ^2 (1)=0.17, p=0.67]; Equibiased verbs: 51.8% SOV [χ^2 (1)=0.08, p=0.8]; SOV-biased verbs: 84% SOV [χ^2 (1)=0.03, p=0.87]; SOV-only verbs: 97% SOV [χ^2 (1)=0.17, p=0.68]). This preference suggests that learners acquired the meaning of case-marking and relied on in interpreting grammatical function assignment.

Learners of the no-case language consistently produced more SOV constituent order than the input for OSV-only and equibiased verbs, but matched the input for the verbs biased towards the dominant SOV order (OSV-only verbs: 19.2% SOV $\chi^2(1)=4.9$, $p<0.05$]; Equibiased verbs: 64% SOV [$\chi^2(1)=4.4$, $p<0.05$]; SOV-biased verbs: 92% [$\chi^2(1)=1.62$, $p=0.2$]; SOV-only verbs: 96% SOV [$\chi^2(1)=0.24$, $p=0.62$]).

These learning outcomes in comprehension are consistent with the pattern observed in Experiment 2, where learners of the no-case language systematically deviated towards the dominant SOV constituent order as well.

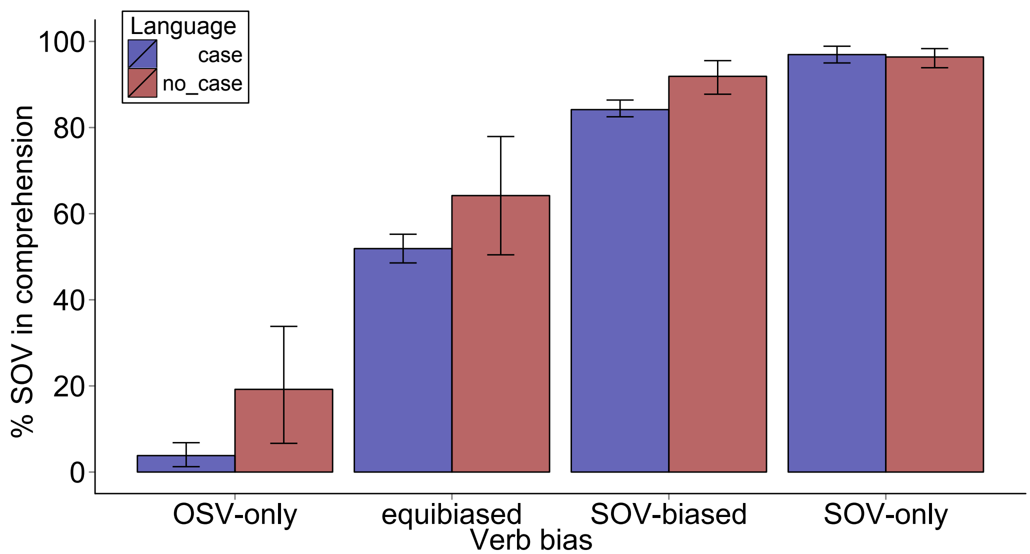

Figure 3.5: Constituent order use in comprehension by language on the final (4th) day of training. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across languages).

Production accuracy

On the final (4th) day of training, learners of both case and no-case languages made the same number (0.2%) of lexical mistakes. All grammatical mistakes on the final day of training (0.6%) were made by learners of the case language. All analyses reported below were conducted on production trials that were grammatically correct (94% of data across all days of training).

3.2.4.2. Constituent order regularization across verb classes

As a first step, we assessed the effect of case-marker presence on the extent of constituent order regularization in the language overall. All analyses presented below are conducted on previously unseen scenes containing familiar verbs; scenes containing novel verbs presented on Day 4 were analyzed separately.

We used a mixed logit model to predict the use of dominant SOV order in participants' productions based on language condition (*case* vs. *no-case* language), day of training (*1-4*), verb class (*OSV-only, equibiased, SOV-biased,* and *SOV-only* verbs) and the interactions between these three factors. The model included the maximal random effects structure that allowed the model to converge, which included by-subject and by-item random intercepts as well as by-item random slopes for day.

There was a significant difference in constituent order use depending on the verb class (see Figure 3.6): Across all days of training, learners' constituent order preferences reflected verb-specific biases in the input languages (learners used significantly more SOV constituent order for the equibiased verbs compared to the OSV-only group ($\widehat{ß} =$ 0.81, $z=12.7$, $p<0.001$), more SOV order for the SOV-biased group compared to the mean of the OSV-only and the equibiased verbs ($\widehat{ß} = 0.79$, $z=27.8$, $p<0.001$), and more SOV order for the SOV-only group compared to the mean of the other groups ($\widehat{ß} = 0.49$, z=21, p<0.001). Thus, learners have successfully acquired the additional complexity in the form of verb-specific biases we introduced into the input languages.

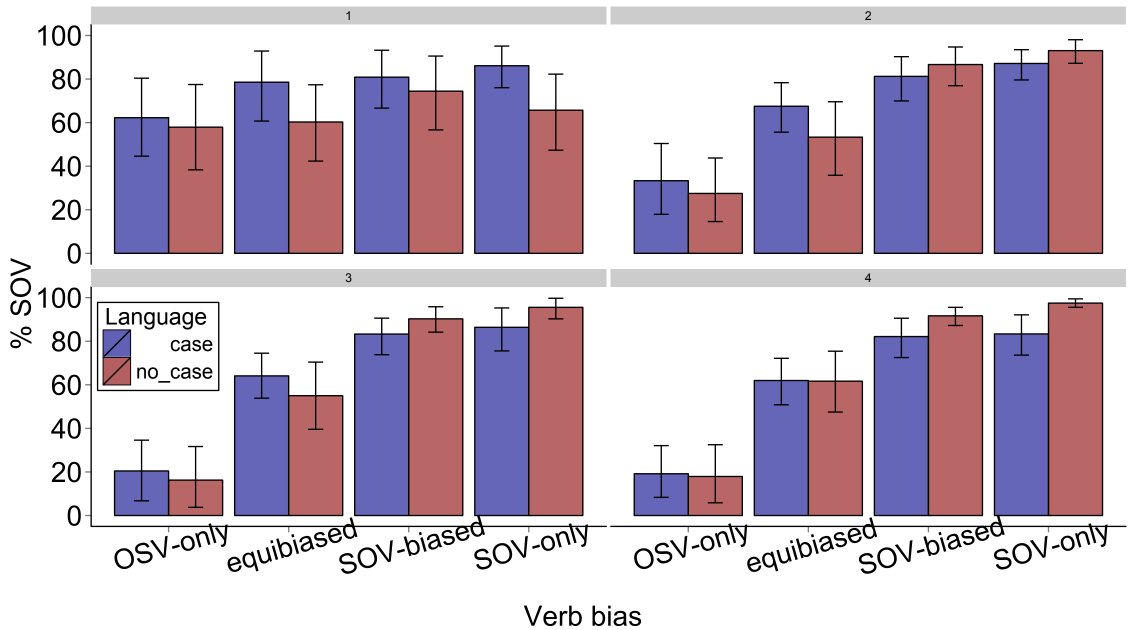

Figure 3.6: Constituent order production by verb class in the two languages. The error bars represent bootstrapped 95% confidence intervals. The numbers on the panels represent days of training.

There was no significant main effect of case-marker presence on constituent order regularization ($\widehat{ß}$ = 0.14, z=0.69, p=0.49), indicating that learners of the two languages did not differ in the overall amount of constituent order regularization across all days of training (see Figure 3.7). Language condition, however, interacted with Day 2 ($\widehat{ß}$=0.19, z=4.48, p<0.001), Day 3 of training ($\widehat{ß}$=0.09, z=3.46, p<0.001), and Day 4 of training ($\widehat{ß}$=0.11, z=5.36, p<0.001), suggesting an increasing tendency to use more dominant constituent order in the no-case language as the training continued. Simple effects test revealed a significant effect of language condition on the final day of training ($\widehat{ß}$=0.46, z=2.16, $p<0.05$): As expected, learners of the no-case language used the dominant constituent order significantly more often than the learners of the case language on the final day of training.

Language condition interacted with the SOV-biased verbs ($\hat{\beta}$=0.18, z=7.14, $p<0.001$) and with the SOV-only verbs ($\hat{\beta}$=0.14, z=7.73, $p<0.001$). Simple effects testing revealed that learners of the no-case language used the dominant constituent order significantly more frequently for the SOV-only verbs ($\hat{\beta}$=0.56, z=2.6, $p<0.05$) compared to the learners of the case language. This preference was more pronounced on the final day of training as suggested by a significant Language x Day x Verb bias interaction ($\hat{\beta}$=0.04, z=3.17, $p<0.001$). The difference between the two language conditions was marginally significant for the SOV-biased group ($\hat{\beta}$=0.36, z=1.7, p=0.08). Somewhat surprisingly, language condition did not interact with the equibiased verbs ($\hat{\beta}$=0.01, z=0.17, p=0.85), suggesting that constituent order use for this verb group, which has highest uncertainty in grammatical function assignment (if case information is not taken into account), did not differ between the case and no-case language.

Further analysis revealed that learners of the case language matched the input frequency of the dominant constituent order on the final day of training (χ^2 (1)=0.05, p=0.81), while learners of the no-case language used the dominant (SOV) constituent order significantly more frequently in their own productions (χ^2 (1)=5.1, $p<0.05$).

Thus, as expected under our hypothesis, learners of the no-case language showed a preference to make constituent order a more informative cue to sentence meaning and regularized it in the language overall on the final day of training, while learners of the case language, where grammatical function was always disambiguated by morphology, tended to maintain constituent order variation.

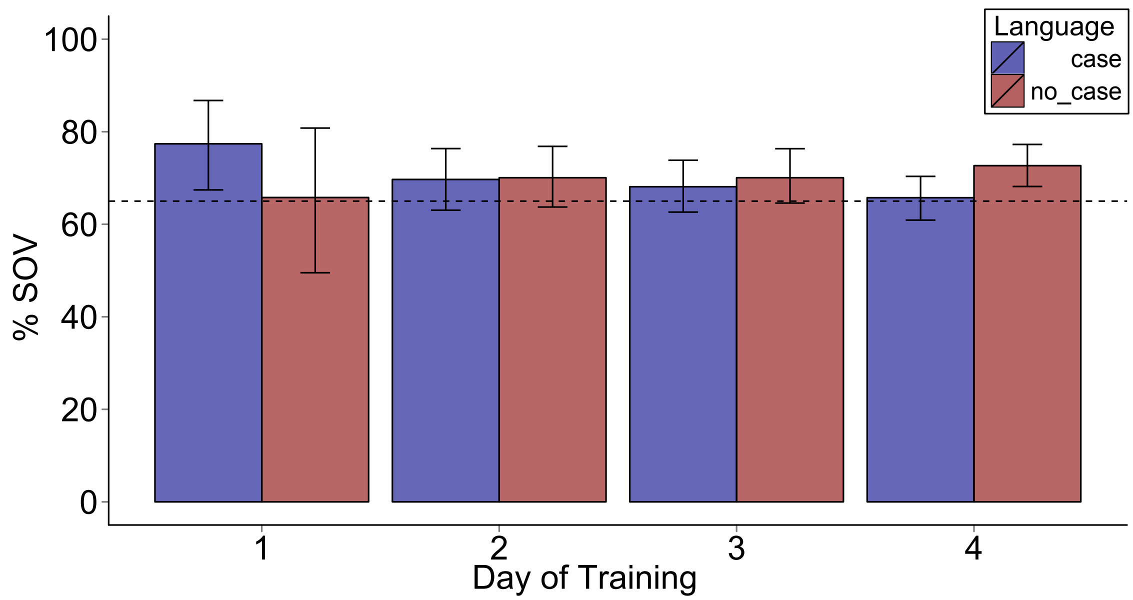

Figure 3.7: Constituent order production by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across languages).

The pattern of differential constituent order regularization depending on the presence of case-marking in the language is also evident in participants' treatment of the two novel verbs presented on Day 4 (see Figure 3.8). While learners of the case language closely matched the overall input frequency of the dominant (SOV) constituent order (χ^2 (1)=0.18, p=0.67), learners of the no-case language used SOV constituent order significantly more frequently than in the input (χ^2 (1)=5.1, $p<0.05$). As revealed by a mixed logit model analysis, the difference between the case and the no-case language for novel verbs was marginally significant ($\widehat{ß}$=0.67, z=1.7, p=0.09), further supporting our hypothesis.

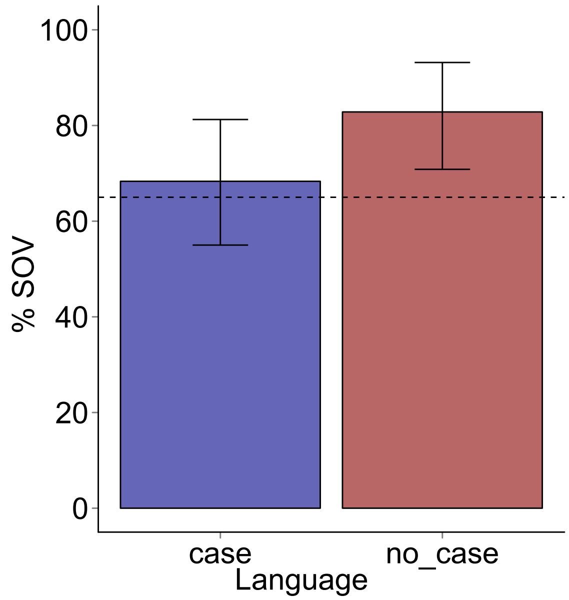

Figure 3.8: Constituent order production for previously unseen verbs in the two languages. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the overall SOV frequency in training (equal across languages).

3.2.4.3. Conditional entropy of constituent order

With our basic prediction confirmed, we investigated whether learners of the no-case also used more sophisticated strategies to reduce uncertainty in grammatical function assignment.

To address this possibility we calculated the entropy grammatical function assignment conditioned on the verb without taking case-marking into account as shown in (2). As expected, ANOVA analysis revealed a significant effect of language on conditional entropy of grammatical function assignment across the four days of the experiment (F(1, 156)=13.8, p<0.0003, see Figure 3.9): Learners of the no-case language converged on a language that had lower conditional entropy of grammatical function assignment than the language produced by the learners of the case language.

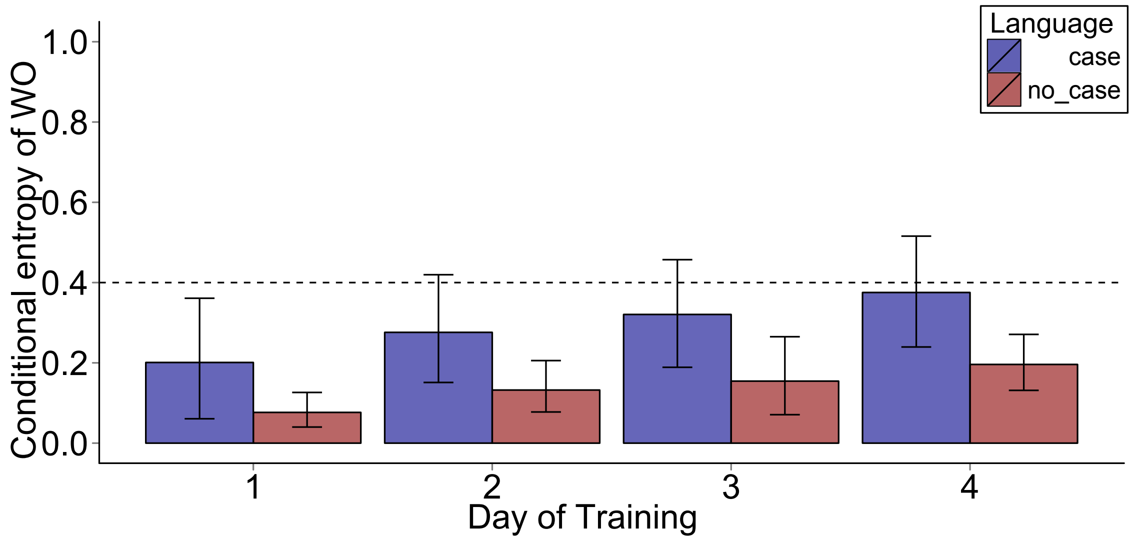

Figure 3.9: Conditional entropy of grammatical function assignment by language. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents conditional entropy of grammatical function assignment in the input (equal across languages).

3.3. General discussion

We find that case-marking and constituent order interact during language acquisition in a manner consistent with the typological correlation between constituent order freedom and the presence of case-marking in a language, further supporting the hypothesis that typological patterns are not due to chance.

In the case language, the case-marker was obligatory in our setup and always unambiguously signaled grammatical function assignment, thus allowing more freedom in constituent order variation. However, if grammatical functions are signaled exclusively by constituent order, as in our no-case language, flexible constituent order does not serve as a sufficiently informative cue to sentence meaning. Put differently, the uncertainty about form-to-grammatical function mappings is much higher in the no-case input language than in the case input language. Learners of the no-case language in Experiment 3 tended to reduce excessive uncertainty of this type by regularizing constituent order both in the language overall and in verb-specific ways.

The observed tendency to regularize constituent order more strongly in the no-case language is unlikely to be due to arbitrary mistakes since participants in Experiment 3 successfully learned the system of verb-specific biases: Constituent order preferences in production were clearly conditioned on the verb class in both languages (cf. Figure 3.6).

The comparison between Experiments 2 and 3 corroborates prior work in linguistic (Hudson Kam & Newport, 2009) and non-linguistic (Ferdinand et al., 2013) domains showing that adult learners are more likely to deviate from the input if the complexity of the system is high. Learners in Experiment 2 reliably reproduced the input frequency of the dominant constituent order, while learners in Experiment 3, where the complexity of the system was increased due to verb-specific biases, were more likely to deviate from the input as a result of a preference to trade off robust information transfer and effort.

One potential caveat in interpreting our results has to do with the structure of the production test. Recall that participants were asked to describe short computer-generated videos. The videos were played to participants and then frozen on the last frame, which stayed on the screen until the last trial began. Participants in our experiment showed a tendency to describe the videos from left to right. For example, in Experiment 2, participants were less likely to use SOV constituent order if the subject was on the right ($\widehat{\beta}$=-0.29, z=-2.20, $p<0.05$). This effect was equally strong in both languages, as suggested by the absence of a language x position interaction ($\widehat{\beta}$=0.03, z=2.55, p=0.79). Since the position of the subject was counterbalanced within the production block (and throughout the experiment) in Experiment 2, it is possible that this bias has 'pulled' participants' SOV responses towards the mean and thus obscured their preference to regularize the dominant constituent order in the absence of case. In Experiment 3, however, the position of the subject and object on the screen was (unintentionally) not counter-balanced: The subject was presented on the left 90% of the time. This was the case for both the case and the no-case language. Still, it is possible that this property of Experiment 3 boosted participants' preference to regularize the dominant order in the no-case language. It is unlikely, however, that the presentation order can fully explain our results since we observed differential constituent order regularization depending on the presence of case for conditions with identical presentation order. In fact, only learners of the no-case language tended to subtly deviate from the input; learners of the case language tended to match the input constituent order distribution.

The changes introduced by learners in the constituent order distribution as a result of a functionally-motivated bias observed in this chapter are, however, smaller than the changes to the case-marking distribution reported in Chapter 2. Case-marking in our experiments was represented by a one-syllabic word 'kah' that followed the noun it modified. Even though the case-marker was more salient in our stimuli than it typically is in natural language (i.e., an unstressed clitic), it was likely to be less perceptually salient than the content nouns it modified. Thus changes to constituent order may result in larger perceptual deviation from the input distributions.

One caveat to this interpretation is, however, that comprehenders do not appear to be very good at keeping track of sequential order information. For example, word order information is often overridden by plausibility (F. Ferreira, 2003), and words in a sentences appear to be easily susceptible to switches (Gibson, Bergen, & Piantadosi, 2013). In fact, some sentence processing accounts assume that verbal working memory involved in sentence processing is not sensitive to serial order information (Lewis, Vasishth, & Van Dyke, 2006).

With these caveats in mind, what conclusions can we draw about the historical change from Old English to Modern English? Experiments reported in this chapter and in Chapter 2 tentatively suggest that case was more susceptible to change than constituent order. In Chapter 2, we found that learners were more likely to introduce changes into case-marking even though the strategy of constituent order fixing was available in the language. In Chapter 3, we found that learners were more likely to regularize constituent order when it was the only cue to sentence meaning and thus introducing changes into the distribution of case-marking was not an option in the language. This effect, however, was a) fairly small and b) only observed if the complexity of the language was relatively high. Thus, one can speculate that a potentially rapid case-marker loss was more likely to precipitate a slower process of constituent order fixing (Marchand, 1951; Sapir, 1921) during the historical change from Old English to Modern English.

To summarize, our findings in Chapters 2 and 3 suggest that the well-documented inverse correlation between constituent order flexibility and the presence of a case system can be explained by learners' preference for grammatical systems that encode linguistic information efficiently. More generally, our results provide additional support for the hypothesis that at least some cross-lexical and grammatical properties of languages represent efficient trade-offs between effort and robust information transmission (Manin, 2006; Maurits et al., 2010; Piantadosi et al., 2011a, 2011b; Qian & Jaeger, 2012), although the results of Experiment 2 and 3 were relatively weak, leaving it to future work to elaborate on them. Our results also contribute to a growing body of work demonstrating the potential of using miniature artificial language learning to study the relationship between learning biases and language structures.

Chapter 4

Using cues to sentence meaning efficiently: Evidence from Optional Case-Marking

In Chapters 2 and 3 we have shown that some typological patterns of cue trade-offs at different levels of linguistic organization (e.g., syntax and morphology) can be explained by preferences for efficient linguistic systems that originate during language acquisition. In this chapter, we explore whether this preference can explain certain cross-linguistically recurring morphological properties such as differential (e.g., Hebrew, Sinhalese) or optional case-marking systems (e.g., Japanese, Korean).

Specifically, we investigate whether language learners are biased toward morphological systems that make efficient use of redundancy in the linguistic signal. In two experiments, we expose learners to miniature artificial languages designed in such a way that they do not use their formal devices (case-marking) efficiently to facilitate robust information transfer. We ask whether learners would alter the input language, providing more linguistic signal precisely for meaning components that are less expected, thereby efficiently trading off the effort required for encoding and decoding of the linguistic message against the speed at which the information is being transmitted. Such deviations from the input could be a vehicle for language change over generations. If these changes shift the input language toward typologically common patterns, this would provide further evidence that gradient linguistic universals can result from biases for efficient communication.

4.1. The Phenomenon: Differential Case-Marking Systems

As a test case, we investigate the acquisition of differential case-marking systems (Aissen, 2003; Bossong, 1985; Mohanan, 1994) found in a large number of natural languages (e.g., Sinhalese, Spanish, Russian, and Hindi). *Differential case-marking languages* mark only certain types of subjects and direct objects, leaving others zero-marked. While morphological case is thus optional in such systems, its occurrence and omission are highly principled and are generally associated with certain semantic properties of the referent such as animacy, definiteness, and person, as shown in (3).

(3)

Animacy scale: human > animate > inanimate

Definiteness scale: personal pronoun > proper name > other

Person scale: 1st, 2nd > 3rd

Referents that are higher on the dimensions in (3) are typically associated with the subject position, while the referents that are lower on those dimensions typically occur as sentential objects. This mapping from scales of referential properties (e.g., human > animate > inanimate) to the grammatical function hierarchy (e.g., subject > object) is sometimes referred to as ‘alignment’. For atypical alignments, grammatical functions are more often signaled by case-marking (Aissen, 2003; Mohanan, 1994; Silverstein, 1976).

Here, we investigate animacy effects. In differential case-marking languages, inanimate subjects and animate objects (less typical alignments) are categorically case-marked, whereas animate subjects and inanimate objects (more typical alignments) are categorically not case-marked (Aissen, 2003; Mohanan, 1994; Silverstein, 1976).

Optional case-marking languages, such as Korean and Japanese, exhibit the same general tendency as differential case-marking languages, but do so gradiently. That is, subjects and objects are more or less likely to be case-marked depending on how typical their referents are for the grammatical function they carry (Lee, 2006).

These animacy effects in optional and differential case-marking can be recast in terms of efficient information transfer through a noisy channel. Consider a simple transitive sentence in a hypothetical language with flexible constituent order (a language in which subject-object-verb and object-subject-verb constituent orders are both permitted, e.g., German or Korean), such as ‘The man the wall hit’. Here, the grammatical functions of ‘man’ and ‘wall’ cannot be identified based on the linear order of elements alone. If the intended message is that the man is hitting the wall, speakers can rely on listeners inferring the correct message because ‘the man’ (animate) is a typical agent (the doer of an action), and ‘the wall’ (inanimate) is a typical undergoer (the referent affected by an action). Case-marking will add little to such a sentence. However, the less the relative animacy of referents itself biases listeners towards the intended message, the more important case-marking becomes. This is most evident when animacy biases the listener towards the wrong interpretation (e.g., if the wall is hitting the man, for example, because it’s falling onto the man). Similarly, case-marking will help to facilitate successful communication when the noun referents rank equally on the animacy hierarchy ('The man the woman hit'). This logic extends to the cross-linguistically more typical case, in which constituent order provides *some* information (e.g., when subjects *tend* to precede objects): case-marking can always be used to further reduce the uncertainty about the intended meaning, but its usefulness is highest if the other cues (e.g., constituent order, animacy) do not bias listeners towards the intended meaning. Thus, under our hypothesis, a referential expression should be more likely to receive overt case-marking when its intended grammatical function is less expected, given other properties of the sentence including animacy (see also Jaeger, 2010).

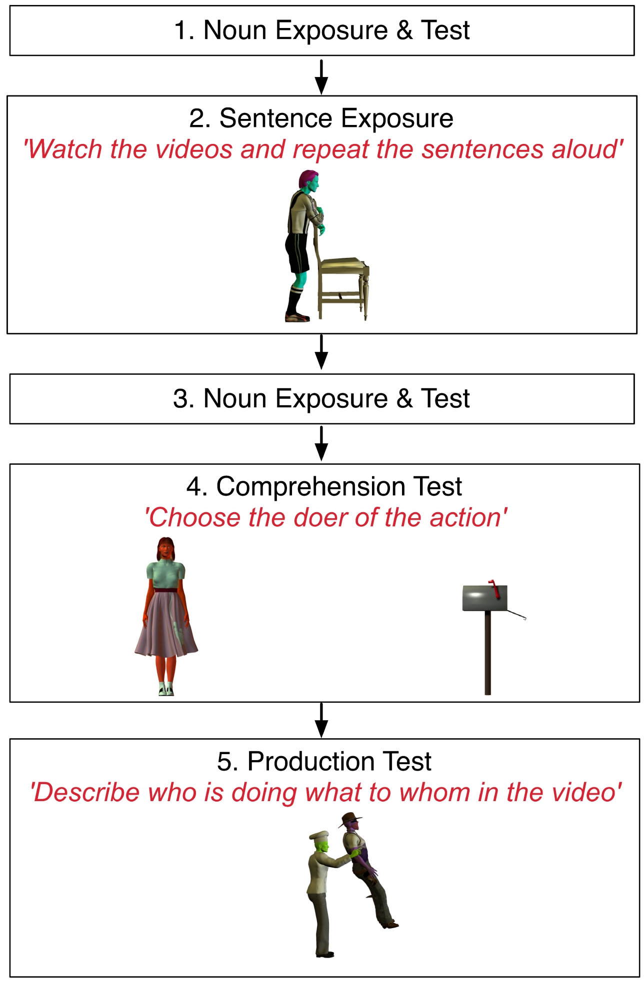

Figure 4.1: Experimental procedure showing still images of the video stimuli used in the experiments.

4.2. The Approach

In two experiments, we expose learners to languages that are inefficient versions of a verb-final language with flexible constituent order and optional case-marking. The miniature languages employed in our experiments resembled naturally occurring languages in that they had a dominant constituent order (subject-object-verb, SOV, 60%

of all sentences) and a less frequent constituent order (object-subject-verb, OSV, 40% of all sentences). Like many verb-final languages with flexible constituent order, our miniature languages contained case-marking. Crucially, however, our miniature languages deviated from naturally occurring languages in that case-marking was *not* conditioned on animacy. In Experiment 4, the grammatical object was optionally case-marked (in 60% of all input sentences). In Experiment 5, the grammatical subject was optionally case-marked (also in 60% of the input). In both experiments, case-marking appeared equally frequently on animate and inanimate noun phrases.

If learners are indeed biased to restructure the input language to increase its communicative efficiency, learners should introduce animacy-contingent case-marking; that is, increasing the use of case-marking on referents that are less likely to carry the grammatical function intended by the speaker, while leaving more expected referent-to-grammatical function assignments zero-marked. Importantly, participants in our experiments were monolingual speakers of English. English has no productive case-marking system. While there are remnants of a former case-marking system preserved in the pronominal system (e.g., he vs. him), English does not case-mark lexical nouns, such as those in our experiments. Crucially, English does not have optional case-marking. So, if observed, the introduction of animacy-contingent case-marking into the artificial language could not be due to transfer from their native language.

4.2.1. Participants

Participants in Experiments 4 and 5 were undergraduate students at the University of Rochester, all of whom were monolingual native speakers of English. Each participant was tested in only one of the two experiments. Participants were paid $5 on days 1-3 of the experiment and $25 upon completion of the fourth and final session. Twenty-nine participants completed Experiment 4, with 1 participant excluded due to experimenter error, 3 participants excluded for failing to achieve a 70% comprehension accuracy requirement (suggesting that overall they had not learned the language sufficiently) and 5 participants excluded for using the case-marker in all or none of their productions on the final day of training (2 used the case-marker in every production and 3 never used it). Thirty-three participants completed Experiment 5, with 4 participants excluded for failing to achieve a 70% comprehension accuracy requirement and 9 excluded for using case-marking in all or none of their productions (7 used the case-marker in every production, 2 never used it). Thus productions from 20 participants were analyzed in each experiment.

4.2.2. Procedure

Experiments 4 and 5 employed identical procedures. They differed only in certain aspects of the input languages presented to participants.

Participants visited the lab four times, each visit on a separate day with at most one day between visits. During each visit, participants saw a mixture of exposure and test blocks. There were two types of exposure blocks and two types of test blocks:

Noun Exposure. Participants viewed static pictures of people and objects one at a time and heard their names in the artificial language (30 trials total). The initial exposure was followed by a series of short vocabulary tests where participants were asked to choose the matching picture (out of two) for the character name they heard and to name the character shown on the screen. Feedback on performance was provided after each trial.

Sentence Exposure. Participants viewed 80 short computer-generated videos depicting transitive actions (one at a time) and heard an accompanying sentence describing the event in the artificial language. Participants were instructed to repeat each sentence aloud to facilitate learning.

Comprehension Test. In each of 80 trials, participants heard a novel sentence in the language, accompanied by two static pictures of the referents described in the sentence, and were asked to identify the doer of the action.

Production Test. Participants were shown a novel transitive scene (80 trials total) and were instructed to describe it in the language learned during the experiment, using a provided verb prompt.

On day 1, participants completed the following blocks: noun exposure, sentence exposure, noun exposure, and a comprehension test. On days 2-4, the sequence of blocks was the same as on day 1, except that a final production test block was added (see also Figure 4.1 above).

4.2.3. Input Languages

The input languages in both experiments contained 8 verbs and 15 nouns. Both input languages had flexible constituent order: subject-verb-object (SOV) order was dominant and occurred in 60% of the input sentences; object-subject-verb (OSV) order was the minority constituent order and occurred in 40% of the input sentences. Both languages had optional case-marking but differed in whether the grammatical object (Experiment 4) or subject (Experiment 5) was optionally case-marked. The case-marker was always *kah* and it always followed the noun whose case it marked. The frequency of case-marking was identical across the two experiments: 60% of objects (Experiment 4) or subjects (Experiment 5) were overtly case-marked and 40% were not. By design, case-marking was always independent of animacy (i.e., animate and inanimate nouns were equally likely to be case-marked). Case-marking did vary by constituent order: 50% of OSV sentences were case-marked and 67% of SOV sentences were case-marked.

In both experiments, the actions and the verbs were compatible with any of the referents being either the agent or theme. There were no differences in subcategorization frequencies between the verbs. That is, the frequency with which a noun was the subject or object did not differ between the verbs. The referents of the nouns and the actions referred to by the verbs differed, however, between languages (the former by design, the latter by necessity since the inanimate agents employed in Experiment 5 strongly constrained the choice of compatible actions).

Experiment 4: Input Lexicon

Verbs: *shen* (CHOP), *daf* (HUG), *kleidum* (HEADBUTT), *slergin* (KICK), *blefee* (PICK UP), *zamper* (POKE), *prog* (PUNCH), *geed* (ROCK).

Nouns: There were 15 nouns (*slagum, tombat, nagid, melnawg, norg, glim, plid, nagid, klamen, dacin, zub, vams, bliffen, rungmat, lombur*), 10 of which represented human characters (MOUNTIE, CHEF, REFEREE, CONDUCTOR, HUNTER, BANDIT, SINGER, WOMAN FROM THE 50S, COWGIRL, and FLIGHT ATTENDANT) and 5 represented inanimate objects (TABLE, CHAIR, MAILBOX, BOOKSTAND, and STATUE). Out of 10 animate nouns, 5 occurred exclusively in the subject position (SINGER, REFEREE, WOMAN FROM THE 50s, MOUNTIE, and CHEF), and the other 5 occurred exclusively in the object position (CONDUCTOR, HUNTER, BANDIT, COWGIRL, and FLIGHT ATTENDANT). Inanimate nouns occurred only in the object position. Each verb occurred twice with each of the subject nouns and once with each of the object nouns. The assignment of nouns to animate or inanimate categories was counterbalanced between the participants for object nouns to prevent any accidental (e.g., phonological or semantic) associations.

Experiment 5: Input Lexicon

Verbs: *shen* (BREAK), *daf* (PUSH), *kleidum* (DRAG), *slergin* (NUDGE), *blefee* (CARRY), *zamper* (CRUSH), *prog* (JUMP OVER), *geed* (KNOCK OVER).

Nouns: There were 15 nouns (*slagum, tombat, nagid, melnawg, norg, glim, plid, nagid, klamen, dacin, zub, vams, bliffen, rungmat, lombur*), 5 of which represented human characters (MOUNTIE, CHEF, REFEREE, CONDUCTOR, HUNTER, and BANDIT) and 10 represented inanimate objects (CAR, JEEP, SHOPPING CART, TRICYCLE, BICYCLE, SCOOTER, MOTORBIKE, BABY CARRIAGE, TRUCK, and BOGIE). Like in Experiment 1, there was no overlap between object and subject nouns: 5 inanimate nouns occurred exclusively in the subject position (CAR, BABY CARRIAGE, SCOOTER, TRICYCLE, and TRUCK), and the other 5 occurred exclusively in the object position (JEEP, SHOPPING CART, BICYCLE, MOTORBIKE, and BOGIE); animate nouns occurred as subjects only. The assignment of nouns to animate or inanimate categories was counterbalanced between the participants for object nouns.

4.2.4. Scoring

In the comprehension test, we assessed how successful participants were at decoding the message we intended to convey. Comprehension responses were scored as correct if the interpretation chosen by a participant matched our intended interpretation. For case-marked (unambiguous) trials, this measure indicated whether participants learned the meaning of case-marking. For non-case-marked trials, this measure indicated participants' reliance on animacy as a cue to sentence meaning. Non-case-marked trials were ambiguous when the referents were matched for animacy. This measure allowed us to assess the possibility that participants noticed the fact that the sets of subject and object animate nouns in Experiment 4 and subject and object inanimate nouns in Experiment 5 were non-overlapping, which disambiguates grammatical function assignment. The results reported below show that participants did not pick up on this information. Presumably, the language was sufficiently complex for them not to notice these gaps in the assignment of grammatical functions (subject, object) to nouns.

In production, we scored lexical errors, constituent order and the presence of case-marking on the object (Experiment 4) or subject (Experiment 5). Occasionally, participants mispronounced the name of one or two referents. Productions where both referents were mispronounced were labeled as incorrect. If only one of the two referents was mispronounced, it was still possible to successfully identify constituent order. Such productions were scored as overall correct but containing a lexical error. Productions that contained grammatical mistakes (i.e., incorrect use of case-marker or constituent order) were labeled as incorrect. Incorrect productions were excluded from all analyses.

Both languages were acquired with a high degree of accuracy, providing evidence that the task was feasible. The total number of grammatical mistakes (incorrect use of case-marker or constituent order) was around 6.5% in Experiment 4 and around 1% in Experiment 5 across all days; lexical mistakes (such as using an incorrect noun or verb) were also rare (7.4% of all productions in Experiment 4 and 6.8% of total productions in Experiment 5).

4.3. Experiment 4

All sentences in Experiment 4 represented simple transitive actions, such as ‘poke’ or ‘hug’, performed by a human actor on either human or inanimate undergoers, which occurred equally often in the exposure. Since the language had flexible constituent order, sentences with human objects were ambiguous if the object was not case-marked, but sentences with inanimate objects could be disambiguated based on animacy even without a case-marker.

If language users indeed try to communicate efficiently, they should restructure the language as they learn it, making it similar to differential object marking found in natural languages. In particular, if language learners are biased towards communicatively efficient linguistic systems, we would expect them to mark animate objects with an overt case-marker more frequently than inanimate objects.

4.3.1. Results

First we examined data from the comprehension test, asking whether a primary function of case-marking is to disambiguate the intended actor and undergoer. As expected, when both referents were animate and there was no case marking on the object, there were many misinterpretations of the intended meaning (53% mean accuracy, not significantly different from chance, $\chi^2(1)=0.28$, $p=0.59$). Overt object case-marking significantly and substantially increased the accuracy of responses (88% accuracy, significantly different from chance, $\chi^2(1)=63.82$, $p<0.0001$).

We then examined the data from the production test, to see what participants learned about the language. We used a mixed logit model (Breslow & Clayton, 1993; for an introduction, see Jaeger, 2008) to predict the presence of case-marking used in participants' productions based on object animacy (*animate/inanimate*), constituent order

(*SOV/ OSV*) and day of training (*2, 3, 4*) as well as all interactions between these factors. The three-way interaction between object animacy, constituent order, and day of training did not result in a significant change ($\chi^2(2)=3.1$, $p=0.21$) in the model fit and was removed from the final model. The model included the maximal random effects structure justified by the data based on model comparison.

Do participants restructure the language in their productions to make more efficient use of its formal devices? Consistent with our hypothesis, participants' productions deviated from the input in that atypical objects were more likely to be case-marked (see Figure 4.2A). Learners used significantly more case-markers on atypical (animate) objects than on typical (inanimate) object across all days of testing ($ß=0.35$, $z=2.27$, $p<0.05$), even though this was not the pattern of their input language. This pattern of conditioning overt case-marking on animacy closely mirrors the pattern commonly found in differential object marking systems (Aissen, 2003; Lee, 2006).

We also found that objects were more likely to be overtly case-marked if the constituent order was OSV ($ß=1.06$, $z=2.14$, $p<0.05$) (see Figure 4.2B). This pattern is opposite to the input distribution, where more case-marking was used on objects in SOV sentences. There are several possible explanations of this result, some of which provide further support for our hypothesis. We postpone the discussion of this finding until after Experiment 5.

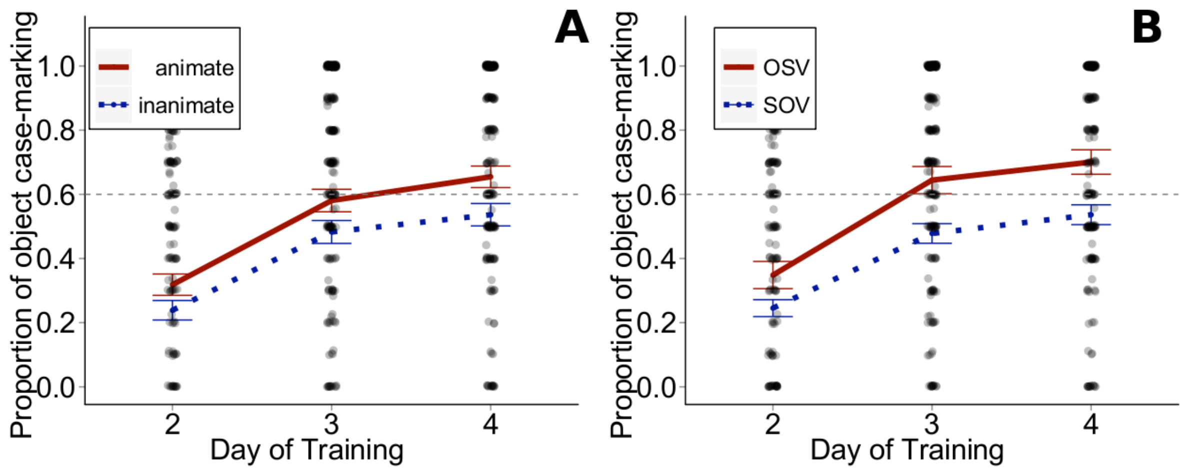

Figure 4.2: Overt case-marking by animacy of object (A) and constituent order (B) in production in Experiment 4. Lines represent condition means, dots represent overall subject means. Error bars represent 95% confidence intervals. The dashed line indicates proportion of case-marking provided in the input (invariant across animacy).

The observed effect of animacy is also compatible with an alternative explanation: the higher proportion of case-marking on animate objects might arise because animate referents attract more visual attention (Yarbus, 1967), which might cause participants to learn case-marking earlier or more successfully for animate referents. This concern was addressed in Experiment 5, which explored optional *subject* case-marking. If the results from Experiment 4 are due to a bias to case-mark the atypical, as we hypothesize, then the opposite pattern should hold for optional subject case-marking. We would expect participants to be more likely to use case-markers on inanimate subjects, while leaving the typical animate subjects more frequently zero-marked. In contrast, if the observed behavior is due to increased attention to animate referents, we would expect participants to case-mark animate referents more frequently in both experiments.

4.4. Experiment 5

The input language in Experiment 5 was the complement of the language used in Experiment 4. In Experiment 5, the animacy of subject varied (50% of subjects were animate and 50% were inanimate); objects were always inanimate. Sentential subjects were optionally case-marked independently of animacy, while objects were always zero-marked. All other aspects of the input grammar were the same as in Experiment 4.

4.4.1. Results

We first analyzed data from the comprehension test, asking about listeners' accuracy in decoding the intended meaning. As in Experiment 4, learners showed chance performance (52% mean accuracy, $\chi^2(1)=0.71$) when the referents were matched for animacy and the subject was not overtly case-marked. Performance was substantially improved and was significantly above chance when subjects were case-marked (94% mean accuracy, $\chi^2(1)=91.7$, $p<0.0001$) or were animate (82% mean accuracy, $\chi^2(1)=35.8$, $p<0.0001$).

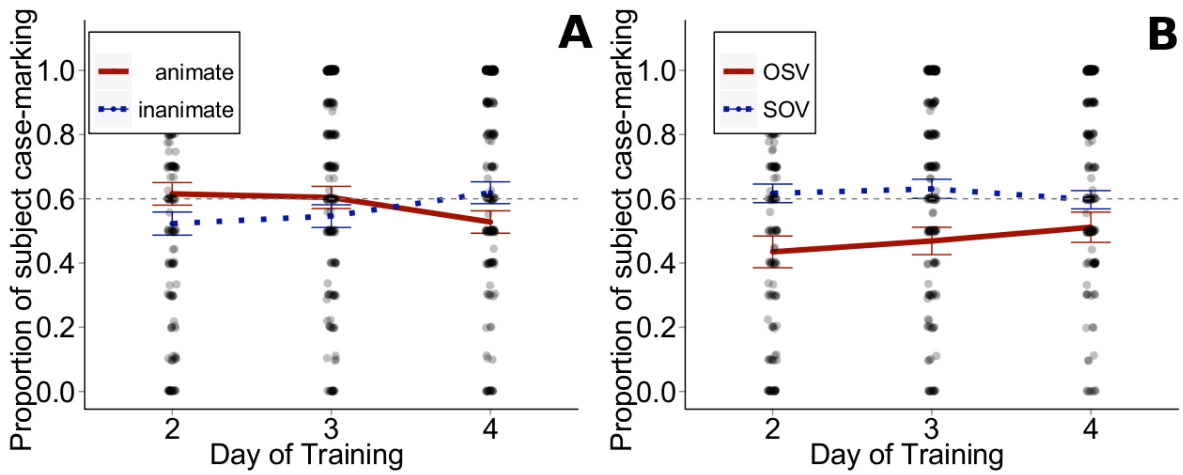

Figure 4.3: Overt case-marking in production by animacy of subject (A) and constituent order (B) in Experiment 5. Lines represent condition means, dots represent overall subject means. Error bars represent 95% confidence intervals. The dashed line indicates proportion of case-marking provided in the input (invariant across animacy).

We used the same procedure in analyzing participants' productions in Experiment 5 as in Experiment 4. The three-way interaction between subject animacy, constituent order, and day of training did not significantly contribute to the model fit (χ^2 (2)=0.73, p=0.69) and was not included into the final model. As in Experiment 4, we included the maximal random effects structure justified by the data based on model comparison.

Figure 4.3A shows the data from participants' productions. On the first day of testing, animate referents were case-marked significantly more frequently than inanimate referents (ß=-0.33, z=-2.5, p<0.05). This behavior is consistent with the alternative hypothesis, that the higher proportion of case-marker use with animate referents may be driven by properties associated with animacy. However, this bias to case-mark animate referents, evident at early stages of learning, gradually weakens as training continues, giving way to a bias toward efficient information transfer, which emerges through language exposure as learners become more proficient. This is evidenced by a significant Day × Animacy interaction ($ß=0.22$, $z=2.87$, $p<0.01$): As expected under our hypothesis, on the final day of training learners show the opposite preference and use more case-marking on atypical inanimate subjects than on animate subjects. This difference in case-marker use on animate and inanimate subjects on the final day of training was, however, not statistically significant ($ß=0.30$, $z=1.16$, $p=0.25$) as revealed by the simple effects test.

We also examined case-marker use in relationship to word order (see Figure 4.3B). In Experiment 4, we observed more frequent object case-marking in the OSV order. Such word-order contingent case-marking could be driven by at least two biases. First, as hypothesized above, OSV order may bias the listener to an incorrect grammatical function assignment; hence, case-marking is used to avoid potential miscommunication. Alternatively, word-order contingent case-marking may reflect a bias to mention disambiguating information as early as possible in the sentence (Hawkins, 2007). For Experiment 5, such a bias toward early disambiguation makes the opposite prediction compared to Experiment 4: in a language with subject case-marking, more frequent case-marking should be observed in the SOV order since this provides information about grammatical function assignment earlier in the sentence.

The results of Experiment 5 suggest that both biases are in play. There was a main effect of word order: overall, significantly more subjects were overtly marked when the constituent order was SOV ($ß=0.93$, $z=2.40$, $p<0.05$), which is indicative of a bias to provide disambiguating information at the earliest possible moment. This bias, however, gradually weakened as training continued, as suggested by the significant Word order × Day interaction ($ß=-0.16$, $z=-4.11$, $p<0.001$). There was no significant preference to differentially case-mark subject referents depending on sentence word order on the final day of training ($ß=0.60$, $z=1.53$, $p=0.13$). This might indicate a point that participants' productions would start to reflect the bias to mark the atypical if training continued.

Importantly, the more complex (one might say, weaker) results of Experiment 5 actually parallel quite nicely the typological data from natural languages. Differential object marking is cross-linguistically highly consistent: Languages with animacy-contingent differential object case-marking tend to follow the pattern found in Experiment 4 (Malchukov, 2008). In contrast, differential subject marking in natural languages is typologically less clear-cut, and this was also true of learners in Experiment 5. The two competing acquisition biases observed in Experiment 5 (a bias to case-mark animate referents and a bias to case-mark less expected referent-to-grammatical function assignments) are manifested typologically as well. Many languages, such as Mangarayi (Merlan, 1982), overtly mark inanimate subjects and leave animate subjects zero-marked, but there are languages (e.g., Samoan) that have been claimed to show the opposite pattern (Mosel & Hovdhaugen, 1992).

4.5. Discussion

It has long been hypothesized that communicative pressures on language can operate during acquisition (Bates & MacWhinney, 1982). The studies presented here provide experimental evidence supporting and clarifying this hypothesis. Our results suggest that language learners are biased towards communicatively efficient linguistic systems and restructure the input language in a way that facilitates information transfer, in line with recent information-theoretic approaches to language production (Genzel & Charniak, 2002; Jaeger, 2010; Levy & Jaeger, 2007). In our experiments this bias affects the acquisition of an optional case-marking system: although case-marking in the input language is independent of animacy, learners showed a tendency to condition case-marking on animacy, with the less expected alignments of animacy and grammatical function (inanimate subjects or animate objects) becoming more likely to be case-marked. Note that learners could instead have generalized case-marking to all nouns, regardless of animacy. This would have maximized the chance of communicative success at the expense of effort and, possibly, processing speed (reducing the rate of information transmission), since case-markers would be produced even when the intended meaning could be inferred in their absence. However, very few participants showed full case-marker generalization, suggesting that the trade-off between successful communication and effort was indeed at work during learning. This conceptually replicates our findings in Chapter 2 with different miniature languages, where learners of non-fixed order languages chose to condition case-marking on constituent order rather than to regularize case-marking in the language to 100%.

The observed bias towards efficient linguistic systems is not reducible to previously documented tendencies of learners to regularize inconsistent structures (Hudson Kam & Newport, 2005, 2009), biases to reduce the representational complexity of linguistic systems (Kirby et al., 2008; Smith & Wonnacott, 2010), or a native language bias since we exposed native speakers of English (a language with no case-marking on nouns) to an artificial language with optional case-marking.

Our results suggest that learners do introduce typologically common patterns into the language. The learning outcomes in our experiments closely mirror natural phenomena, such as the optional case-marking systems found in Japanese and Korean, where animate objects and inanimate subjects are more likely to receive overt case-marking (Lee, 2006). The close correspondence between the patterns observed during acquisition and those found in typological data suggests that some of the properties of natural languages may be shaped by learning biases that stem from a preference for communicatively efficient linguistic systems.

In this way, our results complement previous artificial language learning studies of phonology (Finley & Badecker, 2008; Newport & Aslin, 2004), lexical, and syntax acquisition (Culbertson et al., 2012; Kirby et al., 2008; Morgan et al., 1987; Smith & Wonnacott, 2010) showing behavioral evidence for linguistic universals. Together these and our studies demonstrate the power of the artificial language learning paradigm as a complement to typological work on linguistic universals (cf. Tily & Jaeger, 2011). The biases we have observed during the acquisition of optional case-marking provide a possible mechanism for patterns observed cross-linguistically (Manin, 2006; Piantadosi et al., 2011a, 2011b) and during native language production by adult speakers (Aylett & Turk, 2004; A. Frank & Jaeger, 2008; Genzel & Charniak, 2002; Jaeger, 2006, 2010; Levy & Jaeger, 2007).

As discussed in Chapter 1, we assumed noise-free recognition of words throughout this dissertation. Thus, for all experiments presented so far, robust information transfer in the presence of noise can also be conceived as ambiguity avoidance. The two experiments presented in this chapter can potentially provide an insight into whether learners' behavior is better explained by a strict ambiguity avoidance account or by information transmission in the presence of noise. Recall that in Experiments 4 and 5, grammatical function assignment was lexically-specific: Each noun in our setup was categorically associated with either the subject or object position in the sentence. Thus, for a learner, who has acquired the categorical associations between grammatical function and lexical items in the input, there would be no ambiguity with regard to grammatical function assignment. There would still be differential uncertainty about grammatical function assignment associated with the animacy of the referent. In particular, in Experiment 4, animate referents would be associated with higher uncertainty about grammatical function assignment than inanimate referents; and in Experiment 5 inanimate referents would be associated with higher uncertainty than animate ones. If participants in Experiments 4 and 5 have indeed learned from the input that grammatical function assignment is unambiguously associated with lexical items in our setup, our results would provide further support for the information transmission in noise account.

To test whether learners have acquired lexically-specific grammatical function assignment in the miniature languages, we analyzed learners' responses on comprehension trials where both referents were matched for animacy and case-marking was not present. Knowing these specific associations, learners could unambiguously infer the correct sentence interpretation even in the absence of case-marking. However, as discussed in Section 4.3.1, learners in both experiments were at chance in identifying the intended meaning of the sentence when case-marking was absent and animacy was not a good cue to sentence meaning. This outcome suggests that learners in both experiments failed to acquire the associations between grammatical function assignment and specific lexical items from the input. A likely explanation for this failure is that learners primarily relied on their prior expectations stemming from experience with natural language, where grammatical function is not conditioned on specific lexical items.

Thus, our data cannot currently disentangle whether the learning outcomes observed in our research are better explained by the information-theoretic perspective assumed in this dissertation or by the ambiguity avoidance account. We leave this issue for further research.

Chapter 5

Discussion

One of the central objectives of modern linguistics is to identify the principles that characterize possible human languages. To this end, linguists have examined languages of the world to find patterns that recur across languages (*linguistic universals*). The origins of such recurring patterns have been the subject of long-standing debate in linguistics and cognitive science.

In this dissertation, we focused on the hypothesis that one of the biases shaping languages over time is a preference for communicatively efficient linguistic systems and addressed the possibility that these preferences operate during language acquisition. Using a miniature artificial language learning paradigm, we provided direct behavioral evidence that learners alter the input languages, providing additional cues to the intended meaning of the sentence precisely (and only) when other properties of the sentence would likely cause processing difficulty or misinterpretation. Such deviations from the input language toward a linguistic system that makes more efficient use of redundancy in the linguistic signal could be a vehicle for language change over generations.

In a series of miniature language learning experiments we have shown that these changes consistently shift the input languages toward typologically common patterns, which provides evidence that gradient linguistic universals can result from biases for efficient communication.

In Chapters 2 and 3, we showed that the implicational universal describing the correlation between constituent order freedom and the presence of case-marking in a language (languages with flexible constituent order typically have a case system, while languages with fixed constituent order typically lack case) can be explained by biases operating during language acquisition. From the information-theoretic standpoint, this statistical universal can be explained by a preference for relatively efficient linguistic systems and a bias against linguistic systems that either have excessive uncertainty or abundant cue redundancy. The learning outcomes in our experiments support this prediction. In Chapter 2, we found that learners exposed to three miniature artificial languages with optional case-marking learned case differently, depending on the amount of constituent order flexibility in the language: They were more likely to maintain case-marking in the language when it was not redundant (i.e., when constituent order was not fixed). Chapter 3 provided further – albeit weaker – support for our hypothesis. Consistent with our predictions, we observed differential constituent order regularization depending on the presence of case-marking in a language: At least in one of the two experiments we conducted, learners of the no-case language tended to regularize constituent order, making it a stronger cue to grammatical function assignment, while learners of the case language maintained constituent order flexibility.

In Chapter 4, we extended this line of research to show that a preference for efficient linguistic systems during learning can explain cross-linguistically recurring properties of differential (e.g., Hebrew, Sinhalese) and optional case-marking systems (e.g., Japanese, Korean). The omission of morphological case in differential and optional case-marking systems is highly principled and is generally associated with certain semantic properties of the referent such as animacy, definiteness, and person (Aissen, 2003; Lee, 2006; Mohanan, 1994; Silverstein, 1976). These patterns can stem from pressures associated with efficient information transfer: A referential expression is more likely to receive overt case-marking when its semantic or other properties bias the listener away from the intended grammatical function assignment. We found that learners presented with miniature artificial languages that do not use case-marking efficiently (e.g., do not condition case-marking on animacy of the referent), did not veridically reproduce 'inefficient' morphological systems in the input. Instead learners deviated from the input and used more case-marking for animate objects and inanimate subjects, thus making the case systems in the newly acquired languages more efficient and in line with typological data.

In the remainder of this dissertation, we will address some outstanding questions regarding this research. Specifically, we will consider the possible mechanisms underlying the findings observed in our experiments and discuss the implications of these findings in relation to ambiguity avoidance accounts. We will conclude with a discussion of some promising avenues for further work.

5.1. Outstanding Questions

What are the mechanisms underlying our findings?

Our findings raise questions about the precise nature of the mechanism underlying the biases we observed.

The learning outcomes in our experiments resemble the patterns found in online productions of adult speakers (Aylett & Turk, 2006; A. Frank & Jaeger, 2008; Gomez Gallo et al., 2008; Jaeger, 2010; Kurumada & Jaeger, 2013, submitted; van Son & van Santen, 2005). Most relevant to the work presented in this dissertation is a recent study by Kurumada and Jaeger (submitted). Using a spoken recall paradigm, they found that speakers of Japanese were more likely to produce object case-marking when sentence properties (such as animacy of the object or plausibility of grammatical function assignment) were likely to bias the listener away from the intended interpretation even when the grammatical subject was always overtly case-marked (i.e., when there was no ambiguity in grammatical function assignment). This suggests a possibility that learners' preferences observed in our experiments can originate in the human production system, which is organized to prefer efficient information transfer (Aylett & Turk, 2004; Jaeger, 2010; Levy & Jaeger, 2007; Lindblom, 1990; van Son & Pols, 2003).

This raises the question of how these preferences for efficient information transmission enter the production system. Drawing on findings in motor control (cf. Wei & Kording, 2010), one proposal (Jaeger, 2013; Jaeger & Ferreira, 2013) suggests that speakers implicitly learn to balance robust information transfer and production effort by minimizing task-relevant prediction errors during communication. On this view, speakers continuously monitor the probability of their previous communicative successes across a variety of contexts and integrate this feedback into their subsequent productions. Even though participants in all our experiments did not explicitly engage in a communicative task, they are sensitive to the implicit feedback from the perception of their own productions. It is possible that our learners received little or no implicit negative feedback in those contexts when sentence meaning was easily inferable (e.g., when there was a highly informative cue to grammatical function assignment such as fixed constituent order [Chapter 2] or animacy [Chapter 4]) and thus persisted at using zero-marked forms in these contexts. When other cues to sentence meaning did not provide sufficient information to successfully infer grammatical function assignment (e.g., random constituent order [Chapter 2]), this feedback could have had a more negative effect and prompted learners to use case-marking more frequently in these contexts when they produced their own utterances.

If the learning outcomes in our experiments originate as biases in the production system, certain mismatch between learners' preferences in comprehension and production would be predicted: We might expect learners to deviate from the input in their productions, but to match the input more closely in their comprehension performance or at least, we would not expect comprehension to 'lead the way'. To address this possibility, we compared participants' preferences in production and comprehension. Consider Figure 5.1 that shows learners' constituent order preferences in comprehension and production in Experiment 1 reported in Chapter 2. As Figure 5.1 shows, the pattern in participants' production (top panel) qualitatively closely resembles their comprehension performance (bottom panel). Specifically, in comprehension as well as in production, learners match the input frequencies of constituent order variants more closely if case-marking is present in the sentence and tend to regularize the dominant SOV constituent order when case-marking is absent (see Figure 5.2 for an illustration of this preference on the final day of training).

This close correspondence between production and comprehension performance suggests that learners have acquired a novel grammar that is shared between the two systems and thus our findings are unlikely to originate exclusively in production. There is also more variability in learners' production as indicated by larger confidence intervals, suggesting that the observed preferences were more consistent in comprehension, further supporting the assumption that learners' behavior is not production-specific.

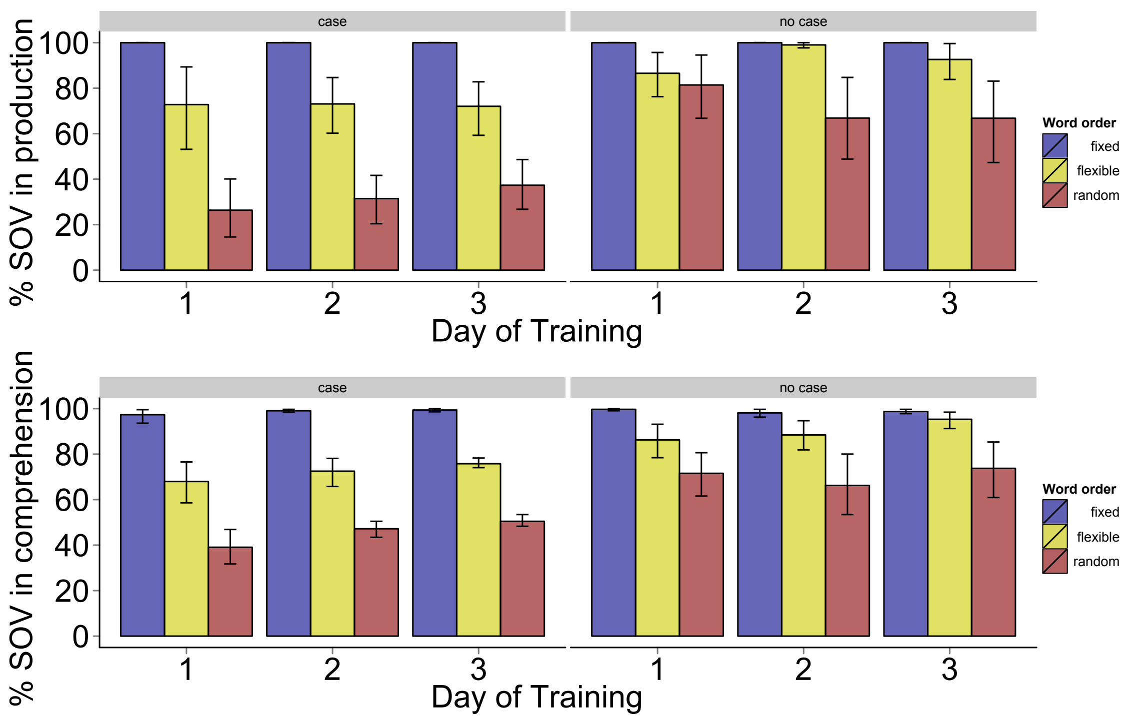

Figure 5.1: Constituent order preferences in Experiment 1 reported in Chapter 2 in production (upper panel) and comprehension (lower panel) on case-marked (left panels) and non-case-marked (right panels) trials. The error bars represent bootstrapped 95% confidence intervals.

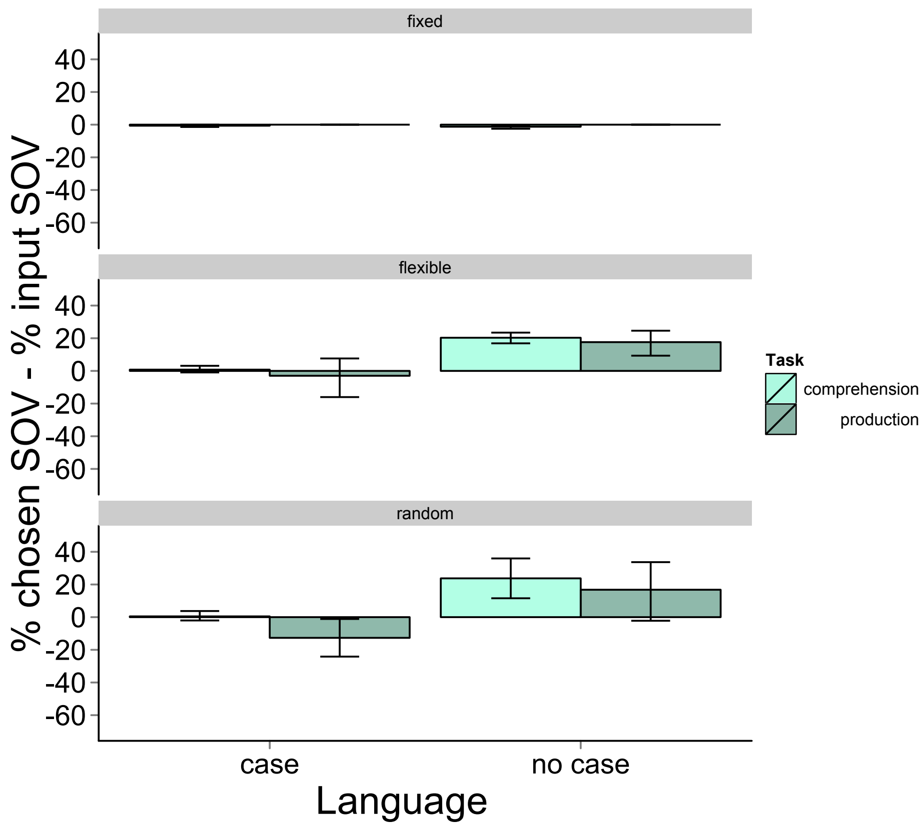

Figure 5.2: Deviations from the input proportion of the dominant constituent order in Experiment 1 reported in Chapter 2 in the fixed (upper panel), flexible (middle panel), and random (bottom panel) constituent order in comprehension and production. The error bars represent bootstrapped 95% confidence intervals.

A more likely possibility is that the patterns observed in our experiments originate in learning that is not specific to production: Learners may have misconstrued some of the sentences they were exposed to, altering the characteristics of the input from which they learned. Indeed, as shown in Figure 5.1, participants' comprehension responses tended to closely match the input in the presence of case-marking, but consistently deviated towards the dominant SOV constituent order in its absence, indicating that some OSV sentences were misinterpreted as SOV when case-marking was not present. This performance suggests that misconstruals were most common in the absence of case-marking and, in particular, when other cues to sentence meaning did not provide sufficient information to successfully infer grammatical function assignment (e.g., in the non-fixed constituent order languages in Chapter 2). As an illustration, let us consider for example how the preference to use more case-marking in OSV sentences observed in Chapter 2 would follow from this account. Consider a hypothetical language where learners are exposed to an equal number of SOV and OSV sentences with and without case-marking as shown in (4):

(4)

SOV case-marked sentence – 50 tokens

OSV case-marked sentence – 50 tokens

SOV non-case-marked sentence – 50 tokens

OSV non-case-marked sentence – 50 tokens

Learners would correctly perceive grammatical function assignment in most case-marked sentences (as evidenced by near-ceiling comprehension performance on case-marked trials reported in Chapter 2). Some non-case-marked OSV trials, however, would be misinterpreted as SOV (as suggested by comprehension preferences shown in Figure 5.1). Let us assume for the sake of the argument that 25 out of 50 non-case-marked OSV tokens were interpreted as SOV by our learners, thus resulting in the perceived distribution as shown in (5). This could have led to a higher perceived proportion of case-marking in OSV sentences (50/75=0.67) compared to SOV sentences (50/125=0.4). Since production preferences appear to closely resemble comprehension performance in our experiment, this perceived pattern could have manifested in production as asymmetric case-marking (i.e., a preference to use more case in OSV sentences) observed in our experiments.

(5)

SOV case-marked sentence – 50 tokens

OSV case-marked sentence – 50 tokens

SOV non-case-marked sentence – 75 tokens

OSV non-case-marked sentence – 25 tokens

Where can such ‘misconstruals’ arise? One possibility is that learners might have misinterpreted some of the sentences already during perception as they watched short videos and heard the accompanying sounds (cf. Guy, 1996; Ohala, 1989). It is possible that learners’ perception of sentences was influenced by their native language experience. In particular, learners could have perceived non-case-marked OSV constituent order as SOV in some sentences since SOV constituent order is more in line with their prior expectations stemming from their native language (English), where it is highly uncommon for the object to precede the subject. This type of explanation is somewhat similar to the idea of the production-perception loop in exemplar-based models (Pierrehumbert, 2002). In these models, exemplars that are too confusable are essentially ‘filtered’ during perception: Exemplars that are not correctly perceived by the listener are not stored in memory and thus do not influence users’ productions.

Certain properties of our experimental design, however, make this possibility somewhat unlikely. First, the meaning of the sentence was always represented by an accompanying video, thereby unambiguously conveying the intended grammatical function assignment. Given that there was also no time pressure, it is relatively unlikely that misinterpretations during perception were sufficiently frequent to create the observed effect.

It is more probable that this type of 'misinterpretation' arises later, when the extracted interpretations are consolidated in memory. A recurrent finding in verbal and non-verbal short-term memory is that items interfere with each other during encoding and retrieval (e.g., Anderson, 1983). This interference is increased if the items that need to be retrieved/encoded are in some way similar to each other (e.g., structurally, semantically, spatially, etc.). This type of interference has been shown to affect sentence comprehension as well (Lewis, 1996; Lewis & Nakayama, 2001; Lewis et al., 2006; Van Dyke & Lewis, 2003): Increasing syntactic or semantic similarity between the target and preceding or following material poses additional difficulty during retrieval (as manifested in longer reading times and increased comprehension errors). There is also some evidence that similarity-based interference poses difficulty during encoding of the target in memory (Gordon, Hendrick, & Levine, 2002). Since all referents in Experiment 1 were human and male, it likely that participants in our experiments experienced this type of interference and thus misassigned grammatical roles in a certain number of comprehension trials. Such misassignments could have introduced changes into the input distribution when it was consolidated in memory.

A similar influence of similarity-based interference has been found in sentence planning (Bock, 1987; Ferreira & Firato, 2002; Gennari, Mircovic, & MacDonald, 2012). For example, Gennari et al. (2012) found that speakers were more likely to use passive if both actor and patient were animate (i.e., semantically similar). Similarity-based interference during production *alone* does not, however, explain our results. Since all referents in our languages in Chapter 2 were animate, this account would predict an equal degree of case-marker regularization in the three languages regardless of the informativity of constituent order.

The hypothesis that the outcomes observed in our experiments indeed arise outside of the production system (whether in perception or memory) is broadly compatible with functional theories of efficient communication. This hypothesis, however, constitutes a different conceptualization of these effects than most current theories of efficient communication, which primarily attribute these findings to the organization of human production system (Aylett & Turk, 2006; Genzel & Charniak, 2002; Jaeger, 2010). If misconstruals in memory or perception indeed underlie the findings observed in our research, current accounts of efficient communication would need to be re-evaluated to account for the precise memory/perception mechanisms that give rise to efficient patterns in human production.

Regardless of the exact mechanisms underlying learners' behavior, our results suggest that some learning biases that shape language structures and language change can be traced back to learners' preferences for efficient linguistic systems. Since the research reported in this dissertation is directly inspired by information-theoretic approaches to language production (Aylett & Turk, 2006; A. Frank & Jaeger, 2008; Jaeger, 2010; Piantadosi et al., 2011a, 2011b; van Son & van Santen, 2005), we have linked learners' biases observed in our experiments to the organization of the human production system in the previous chapters. This assumption would need to be re-evaluated, should further research establish that the memory or perception account of our results is indeed more plausible.

What are the implications of these findings related to ambiguity avoidance?

Computational (Church & Patil, 1982) and psycholinguistic (Bever, 1970; Frazier, 1985) theories of language processing have traditionally assumed that the presence of ambiguity presents additional difficulty during sentence comprehension. Indeed, evidence from online sentence processing suggests that at least some ambiguities pose a processing difficulty: The so-called garden path sentences where ambiguities are initially misanalysed cause a temporary slow-down in processing (Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Trueswell, Tanenhaus, & Kello, 1993).

Somewhat surprisingly, however, a large body of work in sentence production has failed to find conclusive evidence that speakers avoid ambiguity that is rapidly resolved through contextual information and world knowledge (Arnold et al., 2004; V. Ferreira, 2003; Ferreira & Dell, 2000; Roland, Dick, & Elman, 2007; Roland et al., 2006; Wasow & Arnold, 2003). While some studies do find evidence for ambiguity avoidance (Haywood, Pickering, & Branigan, 2005; Jaeger, 2010; Roche, Dale, & Kreutz, 2010; Temperley, 2003), these cases are fairly rare and other factors such as informativity of the linguistic form (Jaeger, 2010) or constituent length (Wasow, 2002) appear to play a more important role in determining speakers' choices in online production.

It is possible that temporary ambiguities are not a good place to look for ambiguity avoidance since they do not lead to long-lasting processing delays and thus may not need to be avoided (cf. Jaeger, 2010). At the same time, languages do seem to avoid global systemic ambiguities that remain unresolved after context and world knowledge are taken into account (Wasow & Arnold, 2003). Specifically, one type of such ambiguity that is being avoided is ambiguity in grammatical function assignment (Wasow, to appear) addressed in this dissertation. Our results support this view. In all experiments presented here, learners were exposed to miniature artificial languages that contained a certain amount global systemic ambiguity since sentences were presented in isolation and case-marking, if present, occurred only on one constituent (either grammatical subject or object depending on the experiment). Thus sentences that did not contain case-marking were potentially globally ambiguous. After a short exposure to novel miniature artificial languages, learners showed a consistent preference to avoid ambiguity in grammatical function assignment that would have remained globally unresolved.

5.2. Future Directions

How can a community of speakers converge on a categorical grammar?

While learners in our experiments show consistent deviations from the input in the direction of more efficient linguistic systems, they typically do not introduce categorical changes into the newly acquired languages (i.e., 0% case-marking for the fixed or 100% case-marking for the random order language in Chapter 2 or, 100% case-marking for animate objects and 0% for inanimate objects in Chapter 4). This is not surprising since language change under normal circumstances is a fairly slow process involving multiple generations of speakers, and thus we expect learners' biases towards communicatively efficient systems to be fairly weak, only slightly shifting the grammar in the preferred direction. How can these weak biases then lead to a categorical language shift?

One influential account suggests that small changes introduced by one generation of learners are amplified as they are passed down to subsequent generations during the process termed cumulative cultural evolution (Christiansen & Chater, 2008; Kirby, 1999; Smith, Kirby, et al., 2003). Indeed, a number of iterative learning studies in the laboratory (Kirby et al., 2008; Reali & Griffiths, 2009; Smith & Wonnacott, 2010) and computer simulations (Kirby, 1999; Smith, Kirby, et al., 2003) suggest that changes introduced by one generation of learners are amplified as they percolate across subsequent generations, gradually causing the linguistic system to become categorical. These experiments, however, leave open two important questions: a) whether a language system would converge on a coherent grammar shared by the entire language community and b) whether the functional biases typically observed in single-generation experiments would survive cultural transmission across generations of learners in a fairly heterogeneous community. In the remainder of this section I will consider these two issues in more detail.

Most of existing iterative learning experiments typically make a simplifying assumption that the input received by the learner comes exclusively from one direct ancestor (Kirby et al., 2008; Reali & Griffiths, 2009; Smith & Wonnacott, 2010). Let us consider what outcomes would be predicted under this assumption for the languages produced by the learners in our experiments if they were to be passed down to new generations of learners. As discussed in Chapter 2, while the majority of learners in our experiments converge on output languages that are expected from the functional standpoint, there is considerable variability among learners with regard to which particular strategy they adopt. Thus, if each learner were to pass down the acquired language to exactly one learner in the subsequent generation (as is typical in iterative learning experiments), the system would eventually result in a heterogeneous language community. In this community, individual speakers or small groups of speakers would have a categorical grammar, but this grammar would not be shared by other individuals. This scenario is clearly inconsistent with natural language change.

The assumption that learners sample the input from only one ancestor is certainly unrealistic in language acquisition outside of the laboratory. Learners are typically exposed to a wide variety of sources including their parents, grandparents, peers, teachers, etc. and as a result receive a representative sample of the language spoken in the community. Even under this more plausible assumption, it is not trivial to predict how a heterogeneous population of speakers in our experiments would eventually converge on a variant shared by the linguistic community. Second-generation learners, who sample their input across a large number of first-generation learners, would receive a fairly heterogeneous sample as participants in our experiments vary widely in their use of grammatical devices.

How could learners derive their own language output from this heterogeneous input? It is possible that language change is primarily driven by the frequency of innovations in the community: The most frequent innovation is maintained in the learners' speech. This would eventually result in a categorical shift of the linguistic system towards using this innovation as the default form in the community. While theoretically possible, this scenario cannot fully account for natural language change since innovations are by definition highly infrequent at some point in history before they spread across the population.

A more plausible alternative is that learners do not sample the input uniformly, but instead employ a sampling strategy that is biased in some way, thus giving a higher weight to some innovations that are not necessarily highly frequent. Such bias is likely to stem from a variety of social factors that are not mutually exclusive (Heine, 2008; Heine & Kuteva, 2006; Nettle, 1999a, 1999b). For example, learners can assign higher weights to the input coming from individuals that have high social status within a community. The innovations introduced by more influential speakers would thus have a better chance of spreading across the community and eventually emerging as the dominant form. Indeed, computational simulations carried out by Nettle (Nettle, 1999a, 1999b) suggest that while functional biases do affect the direction of language change, they might not be strong enough to displace prior forms if language acquisition is not biased by social factors.

Other researchers have pointed out a special role of minority groups in the spread of innovations across the community (e.g., Levine, 1980; Moscovici, 1976). Since minority groups are typically uniform and consistent in their linguistic behavior, innovations have a better chance to become the norm within a small group first and then enter a larger linguistic community. Thus, according to Zipf (1949) words like *volt*, *gas*, *movies* were first shaped by functional pressures within certain groups of individuals and only later entered the lexicon of the larger community. For instance, *voltage* first underwent reduction to *volt* due to its high frequency in the speech of electricians and physicists, who later introduced this word into the general lexicon.

Linguistic communities are not uniformly connected and the structure of the social network also plays a role in the way innovations spread in the population (L. Milroy, 1980; L. Milroy & Milroy, 1978). In particular, non-standard linguistic forms are more likely to persist in communities with dense networks where many speakers know each other and are connected to each other in different capacities (e.g., friend, co-worker, etc.) since these communities are less susceptible to the influence from the outside pulling them towards the norm (L. Milroy, 1980). Thus, innovations that are picked up by central members of a tightly-knit group, who have strong ties to a lot of group members, tend to spread across the group fairly quickly (J. Milroy & Milroy, 1985; Rogers & Schoemaker, 1971).

Future studies are necessary to assess the influence of functional biases on language change during the process of cultural transmission as well as the role of social factors in this process, which so far have unfortunately been mostly studied in a separate literature.

Finally, there is another possibility that we have not considered so far. It is possible that the nature of the innovations introduced by child learners differs from that introduced by adults (e.g., Hudson Kam & Newport, 2005, 2009). In particular, it is possible that child learners introduce greater changes to the input than adult learners (i.e., changes that are more likely to create categorical shifts in the grammar) as a result of functional biases. We leave this hypothesis to be tested in future work.

When do biases for efficient information transfer develop?

How early do biases for efficient information transfer develop? Does one need substantial experience with language to show this preference or is this preference innate or develops very early in life? So far all participants in our experiments were adults, hence our findings do not speak to these questions. The next step in continuing this research program would be to explore young children's preferences in similar tasks to better understand the developmental timeline of these biases.

A large body of literature investigating speaker's preferences in natural language production has to date focused primarily on adult speaker behavior. We are unaware of any studies that directly address the question of communicative efficiency in child speech. However, some studies investigating cross-linguistic development of communicative strategies provide evidence consistent with a claim that preferences for efficient information transfer emerge early during development. In particular, MacWhinney and Bates (1978) found that increased givenness correlated with an increased use of pronominalization and ellipsis in picture descriptions by children acquiring languages as different as English, Italian, and Hungarian. These findings can be interpreted as an indication of an early preference for efficient information transfer: Since given constituents are more predictable, they are likely to undergo reduction in efficient systems. Similarly, Greenfield and Smith (1976) observed that referents containing new information were more likely to be verbally encoded as one-word utterances during the one-word production stage (when children communicate in single words such 'cup' or 'teddy').

These natural language data suggest that the amount of information contained in the linguistic signal may be weighted against production effort very early in development. Future studies are, however, necessary to determine whether this sensitivity goes beyond givennes and how this preference interacts with other biases during language acquisition.

Are processing biases in play during language acquisition?

It has been long argued that processing pressures, such as a preference to maximize the number of linguistic dependencies processed at each point in time, contribute to the way language structures change over time (cf. Hawkins's 'Maximize On-line Processing' principle).

In Chapters 2 and 4, we found a strong preference to use overt case-marking on the object when sentence constituent order was OSV. This pattern is consistent with two accounts: a communicative preference to mark an atypical (i.e., less expected) constituent order or a processing preference to put more informative cues earlier in the sentence. Experiment 2 in Chapter 4 provided preliminary evidence that both preferences are in play during language acquisition. Consistent with the processing account, learners of the subject-marking language initially used more case-marking when sentence constituent order was SOV, suggesting a bias to provide disambiguating cues early during parsing. This preference, however, got weaker as training continued, suggestive of a possible interaction with a bias to mark the atypical (i.e., OSV constituent order). One caveat in interpreting these results is that the languages used in Chapter 4 were not designed to study differential case-marker use depending on constituent order. As such, by design case-marking was not independent of constituent order in the input. Thus, while these findings are suggestive of processing influences in acquisition, they neither disentangle the two competing accounts nor directly probe whether processing biases operate at the same time course as biases for efficient communication.

A follow-up experiment would tease apart these two explanations by comparing two artificial languages that have flexible constituent order (subject-object-verb and object-subject-verb) and either subject or object marking that occurs independently of constituent order. If learning outcomes observed in this dissertation are due to the communicative pressure to overtly mark the atypical, we would expect learners to use more case-marking in OSV sentences in both languages. In contrast, if learners' behavior is primarily driven by a processing preference, we would expect learners to use more case-marking on the first constituent in both languages. Finally, if both preferences are additive as suggested by our preliminary data, we would expect a preference for case-marker use in OSV sentences to be stronger in the object-marking language since here both pressures work in the same direction.

Future work on processing preferences will also expand this research program to study a broader range of typological phenomena. A clear extension of this work would be testing dependency-length minimization accounts (Gibson, 1998; Hawkins, 2004) against behavioral data with a focus on whether learning outcomes follow the predicted cross-linguistic asymmetry in heavy constituent ordering depending on language headedness.

5.3. Conclusions

Summing up, the data presented in this dissertation suggest that there is a strong link between the process of language learning and the structure of natural languages as well as the process of language change. Work presented here provides evidence that biases operating during language acquisition closely align with the patterns observed in synchronic and diachronic typological data. More specifically, at least some of these biases can be traced back to a preference for efficient communicative systems. Moving beyond these findings, our ultimate goal is to understand the mechanism underlying these biases, the timeline of their development, as well as their interactions and how they are reflected in language structures.

Bibliography

Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Nat Lang Linguist Th, 21*(3), 435-483.

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition, 116*(1), 143-148.

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220*, 25-30.

Arnold, J., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding Attachment Ambiguities: The role of Constituent Ordering. *J Mem Lang, 55*(1), 55-70.

Atkinson, Q. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*(6027), 346-349.

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech, 47*(1), 31-56.

Aylett, M. P., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America, 119*, 3048.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: the State of the Art* (pp. 173-218). Cambridge: Cambridge University Press.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362): John Wiley & Sons.

Bickel, B. (2011). Statistical modeling of language universals. *Linguistic typology, 15*, 401-414.

Bickel, B., Witzlack-Makarevich, A., & Zakharko, T. (in press). Typological evidence against universal effects of referential scales on case alignment. In I. Bornkessel-Schlesewsky, A. Malchukov & M. Richards (Eds.), *Scales: A cross-disciplinary perspective on referential hierarchies*. Berlin: de Gruyter.

Blake, B. J. (2001). *Case*. Cambridge: CUP.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*(2), 119-137.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bolinger, D. (1972). *That's That*. The Hague: Mouton.

Bossong, G. (1985). Markierung von Aktantenfunctionen in Guarani. In F. Plank (Ed.), *Relational Typology* (pp. 1-30). Berlin: Mouton.

Bowern, C., Hunley, K., & Healy, M. (2012). Rejection of a serial fonder effects model in biology and linguistics. *Proceedings of the Royal Society B, 279*(1736), 2281-2288.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J Am Stat Assoc, 88*, 9-25.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the Dative Alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., Dingare, S., & Manning, C. D. (2001). Soft constraints mirror hard constraints: Voice and Person in English and Lummi. In M. Butt & T. Holloway King (Eds.), *Proceedings of the Lexical Functional Grammar Conference 2001 (LFG'01)*. Stanford: CSLI Publications.

Brooks, P., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*, 720-781.

Bybee, J. L. (2003). Mechanisms of change in grammaticization: The role of frequency. *The Handbook of Historical Linguistics. Oxford: Blackwell*, 602-623.

Bybee, J. L., & Hopper, P. (2001). Introduction to frequency and the emergence of linguistic structure. *Frequency and the emergence of linguistic structure*, 1–24.

Cangelosi, A., & Parisi, D. (1998). The emergence of 'language' in an evolving population of neural networks. *Connection Science, 10*(2), 83-97.

Cavalli-Sforza, L. (1997). Genes, people, and languages. *Proceedings of the National Academy of Sciences, 94*, 7719-7772.

Choi, Y., & Trueswell, J. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *J Exp Psychol, 106*(1), 41-61.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 119): The MIT press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.

Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In J. L. Dessalles & L. Ghadakpour (Eds.), *The Evolution of Language: 3rd International Conference* (pp. 45-48). Paris, France.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behav Brain Sci, 31*(05), 489-509.

Church, K., & Patil, L. (1982). Copig with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics, 8*, 139-149.

Cook, V. (1988). Language learners' extrapolation of word order in micro-artificial languages. *Language Learning, 38*, 497-529.

Crain, S., Goro, T., & Thornton, R. (2006). Language acquisition is language change. *Journal of Psycholinguistic Research, 35*, 31-49.

Croft, W. (2003). *Typology and Universals*. Cambridge: CUP.

Culbertson, J., & Legendre, G. (2010). *Investigating the evolution of agreement systems using an artificial language learning paradigm*.

Culbertson, J., & Newport, E. (2012). *The role of word order and contiguity in the grammaticalization of case*. Paper presented at the 86th Annual Meeting of the Linguistic Society of America, Portland, OR.

Culbertson, J., & Newport, E. (in prep). Learning biases for word order in children.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*, 306-329.

Cysouw, M. (2010). On the probability distribution of typological frequencies. In C. Ebert, G. Jäger & J. Michaelis (Eds.), *The mathematics of language* (pp. 29-35): Springer.

Cysouw, M. (2011). Understanding transitional probabilities. *Linguistic Typology, 15*, 415-431.

Cysouw, M., Dediu, D., & Moran, S. (2012). Comment on "Phonemic diversity supports a serial founder effect model of language expansion from Africa". *Science, 335*(6069), 657.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*: WW Norton & Company.

Dryer, M. (1989). Large linguistic areas and language sampling. *Studies in Language, 13*, 257-292.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, 81-138.

Dryer, M. (1998). *Why statistical universals are better than absolute universals.* Paper presented at the Proceedings of the 33rd annual meeting of the Chicago Linguistic Society, Chicago, IL.

Dryer, M. (2013). Prefixing vs. suffixing in inflectional morphology. In M. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dryer, M., & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Munich: Max Planck Digital Library.

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.

Evans, N. (1997). Role or cast? Noun incorporation and complex predicates in Mayali. In J. Bresnan & P. Sells (Eds.), *Complex predicates* (pp. 397-430). Stanford.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Piraha: Another look at the Design Features of human language. *Current Anthropology, 46*, 621-646.

Everett, D. L. (2009). Pirahã culture and grammar: A response to some criticisms. *Language, 85*, 405-442.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541-553.

Fedzechkina, M., Jaeger, T., & Newport, E. (2011). *Functional biases in language learning: Evidence from word order and case-marking interaction.* Paper presented at the CogSci 2011.

Fedzechkina, M., Jaeger, T., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proc Natl Acad Sci USA, 109*(44), 17897-17902.

Fedzechkina, M., Jaeger, T., & Newport, E. (2013). *Communicative biases shape structures of newly acquired languages.* Paper presented at the CogSci '13, Berlin, Germany.

Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society, 436-441.*

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164-203.

Ferreira, V. (2003). The persistence of optional complementizer production: Why saying “that” is not saying “that” at all. *Journal of Memory and Language, 48*, 379-398.

Ferreira, V., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychol, 40*(4), 296-340.

Ferreira, V., & Firato, C. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review, 9*(4), 795-800.

Ferrer i Cancho, R., Debowski, L., & Moscoso del Prado Martin, F. (2013). Constant conditional entropy and related hypotheses. *Journal of Statistical Mechanics, L07001.*

Ferrer i Cancho, R., & Díaz-Guilera, A. (2007). The global minima of the communicative energy of natural communication systems. *J Stat Mech*, P06009.

Finley, S. (2011). Generalization to novel constraints in artificial grammar learning. In L. Carlson, C. Hoeschler & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 318-323). Austin, TX: Cognitive Science Society.

Finley, S., & Badecker, W. (2008). Substantive biases for vowel harmony languages. In N. Abner & J. Bishop (Eds.), *27th West Coast Conference on Formal Linguistics* (pp. 168-176). Los Angeles, CA.

Fodor, J. D. (2001). Setting syntactic parameters. *The Handbook of contemporary syntactic theory*, 730-767.

Frank, A., & Jaeger, T. (2008). *Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production.* Paper presented at the The 30th Annual Meeting of the Cognitive Science Society Austin, TX.

Frank, M., Simmons, K., Yurovsky, D., & Pusiol, G. (2013). *Developmental and postural changes in children's visual access to faces*. Paper presented at the Proceedings of the 35th annual meeting of the cognitive science society, Berlin, Germany.

Frank, M., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PLoS one, 8*(1), e52500.

Frazier, L. (1985). Syntactic complexity. In D. R. Dowty, L. Karttunen & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129-189). Cambridge, UK: Cambridge Univ Press.

Friedman, V. (2006). Balkans as a linguistic area. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 1, pp. 657-672). Oxford: Elsevier.

Gennari, S. P., Mircovic, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: a cross-linguistic investigation. *Cognitive Psychology, 65*, 141-176.

Genzel, D., & Charniak, E. (2002). *Entropy rate constancy in text.* Paper presented at the Proc ACL-2002, Philadelphia.

Gerken, L. A. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, 67-74.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, P. Marantx & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-112). Cambridge, MA: MIT Press.

Gibson, E., Bergen, L., & Piantadosi, S. (2013). The rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, 110*(20), 8051-8056.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Sci, 34*(2), 286-310.

Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1005–1028). Hillsdale: Erlbaum.

Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntaz. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187-219). Amsterdam: John Benjamins.

Givón, T. (1991). Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Stud Lang, 15*(2), 335-370.

Goldberg, A. (1995). *A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Goldberg, A. (2013). Substantive learning bias or an effect of similarity? Comment on Culbertson, Smolensky, Legendre (2012). *Cognition, 127*(3), 420-426.

Goldin-Meadow, S., & Mylander, C. (1983). Gestural communication in deaf children: Noneffect of parental input on language development. *Science, 221*(4608), 372-374.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature, 391*(6664), 279-281.

Goldin-Meadow, S., Mylander, C., de Villiers, J., Bates, E., & Volterra, V. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. *Monographs of the Society for Research in Child Development*, 1-151.

Gomez Gallo, C., Jaeger, T., & Smyth, R. (2008). Incremental Syntactic Planning across Clauses *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci08)* (pp. 845-850).

Gordon, P., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 425-430.

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.

Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development*: Academic Press New York.

Griffiths, T., & Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cogn Sci, 31*, 441-480.

Grodner, D. J., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29*, 261-290.

Guy, R. (1996). Form and function in linguistic variation. In R. Guy, D. Feagin, D. Schiffrin & J. Baugh (Eds.), *Towards a social science of language: Papers in honot of William Labov* (pp. 221-252).

Haiman, J. (1983). Iconic and economic motivations. *Language, 56*, 515-540.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model *North American Chapter Of The Association For Computational Linguistics (NAACL)* (pp. 1-8). Morristown, NJ: Association for Computational Linguistics.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft, 18*(2), 180-205.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hawkins, J. A. (1983). *Word order universals*. New York, NY: Academic Press.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

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

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas Psychol, 25*, 87–107.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16*(5), 362-366.

Heine, B. (2008). Contact-induced word order change without word order change. In P. Siemund & N. Kintana (Eds.), *Language contact and contact languages*. Amsterdam: Benjamins.

Heine, B., & Kuteva, T. (2006). *The changing languages of Europe*. Oxford: OUP.

Hockett, C. (1958). *A course in modern linguistics*. New York: MacMillan.

Hockett, C. (1960). The origin of speech. *Scientific American, 203*, 88-111.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev, 1*(2), 151-195.

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychol, 59*(1), 30-66.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Lang Cognitive Proc, 24*(6), 876-909.

Hyams, N. (1983). *The pro drop parameter in child grammars.* Paper presented at the Proceedings of the West Coast Conference on Formal Linguistics.

Jackendoff, R. (1977). *X-Bar syntax: A study of phrase structure*. Cambridge: MIT Press.

Jaeger, T. (2006). *Redundancy and syntactic reduction in spontaneous speech.* (PhD dissertation), Stanford University, Stanford, CA.

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang, 59*, 434-446.

Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychol, 61*, 23-62.

Jaeger, T. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology, 4*(230).

Jaeger, T., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*.

Jaeger, T., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effects models for genetic and areal dependencies in linguistic typology: Commentary on Atkinson. *Linguistic Typology, 15*(2), 281-319.

Jaeger, T., & Tily, H. (2011). On language 'utility': processing complexity and communicative efficiency. *Wiley Interdiscip Rev: Cognitive Science, 2*(3), 323–335.

Karuza, E., Farmer, T., Fine, A. B., Smith, F., & Jaeger, T. (2014). *On-line measures of prediction in a self-paced statistical learning task*. Paper presented at the Proceedings of the 36th annual meeting of the Cognitive Science Society, Quebec City, Canada.

Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*: Oxford University Press, USA.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc Natl Acad Sci USA, 105*(31), 10681.

Kleinschmidt, D., & Jaeger, T. F. (under review). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review*.

Kornfilt, J. (2003). Scrambling, subscrambling and case in Turkish. In S. Karimi (Ed.), *Word order and Scrambling* (pp. 125-155). Malden, MA: Blackwell.

Kurumada, C., & Jaeger, T. (2013). *Communicatively efficient language production and case-marker omission in Japanese.* Paper presented at the CogSci '13, Berlin, Germany.

Kurumada, C., & Jaeger, T. (submitted). Communicative efficiency in language production: Optional case-marking in Japanese.

Labov, W. (1963). The social motivation of a sound change. *Word, 19*, 273-309.

Langacker, R. (2000). *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Lee, H. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean. *J East Asian Linguis, 15*, 69-96.

Lehnert, M. (1957). The interrelation between form and function in the development of the English language. *Zeitschrift für Anglistik und Amerikanistik, 5*, 43-56.

Levine, J. (1980). Reaction to opinion deviance in small groups. In P. Paulus (Ed.), *Psychology of group influence*. Hillsdale, NJ: Erlbaum.

Levy, R. (2005). *Probabilistic models of word order and syntactic discontinuity.* (Ph.D.), Stanford University, Stanford.

Levy, R. (2011). *Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results*.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106*(50), 21086-21090.

Levy, R., & Jaeger, T. F. (2007, December 4-7, 2006). *Speakers optimize information density through syntactic reduction.* Paper presented at the Advances in NIPS, Cambridge, MA.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*(1), 93-115.

Lewis, R. L., & Nakayama, M. (2001). Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 85-113). Stanford, CA: CLSI Publications.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*(10).

Lindblom, B. (1990). On the communication process: Speaker-listener interaction and the development of speech. *Augmentat Altern Commun, 6*(4), 220-230.

MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. *J Psycholinguist Res, 12*(5), 467-478.

MacWhinney, B., & Bates, E. (1978). Sentential Devices for Conveying Givenness and Newness: A Cross-Cultural Developmental Study. *Journal of Verbal Learning and Verbal Behavior, 17*, 539-558.

MacWhinney, B., & Bates, E. (1989). *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.

Magnuson, J., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word recognition and learning: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202-227.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua, 118*, 203-221.

Manin, D. (2006). Experiments on predictability of word in context and information rate in natural language. *J Inform Processes, 6*(3), 229-236.

Marchand, H. (1951). The syntactical change from inflectional to word order system and some effects of this change on the relation verb-object in English. *Anglia, 70*, 70-89.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. (1999). Rule-learning in seven-month-old infants. *Science, 283*, 77-80.

Maurits, L., Perfors, A., & Navarro, D. (2010). Why are some word orders more common than others? A uniform information density account. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 23* (pp. 1585-1593). Vancouver, BC.

Merlan, F. (1982). *Mangarayi* (Vol. 4). Amsterdam: North-Holland.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network, speaker innovation. *Journal of Linguistics, 21*(2), 339-384.

Milroy, L. (1980). *Language and social networks*. Oxford: Blackwell.

Milroy, L., & Milroy, J. (1978). Belfast: change and variation in an urban vernacular. In P. Trudgill (Ed.), *Sociolinguistic patterns in British English*.

Mohanan, T. (1994). *Argument Structure in Hindi*. Stanford: CSLI.

Morgan, J. L., Meier, R. P., & Newport, E. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychol, 19*(4), 498-550. doi: 10.1016/0010-0285(87)90017-x

Morolong, M., & Hyman, L. (1977). Animacy, objects and clitics in Sesotho. *Studies in African Linguistics, 8*, 199-218.

Moscovici, S. (1976). *Social influence and social change*. London: Academic Press.

Mosel, U., & Hovdhaugen, E. (1992). *Samoan reference grammar*. Oslo: Scandinavian University Press.

Nettle, D. (1999a). Is the rate of linguistic change constant? *Lingua, 108*, 119-136.

Nettle, D. (1999b). Using Social Impact Theory to simulate language change. *Lingua, 108*, 95-117.

Newport, E. (1981a). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Lang Scie, 10*, 147-172.

Newport, E. (1981b). Constraints on structure: Evidence from American Sign Language and language learning. In W. A. Collins (Ed.), *Aspects of the development of competence. Minnesota Symposium on Child Psychology* (Vol. 14, pp. 93-124): Erlbaum.

Newport, E. (1982). Task-specificity in language learning? Evidence from speech perception and American Sign Language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: CUP.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

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.

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357.

Nyogi, P., & Berwick, R. (1997). Evolutionary consequences of language learning. *Linguistics and Philosophy, 20*, 697-719.

Ohala, J. (1989). Discussion of Lindblom's 'Phonetic invariance and the adaptive nature of speech'. In B. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 175-183). London: Academic Press.

Pajak, B. (2010). *Perceptual advantage from generalized linguistic knowledge.* Paper presented at the Proc of the 33rd Annual Conference of the Cognitive Science Society.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*, 306-338.

Piantadosi, S., Tily, H., & Gibson, E. (2011a). The communicative function of ambiguity in language. *Cognition, 122*, 280-291.

Piantadosi, S., Tily, H., & Gibson, E. (2011b). Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA, 108*(9), 3526.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology, 7*, 101-139.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pusiol, G., Soriano, L., Fei-Fei, L., & Frank, M. (2014). *Discovering the signatures of joint attention in child-caregiver interaction*. Paper presented at the Proceedings of the 36th annual meeting of the Cognitive Science Society, Quebec City, Canada.

Putnam, H. (1971). The 'innateness hypothesis' and explanatory models in linguistics. In J. Searle (Ed.), *The philosophy of language*. London: Oxford University Press.

Qian, T., & Jaeger, T. F. (2012). Cue Effectiveness in Communicatively Efficient Discourse Production. *Cognitive Sci, 36*, 1312-1336.

Rafferty, A., Griffiths, T., & Klein, D. (in press). Analyzing the rate at which languages lose the influence of a common ancestor. *Cognitive Science*.

Reali, F., & Griffiths, T. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition, 111*(3), 317-328.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior, 6*, 855-863.

Reeder, P., Newport, E., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*(30-54).

Roche, J., Dale, R., & Kreutz, R. (2010). *The resolution of ambiguity during conversation: More than mere mimicry?* Paper presented at the Proceedings of the 31st Annual Conference of the Cognitive Science Society, Austin, TX.

Rogers, E., & Schoemaker, F. (1971). *Communication of innovations*. New York: Free Press.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*(3), 348-379.

Roland, D., Elman, J. L., & Ferreira, V. (2006). Why is *that?* Structual prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition, 98*(3), 245-272.

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., . . . Gorniak, P. (2006). The human speechome project *Proceedings of the 28th Annual Cognitive Science Conference*.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926.

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

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science, 12*(4), 323-328.

Senghas, A., Coppola, M., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: The emergence of grammatical devices. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University conference on language development* (Vol. 2, pp. 550-561). Somerville, MA: Cascadillia Press.

Shannon, C. (1948). A mathematical theory of communications. *Bell Syst Tech J, 27*(4), 623-656.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112-171). Canberra: Australian Institute of Aboriginal Studies.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49*(4), 370-407.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, Winston.

Slobin, D. (1977). Language change in childhood and in history. In J. Macnamara (Ed.), *Language Learning and Thought* (pp. 185-214). New York: Acad. Press.

Slobin, D., & Bever, T. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12*(2), 229-265.

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems, 6*(4), 537-558.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial life, 9*(4), 371-386.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*(3), 444-449.

Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical biases and referential scenes in child and adult sentence processing. *Cognit Psychol, 49*(3), 238-299.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*(2), 276–309.

St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*(7), 1317-1329.

Steels, L. (1995). A self-organizing spation vocabulary. *Artificial Life, 2*(3), 319-332.

Steels, L. (2006). *How to Do Experiments in Artificial Language and Why.* Paper presented at the Proceedings of the 6th International Conference on The Evolution of Language (EVOLANG6), London.

Szmrecsanyi, B. (2004). *On operationalizing syntactic complexity.* Paper presented at the JADT, Louvain-la-Neuve, Belgium.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language, 73*(3), 464-484.

Thompson, S., & Newport, E. (2007). Statistical learning of syntax: The role of transitional probability. *Learning and Development, 3*, 1-42.

Tily, H. (2010). *The role of processing complexity in word order variation and change.* (PhD), Stanford University.

Tily, H., Frank, M., & Jaeger, T. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hoelscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1364-1369). Boston, MA: Cognitive Science Society.

Tily, H., & Jaeger, T. (2011). Complementing quantitative typology with behavioral approaches: Evidence for typological universals. *Linguistic Typology, 15*(2), 497-508.

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

Trueswell, J., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*(3), 285-316.

van Son, R. J. J. H., & Pols, L. C. W. (2003). *How efficient is speech?* Paper presented at the Proc of the Institute of Phonetic Sciences, Amsterdam.

van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for effcient communication. *Speech Commun, 47*, 464-484.

Wasow, T. (2002). *Postverbal Behavior*. Stanford: CSLI Publications.

Wasow, T. (to appear). Ambiguity avoidance is overrated.

Wasow, T., & Arnold, J. (2003). Post-verbal constituent ordering in English. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English* (pp. 119-154): Walter de Gruyter.

Wei, K., & Kording, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Front Comput Neurosci, 4*(11).

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science, 30*(5), 945-982.

Wonnacott, E., Newport, E., & Tanenhaus, M. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychol, 56*(3), 165-209.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.