Title
The Representation of Structure in Language: When is there more than meets the eye?

Author
Lau, Sin Hang

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Representation of Structure in Language: When is there more than meets the eye?

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Experimental Psychology

by

Sin Hang Lau

Committee in charge:

Professor Victor Ferreira, Chair
Professor Leon Bergen
Professor Judith Fan
Professor Andrew Kehler
Professor Edward Vul

2023

The Dissertation of Sin Hang Lau is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

I would like to dedicate this dissertation to my family and friends who were always there to celebrate my achievements and supported me through many challenges in my career and personal life.

I would also like to dedicate this dissertation to my teachers and mentors, past and present, without whom I would have never experienced the immense joy of learning and sharing knowledge with others.

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE iii

DEDICATION iv

TABLE OF CONTENTS v

LIST OF FIGURES vi

LIST OF TABLES vii

ACKNOWLEDGEMENTS viii

VITA ix

ABSTRACT OF THE DISSERTATION x

INTRODUCTION 1

CHAPTER 1 8

CHAPTER 2 74

CHAPTER 3 117

CONCLUSION 154

LIST OF FIGURES

Figure 1.1. Example stimuli in the first learning phase. 21

Figure 1.2. Example stimuli in the second learning phase (L-R: “give”, “hand”, and “show”)... 22

Figure 1.3. Examples of the final frame of each event description in the third learning phase.... 23

Figure 1.4. Proportion of responses produced by each exposure group with the three verbs in the picture description task 29

Figure 1.5. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task 32

Figure 1.6. Proportion of responses produced by each exposure group with the three verbs in the picture description task 36

Figure 1.7. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task 39

Figure 1.8. Proportion of responses produced by each exposure group with the three verbs in the picture description task 43

Figure 1.9. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task 45

Figure 1.10. Experiment 1 Results Broken Down by Prompt 71

Figure 2.1. Proportion of high attachment sentence continuations following high- or low-attachment morphological or sentence primes 87

Figure 2.2. Mean rating (in z-score unit) for ambiguous noun phrases following high- or low-attachment morphological or sentence primes 92

Figure 2.3. Mean rating (in z-score unit) for ambiguous target items following high- or low-attachment primes 97

Figure 2.4. Mean rating (in z-score unit) for ambiguous target items following high- or low-attachment primes 98

Figure 3.1 Average production times for trials with one tone only, two tones that covaried with both C and V, C alone, V alone, and neither 135

Figure 3.2 Average production times for trials for various CV repetition patterns that required reattaching versus not 136

Figure 3.3 Average production times for the eight conditions of CVs, Tones, and Repetition Patterns 143

LIST OF TABLES

Table 1.1 Summary and Example Sentences of the Exposure in the Third Learning Phase by Group 23

Table 1.2. Summary and Example Sentences of the Exposure in the Learning Phase by Group. 34

Table 1.3. Exposure Phase Materials 61

Table 1.4. Test Phase Picture Description Task Materials 62

Table 1.5. Test Phase Acceptability Judgment Task Materials 63

Table 1.6. Number of excluded production responses in each condition 65

Table 1.7. Detailed Summary of Production Responses Across Experiments 66

Table 1.8. Full Output of Models for the Log-odds of Producing a Given Structure in the Picture Description Task 67

Table 1.9. Full Output of Models for Acceptability of a Given Structure in the Acceptability Judgment Task 69

Table 2.A1. Prime Materials 107

Table 2.A2. Target Materials 110

Table 2.B1 Prime Materials 113

Table 2.B2 Target Materials 113

Table 2.C1 Prime Materials 114

Table 2.C2 Target Materials 115

Table 3.1 Sounds in Each Phoneme Set 127

Table 3.2 Example trials of the 27 Conditions from Phoneme Set 1 128

Table 3.3 Average Production Times of the 27 Conditions and Marginal Means for Each C, V, and Tone Repetition Patterns 132

Table 3.4 Sounds in Each Phoneme Set 138

Table 3.5 Example trials of the 8 Conditions from Phoneme Set 1 139

Table 3.6 Example trials of the 8 Conditions from Phoneme Set 1 142

ACKNOWLEDGEMENTS

I would like to thank Victor Ferreira, my advisor for the last five and a half years. His unwavering support for my research ambitions and endeavors beyond my academic work was invaluable to my growth as a researcher. I would also like to thank the rest of my committee for their insightful comments.

I would like to acknowledge the members of the Language Production Lab, past and present, whose feedback, collaboration, and support contributed tremendously to my work. Their passion for research rejuvenated me when I felt discouraged, and their creative ideas inspired me always. It has truly been an honor and privilege to be a member of this lab.

Finally, I would like to thank my family for their sacrifices, both big and small, which enabled me to pursue my wildest dreams.

Chapter 1, in full, is a reprint of the material as it appears in Cognition, 2021, Lau, Sin Hang; Momma, Shota; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material currently being prepared for submission for publication. Lau, Sin Hang; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material currently being prepared for submission for publication. Lau, Sin Hang; Li, Chuchu; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

VITA

2016 Bachelor of Arts in General Linguistics, The University of Hong Kong

2020 Master of Arts in Experimental Psychology, University of California San Diego

2023 Doctor of Philosophy in Experimental Psychology, University of California San Diego

PUBLICATIONS

Lau, S. H., Li, C., & Ferreira, V. S. (in prep). Lexical tone is different and special: Evidence from a speeded repeated production task.

Lau, S. H. & Ferreira, V. S. (in prep). Do people structure words and sentences using shared mechanisms?

Lau, S. H., Momma, S., & Ferreira, V. S. (2021). Learning structural alternations: What guides learners' generalization?. *Cognition, 215*, 104828.

Lau, S. H., Huang, Y., Ferreira, V. S., & Vul, E. (2019). Perceptual features predict word frequency asymmetry across modalities. *Attention, Perception, & Psychophysics, 81*(4), 1076–1087

Lau, S.H., & Hwang, H. (2016). The effects of frequency on pronoun production. *Journal of Cognitive Science, 17*(4), 547–569.

AWARDS

J. Frank Yates Student Conference Award, Psychonomic Society, 2020

Runner-up for the Student Poster Award, the International Workshop on Language Production, 2018

ABSTRACT OF THE DISSERTATION

The Representation of Structure in Language: When is there more than meets the eye?

by

Sin Hang Lau

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2023

Professor Victor Ferreira, Chair

Linguistic structures that appear to be different on the surface may be linked. This dissertation contains three sets of studies that investigate the cognitive mechanisms behind the relationships among linguistic structures on the syntactic, morphological, and phonological levels. Chapter 1 examines how speakers learn to generalize word orders in sentences in three artificial language learning experiments. Results suggested that learners have linguistic biases

that mirror typological differences, which help them go beyond simple statistics tracking. Chapter 2 investigates whether structuring mechanisms are shared across linguistic units of different grain sizes. In three structural priming experiments, we tested whether priming of attachment preferences occurs between words and sentences. Results showed that priming only occurred within- but not across-grain size, suggesting that structuring mechanisms for words and sentences are not shared. Chapter 3 explores phonological representation in the production of tonal languages. In two speeded repeated production experiments, we confirmed that Mandarin Chinese speakers use syllables (rather than segments such as consonants and vowels) as basic planning units. Additionally, we discovered that lexical tone is special in phonological processing, in that it is integrated with syllables relatively late in processing and in ways that are different from how segments in non-tonal languages are represented and combined.

INTRODUCTION

Language contains many kinds of structures that appear to be very different on the surface. For example, sentences that convey similar messages can have different word orders, hierarchical structures in words and sentences contain building blocks of different grain sizes, and segments (consonants and vowels) and tone contribute to different aspects of sounds in speech. Yet, there are distributional patterns and experimental evidence that suggest some of these structures may be linked cognitively, in that they may have originated from some latent cognitive biases, or have shared representation and processing mechanisms. This dissertation aims to find these seemingly distinct but potentially linked linguistic structures and examine the cognitive factors that explain possible relationships between them.

On the sentence level, very similar messages can be expressed with grammatical structures of different word orders (i.e., the order of the subject, verb, and object). This flexibility arises through what are termed structural alternations. One example is the dative alternation in English. Even though word orders in English are predominantly SVO (subject-verb-object), speakers can sometimes choose between the prepositional dative structure (e.g., “the girl gave the book to the boy) and the double object structure (e.g., “the girl gave the boy the book”). Importantly, structural alternations can be observed both within and across languages with different degrees of flexibilities and constraints (see Goldberg, 2011 for a corpus study on the distributional pattern of datives in English and Whong-Barr & Schwartz, 2002 for a cross-linguistic comparison on the differences in constraints on dative structures).

Generally, languages of the world are categorized into fixed versus free word order languages. There are two important typological differences between these languages in regards to structural alternations: First, fixed word order languages (e.g., English) tend not to allow subject-crossing alternations (i.e., where the subject intervenes between the verb and the object(s), also termed “scrambling”; Ross, 1967), whereas free word order languages do (e.g., “the book the boy the girl gave” is acceptable in Korean, a predominantly SOV language). Second, structural alternations are lexically-constrained in fixed word order languages (i.e., both dative structures are acceptable for verbs like “give”, but only the prepositional dative structure is acceptable for verbs like “donate”), but not in free word order languages (i.e., both dative structures are permitted across all verbs in Korean). From a language-learning perspective, learners are faced with the challenge of figuring out which type of alternations they are encountering, and to condition their generalization accordingly. This raises the question of whether these two seemingly distinct typological features are linked. Specifically, does exposure to subject-crossing cue learners to not be verb-wise conservative? And if so, what explains their relationship?

In Chapter 1, we investigate whether learners indeed condition their generalization according to typologically relevant features when learning a novel language; and if they do, what guides their generalization. In three artificial language learning experiments, participants learned an English-Korean hybrid language, then performed a picture description task (production) and an acceptability rating task (comprehension). In the learning phase, all participants saw alternations with one verb (“give”), while they only saw the canonical structure with no other alternations with two other verbs (“hand” and “show”). Critically, one group of participants had no exposure to subject-crossing alternations, whereas the other group did. Results showed that participants who were only exposed to non-subject-crossing alternations generalized in a verb-wise conservative manner, such that they rated the alternations they saw as more acceptable with “give” than with “hand” and “show”. In contrast, participants who were exposed to subject-crossing alternations accepted the alternations they saw about equally across all verbs. Production data trended in the same direction. These patterns mirror typological differences between languages with relatively fixed versus free word orders, suggesting that typological features may be linked by a latent linguistic bias that guides learners' generalization beyond simple statistics tracking. Specifically, we speculated that there is a cognitive preference for a clear subject-predicate distinction (i.e., keeping the verb and the object(s) together, separate from the subject). When this preference is violated, learners may attribute the driving force of alternations to factors other than verb argument structure, such as pragmatics, and are thus less likely to constrain their generalization to a certain verb.

Structural alternations demonstrate that similar messages can be conveyed by linguistic structures that are sequenced very differently on the surface, but sometimes the opposite happens in language – the same sequence of linguistic units can be analyzed in different ways, leading to different meanings. For example, the relative clause in the sentence "I met the students of the teacher who played the violin" can be analyzed to have a high attachment (HA; i.e., the students played the violin) or a low attachment reading (LA; i.e., the teacher played the violin). Likewise, on the morphological level, "social psychologist" can also be analyzed to have an HA (i.e., someone who studies social psychology; [[social psycholog(y)][ist]]) or LA organization (i.e., a psychologist who is social; [[social][[psycholog(y)][ist]]]). It appears that there are some similarities between attachment structures in words and sentences, but the linguistic units involved vary in grain size. This leads to the question of whether structuring mechanisms are shared across linguistic levels.

Chapter 2 examines whether people structure words and sentences using shared mechanisms with a structural priming paradigm, which is a methodology that has been widely used to establish evidence for shared structural representation within the linguistic domain (e.g., Desmet & Declercq, 2006, Scheepers, 2003), as well as shared structuring mechanisms across domains (see Van de Cavey & Hartsuiker, 2016 for shared mechanisms between music, math, and language). Experiment 1 tested whether attachment manipulation in sentence production and free recall of noun phrases affected subsequent sentence production attachment preferences; Experiment 2 tested whether sentence production and noun phrase comprehension affected subsequent noun phrase comprehension; and finally, Experiment 3 tested whether sentence and noun phrase comprehension affected subsequent comprehension on both levels. Overall, we observed within-level but not across-level priming. That is, attachment preferences in sentences only affected later preferences in sentences but not words, and vice versa, suggesting that the structuring mechanisms are not shared between linguistic units of different grain sizes. We speculate that the mechanisms are only shared across different cognitive domains in very abstract and general terms, and that these mechanisms (at least as reflected by structural priming) are sensitive to differences in online processing dynamics across different linguistic levels. In particular, the meaning of a noun phrase may not be computed by combining each of the morphemes on-the-fly, whereas the attachment site of a relative clause in a sentence is more likely to be determined online.

Differences in structures are not only found in comparisons across different linguistic levels, they can also be found in cross-linguistic comparisons within the same linguistic level. On the phonological level, a distinctive feature of tonal languages is that the same sequence of segments (i.e., consonants and vowels) can have different meanings depending on which pitch variation (i.e., tone) it is produced with. For example, in Mandarin Chinese, “ma” could mean “mother”, “hemp”, “horse”, or “to scold”, depending on whether it is produced with a high and flat, rising, falling then rising, or falling pitch, respectively. However, how tone is represented in speech planning (i.e., whether it is attached to the representation of the vowel or the syllable as a whole, or represented independently), as well as the timing at which it is integrated in phonological encoding is not well-understood.

Chapter 3 first explores what the basic speech planning units are in Mandarin Chinese, then examines how these units are integrated in phonological encoding. We reported two speeded repeated production experiments, in which participants produced four-word sequences out loud as many times as possible within eight seconds. Importantly, we manipulated the sequences such that the potential units involved in speech planning repeated at different frequencies across trials. The average production time per word was measured and interpreted as an indicator of production difficulty, with the initial assumption that reusing the same basic planning unit should generally facilitate production. The results confirmed existing findings in the literature (e.g., Chen et al., 2002; O'Seaghdha et al., 2010), showing that the basic planning units in Mandarin are syllables instead of segments (as in non-tonal languages like English). Moreover, we found that the number of unique planning units involved in a sequence did not reliably predict speech rate (i.e., about equal speech rate for ba2 di1 da1 bi2 and ba1 ba1 ba1 ba1). Rather, we reported the novel finding that what reliably predicted speech rate instead was how the syllable(s) and tone(s) in a sequence were combined. Specifically, speech rate was significantly slower in sequences where one syllable was paired with more than one tone (e.g., ba1 ba2 ba1 ba2). We proposed a reattachment hypothesis, which asserts that the process of detaching a tone from a syllable and reattaching another tone to it is a costly process that slows production. This hypothesis suggests that tone is integrated with segments relatively late in phonological processing.

Overall, the goal of this dissertation is to find structures that may be implicitly linked and investigate the relationship between such structures, in order to shed light on the cognitive mechanisms behind language. Across eight experiments, we show that typological patterns on the sentence levels are linked by a cognitive bias regarding subject-predicate structure (Chapter 1), structuring mechanisms across words and sentences are not shared due to potential differences in online processing dynamics (Chapter 2), and tone is represented and processed differently from segments (Chapter 3). Altogether, these results further our understanding of linguistic structures on the syntactic, morphological, and phonological level.

References

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*(4), 751-781.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language, 54*(4), 610-632.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131-153.

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282-302.

Ross, J.R., 1967. *Constraints on variables in syntax*. PhD Dissertation. MIT, Cambridge, MA

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition, 89*(3), 179-205.

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition, 146*, 172-184.

Whong-Barr, M., & Schwartz, B. D. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, *24*, 579–616.

CHAPTER 1

Learning Structural Alternations:

What guides learners' generalization?

Sin Hang Lau, Shota Momma, Victor S. Ferreira

Author Note

Sin Hang Lau, Department of Psychology, University of California, San Diego; Shota Momma, Department of Linguistics, University of Massachusetts Amherst; Victor S. Ferreira, Department of Psychology, University of California, San Diego

Correspondence concerning this article should be addressed to Sin Hang Lau, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: aubreylau@ucsd.edu

Abstract

Some word-order alternations observed across the world's languages are constrained by specific verb choice, whereas one type of word-order alternation (i.e., scrambling) frequently seen in free word order languages is not lexically-dependent on the verb. Three novel-language learning experiments explore whether speakers latently respect this generalization. If learners show conservativeness that closely reflects statistics from the input, then it would support usage-based and statistical accounts; alternatively, if learners have linguistic biases that allow them to generalize beyond statistics and show generalization similar to typological patterns, then it would support an internal bias account. In each of the three experiments, two groups of English monolinguals learned a Korean-English hybrid language with structural alternations analogous to those found in different categories of natural languages, as defined by whether the language allows scrambling and whether alternations are lexically-dependent on the verb. Learners' generalization patterns in subsequent picture description and acceptability judgment tasks were analyzed. Comprehension data consistently showed that the group which learned alternations found in natural languages with relatively rigid word orders tended to be more verb-wise conservative than the group that learned alternations found in languages with relatively free word orders. Production data trended in the same direction as the comprehension data. Thus, our results suggest that learners have linguistic biases that mirror typological differences that help learners go beyond simple statistics tracking.

Keywords: structural alternation, learning, verb argument structure, sentence production, artificial language

Learning Structural Alternations:

What guides learners' generalization?

Languages of the world show great flexibility in how they convey intended messages. Specifically, very similar messages can be expressed by sentences with different word orders (i.e., the order of the subject, verb, and object in a sentence), both across and within languages. When multiple grammatical structures can describe the same event, they are referred to as *structural alternations*. However, there are limitations on such structural flexibility. For example, English allows flexibility in word orders only with certain verbs. Specifically, even though English speakers can alternate between "the girl gave the book to the boy" and "the girl gave the boy the book" to convey similar messages, this structural alternation is not possible with verbs like "donate": "The girl donated the museum the book" is not acceptable to many English speakers. Hence, language learners, regardless of whether they are infants learning their first language, adults learning a second language, or even adults learning to use a new verb in their native language must face the challenge of figuring out what structures are permitted under what context and make generalizations based on limited input. How, then, do learners learn to use structural alternations appropriately? What guides their generalization? We first explore how languages differ, then investigate two classes of learning hypotheses that make different predictions on how learners make generalizations about structural alternations, focusing on structural alternations involving word-order changes.

The differences in word order across languages not only impact the surface order of sentence elements, but also have implications for what linguistic features learners can depend on to infer meaning (Bates & MacWhinney, 1989). In general, languages can be categorized into relatively *fixed* and *free word order* languages. English is often described as a prototypical fixed word order language, because changing word orders results in either unacceptability or a clear change in meaning. For instance, “Mary John pushed” is not an acceptable English sentence (without strong prosodic marking). Speyer’s (2010) diachronic analysis of Old, Middle, and Modern English showed a decline in object topicalization (e.g., “Mary John pushed”), which is argued to be a possible consequence of English word order becoming more rigid over time. Meanwhile, “Mary pushed John” and “John pushed Mary” convey two clearly different meanings. Even though this rigidity of English may seem inconvenient and challenging to learners, it also allows them to reliably use word order as a cue to interpret meaning: Because “Mary” comes before the verb “pushed” in an active sentence, she must be the agent of the action; “John” comes after the verb, so he must be the undergoer of the action. Another key constraint on structural alternations in relatively fixed word order languages was demonstrated earlier with the prepositional dative structure (e.g., “the girl gave the book to the boy”) and the double object structure (e.g., “The girl gave the boy the book”). Although this alternation is acceptable for many verbs (e.g., “give”, “hand” and “show”), only the prepositional dative option is permitted by other verbs (e.g., “donate”, “submit”, and “transfer”). Also, even for verbs that allow both structures, the use of the two forms of dative is not evenly distributed. Specifically, English corpus evidence shows that verbs that allow alternations behave drastically differently than those that do not. Given a discourse context that is suitable for both forms of datives in English, the probability of observing an alternating verb with the prepositional dative is .04 on average, compared to .83 in non-alternating verbs (Goldberg, 2011). In other words, dative alternations in English are *lexically-dependent* on the verb. We refer to dative alternations that switch the order of the direct and indirect objects as *local alternations*, to reflect that the alternating elements inside the verb phrase (the verb and the objects) stay together and are separate from the subject in these word orders.

Relatively free word order languages, on the other hand, tend to have much fewer structural constraints. Korean, for example, is a language that has the predominant word order of subject-object-verb (SOV). However, object-subject-verb (OSV) is also permitted. Even though the surface orders are different, the propositional meaning and truth conditions of the sentence remain unchanged. The phenomenon of having structural alternations that separate elements inside the verb phrase (i.e., where the object(s) separate from the verb), which we refer to as *subject-crossing alternations* (because an object "crosses" the subject, such that the subject intervenes linearly between that object and the verb), is an instance of what is termed "scrambling" (Ross, 1967; see Yamashita, 2002 for a review of the linguistic properties of scrambling in Japanese). Critically, scrambling alternations are not constrained by particular verbs. Likewise, local dative alternations which English permits on a lexically-constrained basis are also not constrained by verbs in Korean (see Whong-Barr & Schwartz, 2002 for a cross-linguistic comparison of English, Japanese, and Korean datives, showing more semantic constraints in English datives than in Korean). In contrast to English, Korean allows much more structural flexibility, with the result that learners cannot reliably infer propositional meaning from word order. Instead, Korean has *case markers*, which are grammatical components that are attached to and undergo movement with nouns for the purpose of indicating their roles: The *nominative* case marker is attached to the subject "the boy", the *accusative* marker to the direct object "the book", and the *dative* marker to the indirect object "the girl", such that the comprehender can easily construct the meaning of "the boy gave the book to the girl" regardless of surface word order.

We have highlighted two important typological differences relevant to the current work: First, relatively fixed word order languages like English tend not to allow subject-crossing dative alternations, whereas free word order languages like Korean tend to allow them. Second, word order alternations seen in relatively fixed word order languages like English are lexically constrained, but word order alternations due to scrambling seen in relatively free word order languages, such as subject-crossing alternations, are not lexically constrained. From the language learning perspective, the challenges for learners are to figure out which type of alternations they are encountering and to condition their generalizations accordingly.

One possible strategy for learning and generalizing correctly would be to start conservatively on an item-by-item basis, then generalize later. This strategy is similar to what has been proposed under *usage-based accounts* (Tomasello, 2000, 2003). For example, a learner of English very rarely sees a subject-crossing alternation such as "The book the boy gave to the girl", and some alternations they do see are verb-wise restricted ("The boy donated the book to the girl, but "*The boy donated the girl the book"). Based on these data, they are unlikely to produce subject-crossing alternations but are likely to infer that alternations in English are lexically-constrained. At the same time, because a learner of Korean sees both local and subject-crossing alternations with many different kinds of verbs, they are more likely to infer that the alternations they see will not be verb-wise restricted.

Evidence for usage-based accounts comes from developmental findings suggesting that children largely restrict their use of a structure to the verb context that they learned it in. Brooks and Tomasello (1999) showed that when children just under the age of three years learned a novel verb in an active sentence and another in a passive sentence, only a small number of children were able to use the latter verb in an active sentence. Similar conservative patterns were also reported in a different pair of sentence frames (e.g., transitive or intransitive sentence frames; Tomasello & Brooks, 1998), as well as in non-experimental paradigms such as observational studies (Braine & Bowerman, 1976) and corpus analyses (Goldberg, Casenhiser, & Sethuraman, 2004).

The opposite strategy still within a usage-based-type of account would be to first generalize all structural alternations to all verbs, until there is direct negative evidence that suggests the learner should do otherwise. For example, a learner of English may see an alternation with a small subset of verbs ("The boy gave/handed/showed the book to the girl" vs. "The boy gave/handed/showed the girl the book"), and so they first assume that the same alternation will be allowed by all verbs until they are corrected. Because it is not the case that all verbs in English allow such an alternation ("The boy donated the book to the girl", but "*The boy donated the girl the book"), and that the learner will often over-generate, it is possible that they will be given negative feedback and thus learn not to use that alternation with that particular verb. On the other hand, because alternations in Korean are indeed not lexically-dependent on the verb, even if a learner of Korean first assumes that the alternations they see will not be verb-wise restricted, they will not receive negative feedback and will thus continue to generalize liberally. However, this is an improbable strategy, as learners seldom receive the systematic feedback that this strategy requires — this is an instance of the lack of negative feedback problem (Brooks & Tomasello, 1999). In the case of dative alternations in English, children rarely overgeneralize the double object dative structure to verbs like "donate", and if they do, they are often not corrected (Baker, 1979).

Addressing the lack of negative feedback problem, statistical accounts of language learning modify the usage-based account to argue that learners have sophisticated ways of tracking statistics. Specifically, learners are able to infer indirect negative evidence by tracking the frequency of different types of positive evidence alone (Perfors, Tenenbaum, & Wonnacott, 2010; Stefanowitsch, 2008; Wonnacott, Brown, & Nation, 2017; Wonnacott, Newport, & Tanenhaus, 2008). In a series of artificial language experiments, Wonnacott et al. (2008) exposed learners to twelve novel verbs and two constructions, varying how many verbs allowed alternations, as well as the frequency of the verbs. The results suggested that learners keep track of both verb-specific (i.e., how often a given structure occurs with a specific verb) and verb-general statistics (i.e., how often a structure occurs across different verbs). With these two competing statistics, learners keep track of two additional statistics that influence to what extent verb-specific and verb-general statistics are used. First, verb frequency affects the degree to which learners pay attention to verb-specific statistics, in that they are more likely to ignore verb-specific statistics of low-frequency verbs. Second, learners keep track of language-global data (i.e., how many verbs allow alternation relative to how many verbs do not) and are more likely to ignore verb-specific statistics if many verbs alternate in the language. If learners indeed keep track of these statistics, the data we quoted from Goldberg (2011) can clearly signal to learners that English is lexically-conditioned. Related to the idea of language-global data in Wonnacott et al. (2008), work by Thothathiri and colleagues discussed a similar idea in terms of cue validity (Thothathiri & Braiuca, 2020; Thothathiri & Rattinger, 2016). For example, in learning structural alternations, learners of a language with many verb-specific biases (i.e., verbs are highly predictive cues) show a greater degree of conservatism, compared to learners of a language in which more verbs appear in all possible constructions (i.e., verbs are not predictive cues). Additionally, there is evidence that learners are able to use high-level pragmatic reasoning in statistical learning. Perek and Goldberg (2015, 2017) highlight the influence of meanings and discourse functions of constructions on how they are used and generalized. For example, the double object dative is more common when the goal is more given and topical in discourse than the theme, whereas the prepositional dative is much less constrained and could have meanings other than "transfer". Thus, the double object dative is more likely to be used in specific events where the goal is salient, whereas the use of prepositional dative is much more general based on the less restrictive meaning.

All of the above perspectives argue that learning is largely experience-based, which makes the prediction that learners should show conservativeness, or the lack thereof, in a way that closely reflects their input statistics. However, there are findings from different domains of language learning that suggest learners may have linguistic biases that allow them to generalize beyond input statistics (note that though such biases have been proposed to derive from innate mechanisms, Chomsky, 1965, they could derive also from other perhaps cognitive primitives — we discuss these later). Here, we term this an *internal bias account*. In learning the morphology and syntax of an artificial language similar to Japanese (a free word order language), Fedzechkina, Jaeger, and Newport (2012) showed that English monolinguals constrained their use of case markers in a native-speaker-like manner. In particular, the input to learners only included distributional information to signal that one case marker, either the one for the subject or the object, could sometimes be dropped. However, learners' subsequent production patterns showed sensitivity to an attested alignment (in relevant case-marking languages) between grammatical roles and animacy (i.e., subjects are expected to be animate, and objects inanimate), such that they were more likely to drop the case marker for nouns whose roles can be easily inferred (an animate subject or an inanimate object). Critically, the input distribution never signaled to the learners that animacy was a relevant aspect in the optional use of case markers.

Berent, Harder, and Lennertz (2011) showed that in phonology, four-year-olds also show sensitivity to constraints unattested in their native language. In both of these examples, learners were able to behave in ways that showed sensitivity to information not explicitly encoded in the input. This suggests that language learning is sometimes not simply statistics tracking. Instead, learners may have internal biases that guide their statistical analyses of limited input.

Though some recent results in related psycholinguistic domains are in line with this internal bias account (Berent et al., 2011; Fedzechkina et al., 2012), it is unclear whether this is the case for word order and structural alternation learning. As far as we know, the current work is the first to investigate whether a usage-based account or an internal bias account can better explain how learners reach appropriate generalizations about word order alternations. In three novel language learning experiments, we taught two groups of English monolinguals a typologically different language from their native language (English) with a different basic word order (SOV) and a case-marking system, with three different verbs ("give", "hand", and "show"). All participants learned structural alternations in sentences using the verb "give", but only the canonical structure with "hand" and "show". One group saw word-order variation across stimuli consistent with a grammar similar to those in fixed word order languages, such that local alternations preserved the relative order between the subject and the objects (i.e., a grammar that allows local alternations but not subject-crossing alternations). Meanwhile, the other group saw word-order variation across stimuli consistent with a grammar similar to those in free word order languages, in which alternations can sometimes change the relative order between the subject and the objects (i.e., a grammar that allows both local alternations and subject-crossing alternations). Both groups then performed a picture description task and an acceptability judgment task. Importantly, alternations seen by participants in the first group resemble dative alternations in English, because they can be analyzed as changing the constituent order only within a verb phase. In contrast, alternations seen by participants in the second group resemble the scrambling alternation in Korean, because they can be analyzed as involving subject-crossing without changing the grammatical functions of constituents. The critical question is whether learners are able to generalize alternations appropriately in each grammar, despite the fact that alternations in both grammars occur only with the single verb ("give") in the input.

If learners constrain word-order alternations on a verb-by-verb basis according to the statistical distribution in the input, both groups of learners should largely restrict their production and acceptability of alternated structures to the verb that they learned the alternations with (i.e., "give") and generalize to a smaller extent to the other verbs (i.e., "hand" and "show"). That is, if learners' production behavior and acceptability judgments closely mirror their input statistics, that would support some usage-based and purely statistical accounts. In contrast, if learners are biased to learn that whether word order alternations should or should not be constrained by verb choice depends on whether they experience any subject-crossing alternations, that should lead to qualitatively different generalization patterns across the groups. Specifically, the group that saw a form of scrambling (i.e., subject-crossing alternations) with "give" — the critical evidence of structural flexibility — should be liberal in generalizing structural alternations to "hand" and "show", showing lexical independence. However, the group that only saw local alternations should remain conservative, limiting generalization and so using (local) alternations less with "hand" and "show", showing lexical dependence. In other words, if learners go beyond the input and show generalizations similar to typological patterns, then it would support an internal bias account.

Experiment 1

Method

Participants

Fifty-two undergraduates from the University of California, San Diego participated in the experiment in exchange for course credit. Four of them were excluded from our analyses because they did not produce more than 50% usable data. All participants indicated that they were native English speakers with little to no exposure to other languages.

Materials

Sixty-four line drawings of dative events involving two cartoon characters and an object were used in the learning phase (32 for "give", 16 for "hand", and 16 for "show" events), while another unique set of 48 were used in the test phase (16 for each type of event). To help participants better understand the events during the learning phase, we manipulated the objects involved and the postures of the characters. "Give" events were paired with valuable objects such as a crown or a guitar, with the recipient extending both arms; "hand" events were paired with ordinary objects such as a cup or an apple, with the recipient extending only one arm; "show" events were paired with big objects such as a stroller or a tent, with the observer looking awestruck. This information was not explicitly explained to the participants.

We took extra caution in designing our materials to prevent learners from forming associations between character-object combinations and structures. First, the aforementioned object-verb pairing designed to help participants in the learning phase was no longer present in the test phase, as we used a completely different set of objects and cycled the objects through all verbs the same number of times. The objects at test were of different categories (e.g., ruler, carrot, shirt), so that there was no semantic association between the objects and the verbs. Second, we used eight characters in total repetitively in a balanced manner throughout different phases (exposure and test) of the experiment. In each phase, each character was the agent and the goal of each verb the same number of times, paired with a different character or object each time. The left-right position of the agent in the scene was counterbalanced, so that participants could not associate the role of the characters with the positions they were in. The full list of materials can be found in Appendix A.

We used a two-way mixed design, with exposure group as a two-level between-subject factor and verb as a three-level within-subject factor. We randomly assigned participants to one of two groups—the *non-scrambling group* or the *scrambling group*. The number of participants in each group was balanced (i.e., 24 each). The two groups differed by the types of structural alternation exposure they received. Both groups only saw the canonical word order for "hand" and "show" events, but they saw qualitatively different types of structural alternations for "give" events (see the procedure of the third learning phase, Figure 1.3, and Table 1.1 below for details). In the test phase, both groups first performed a picture description task, then an acceptability judgment task. The combination of these two tasks allowed us to examine if participants exhibited similar generalization patterns in production and comprehension. We arranged the production and acceptability tasks in this order as we believe that performance of the comprehension task may potentially alter production performance more so than vice versa.

Procedure

We taught participants a Korean-English hybrid language through four learning phases. The language was constructed with Korean verbs (*joo-utt-dah* for "give"; *gun-neo-joo-utt-dah* for "hand"; and *bo-yeo-joo-utt-dah* for "show"), case markers (*-ga* for nominative case (NOM), which denotes the agent; *-eul* for accusative case (ACC), which denotes the theme; and *-aegae* for dative case (DAT), which denotes the goal), word order (flexible ordering of the subject, direct object, and indirect object as long as the verb is sentence-final), but English nouns. The goal of this design was to provide participants with a genuine language learning experience, though we used English nouns to make the learning task easier. Korean was chosen, as its word order is clearly different from English, such that participants could not easily transfer their existing knowledge of English to this hybrid language.

In the first learning phase, participants studied the English names of eight cartoon characters (e.g., painter, robber, and dancer) and three events ("give", "hand", and "show") with a randomized stack of cards (eight cards for characters, 32 for "give", and 16 each for "hand" and show", see Figure 1.1 for examples). The purpose was to familiarize the participants with the characters and the events, in order to facilitate later Korean learning.

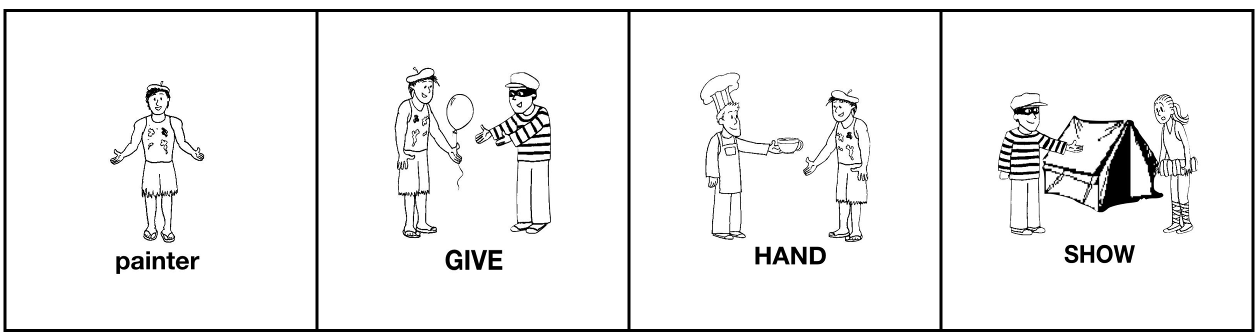

Figure 1.1. Example stimuli in the first learning phase.

The rest of the experiment was conducted using Psychopy (Peirce, 2007). In the second learning phase, participants studied the same set of event pictures one-by-one in a slideshow. The pictures were labelled with Korean verbs instead of English and paired with audio recordings of a native Korean speaker pronouncing the verb on the picture (see Figure 1.2 for examples). Participants studied each picture once, for two seconds each, and they were instructed to pay attention to the meaning and the pronunciation of the verbs.

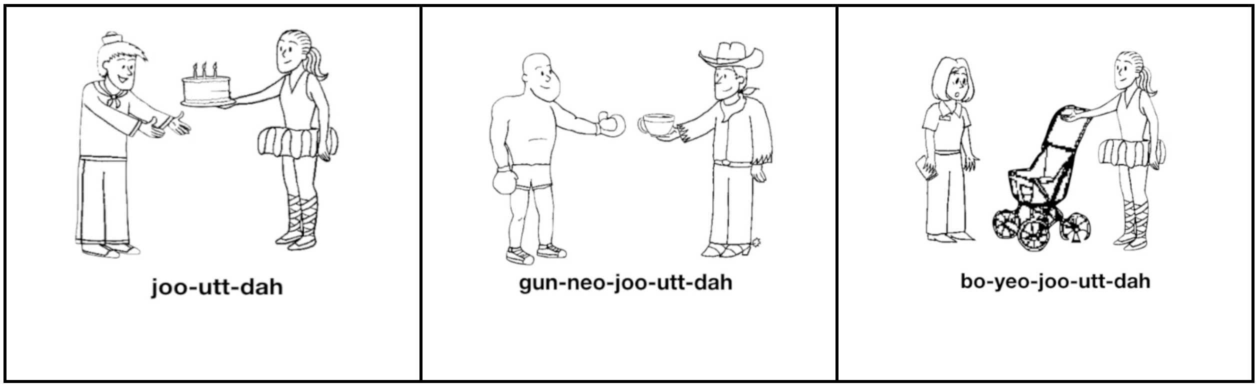

Figure 1.2. Example stimuli in the second learning phase (L-R: “give”, “hand”, and “show”).

In the third learning phase, participants learned to describe the same set of events in complete sentences using the hybrid language. In particular, they were instructed to pay attention to the word orders and case markers. In the instructions, we specifically told the participants that this language places the verb at the end, has different word orders from English (without further specification), and attaches special markers to nouns to indicate their roles (i.e., who is doing the action, who is undergoing the action, and which is the object involved in the action). The pictures were labelled with text cues and paired with audio recordings of a native Korean speaker describing the scenes in complete sentences. Participants were instructed to repeat out loud each sentence in the exact order as in the audio recording, and they were corrected by the experimenter if they forgot the word order or used an order that was different from the recording. The verb of the sentence was always displayed at the bottom of the screen, while the display times of other parts of the sentence were synchronized with the times when the Korean speaker said those parts (see Figure 1.3 for examples of the last frame of each event). In short, participants learned to produce complete sentences in a karaoke manner.

The critical manipulation was that the two experimental groups saw different types of structural alternations in “give” events. Both groups saw 16 “give” events described with the canonical order (NOM-DAT-ACC). The non-scrambling group saw another 16 instances of "give" described with a local structural alternation within the verb phrase (NOM-ACC-DAT, similar to alternations in fixed order languages), whereas the scrambling group only saw eight of the local alternation but also saw eight of a subject-crossing alternation (DAT-NOM-ACC, an alternation only allowed in free order languages). Both groups only saw 16 instances of the canonical order each for "hand" and "show". Table 1.1 summarizes this exposure phase and provides example sentences based on the events in Figure 1.3.

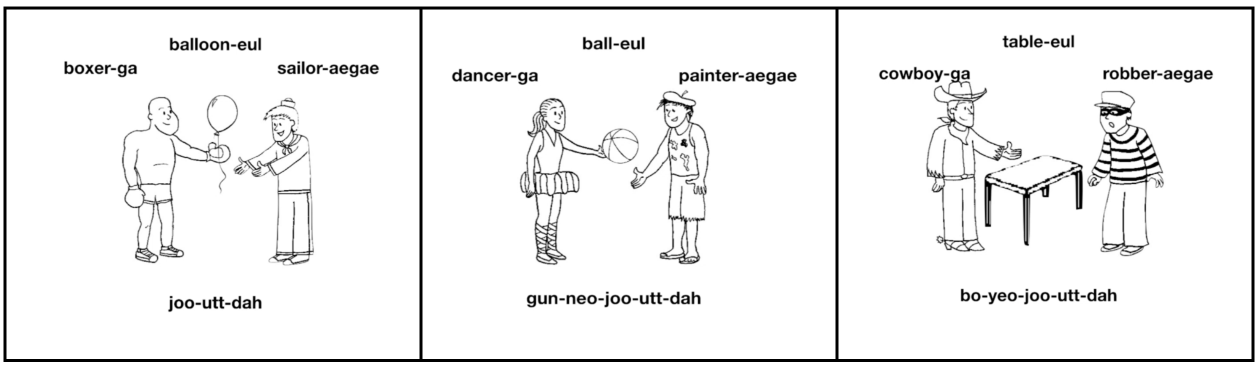

Figure 1.3. Examples of the final frame of each event description in the third learning phase.

Table 1.1

Summary and Example Sentences of the Exposure in the Third Learning Phase by Group

Verb	Structure		Example Sentence	Number of Instances Seen by Group	
				Non-scrambling	Scrambling
give	NOM-DAT-ACC	Canonical	Boxer-ga sailor-aegae balloon-eul joo-utt-dah.	16	16
	NOM-ACC-DAT	Local Alternation	Boxer-ga balloon-eul sailor-aegae joo-utt-dah.	16	8
	DAT-NOM-ACC	Subject-crossing Alternation	Sailor-aegae boxer-ga balloon-eul joo-utt-dah.	0	8
hand	NOM-DAT-ACC	Canonical	Dancer-ga painter-aegae ball-eul gun-neo-joo-utt-dah.	16	16
show	NOM-DAT-ACC	Canonical	Cowboy-ga robber-aegae table-eul bo-yeo-joo-utt-dah.	16	16

Before performing the target tasks, we created a quiz with eight pictures for each verb from the previous phase and asked participants to produce a complete sentence description for each of them without the aid of the audio recordings. The experimenter only intervened if the participant did not mention all components of the picture, misinterpreted the events, or used the wrong case markers. However, participants were never given advice or corrections on word orders. They proceeded to the test phase once they achieved 80% accuracy.

The test phase consisted of a picture description task and an acceptability judgment task. In the picture description task, we presented participants with 48 novel pictures (16 per verb) in a fully randomized order and instructed them to describe the pictures to a hypothetical Korean-English hybrid language speaker. Participants first heard an audio prompt of the Korean-English speaker asking about an element of the picture, either the theme (e.g., “What’s happening with the spoon?”) or the goal (e.g., “What’s happening with the lady?”). Existing research shows that sentence elements that are previously mentioned are more accessible and thus more likely to be uttered earlier than new and less accessible elements (e.g., Bock & Irwin, 1980). Pilot data showed that without this audio prompt, participants were likely to use the canonical word order for all trials. Hence, taking advantage of the givenness/accessibility effect, we hoped to encourage the use of structural alternations using audio prompts that mention either the direct or indirect object across all verbs. Participants responded by describing the picture in a complete sentence using the hybrid language. All spoken responses were recorded using the built-in microphone.

In the acceptability judgment task, we presented the same 48 pictures to participants in a fully randomized order and provided them with a written sentence description under the picture. The description was either the canonical word order, a local alternation, a subject-crossing alternation, or an ungrammatical verb-initial sentence. Each structure occurred equally with each verb. Participants rated the acceptability of the sentence description on a 1 to 7 Likert scale.

Coding and Analyses

Each spoken response from the picture description task was coded in terms of word order, which indicated whether the response was a canonical word order (NOM-DAT-ACC), a local alternation (NOM-ACC-DAT), or a subject-crossing alternation (any utterance with nominative case not first). Because the majority of subject-crossing alternations responses were the one that participants were exposed to (DAT-NOM-ACC), we grouped all non-nominative-first responses into the subject-crossing response category. See Appendix C for a detailed summary of the production responses. Responses in which the participant did not use the hybrid language, did not mention all components of the picture, misinterpreted the event, misused the case markers, or misplaced the verb to any location other than the end of the sentence were excluded. As noted above, 4 participants were excluded from our analyses as a result of failing to provide more than 50% analyzable data. The remaining 48 participants were evenly distributed across experimental conditions. We further excluded 79 spoken responses from the remaining participants based on the criteria above (see Appendix B for the distribution of exclusions across conditions), leaving 2,225 analyzable responses for this task. There were no missing data or response exclusions in the acceptability judgment task (2,304 analyzable responses). All raw acceptability ratings were transformed into z-score units based on each participant's mean and standard deviation prior to analysis. The data can be found at https://osf.io/qudy9/.

The statistical analyses were conducted using R (R Core Team, 2014) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2014). In the picture description task, we analyzed speakers' choice of word order as a function of the different verbs, and whether that choice differed by exposure group. Specifically, we assessed their likelihood of producing local and subject-crossing alternations separately. We built two separate generalized linear mixed effects models with the dependent variable being whether the produced word order was the structure of interest. Verb (give, hand, and show), group (scrambling or non-scrambling exposure group), and their interaction (verb x group) were entered into the model as fixed effects. We used the maximal random effects structure given the experimental design (Barr, Levy, Scheepers, & Tily, 2013), which included intercepts for participants and items, as well as a by-subject random slope for the effect of verb and a by-item random slope for the effect of exposure group.

With a similar method, we examined whether participants' acceptability ratings differed by structures, verbs, and exposure groups. Once again, we were particularly interested in ratings for local and subject-crossing alternations. We built two linear mixed effects models with the ratings (z-scores) of local and subject-crossing alternations as the dependent variable. In each model, verb (give, hand, and show), group (non-scrambling or scrambling exposure group), and their interaction (verb x group) were the fixed effects. We used the same random effects structure as above.

All categorical predictors were transformed into numeric contrasts using the sum-coding method, which enabled us to test for main effects in the presence of interactions (Levy, 2018). We obtained p-values by performing chi-square tests on the maximal model against a nested model that only differed by the fixed effect in question. Since verb is a three-level factor, we conducted pairwise comparisons to understand the difference between "give" and other verbs. All p-values reported in the pairwise comparisons have been adjusted with the Tukey method to account for multiple comparisons. These comparisons demonstrated the degree to which participants generalized alternations from the learned context to novel contexts. Appendix D lists the full output of all models in the current work.

Results and Discussion

Recall that usage-based and statistical accounts predict that learners should produce and accept alternations to an extent that largely reflects their input statistics (i.e., both groups of participants should produce and accept alternations more with “give” than with “hand” and “show”). In contrast, the internal bias account views local and subject-crossing alternations as qualitatively different, in that the latter may signal a greater degree of flexibility in the language (i.e., the non-scrambling group should produce and accept alternations more with “give” than with “hand” and show”; the scrambling group should produce and accept all alternations with all verbs).

Picture Description Task

In this production task, we measured whether the type of exposure speakers received affected their likelihood of producing different types of alternations, and whether they used alternations with verbs that they had never experienced alternations in. Figure 1.4 shows the proportions of responses produced by each group with each verb, color coded by structure.

For the production of subject-crossing alternations, participants who saw scrambling produced 29.5% more subject-crossing alternations (33.1%) than participants who did not see scrambling (3.6%), leading to a significant main effect of exposure group, $\chi^2(1) = 7.74$, $p = .005$. Across groups, participants were about equally likely to produce subject-crossing alternations with each of the three verbs (18.2% for “give”, 18.8% for “hand”, 18.2% for “show”); the main effect of verb was not significant, $\chi^2(2) = 0.68$, $p = .71$. The difference in the proportions of subject-crossing alternations produced by participants who saw scrambling and participants who did not were about equal for each verb—a difference of 30.1% for "give", 30.1% for "hand", and 27.5% for "show"; the interaction between exposure group and verb was not significant, $\chi^2(2)$ = 1.76, p = .41. In sum, participants who saw scrambling (but only with "give") produced more subject-crossing alternations (with all verbs), compared to participants who did not see scrambling.

For the production of local alternations, participants who did not see scrambling were 16.1% more likely to produce local alternations (38.0%) than participants who saw scrambling (21.9%); though the main effect of group was not statistically significant, $\chi^2(1)$ = 2.40, p = .12. Participants were about equally likely to produce local alternations across the three verbs (32.9% for "give", 24.3% for "hand", 32.6% for "show"); the main effect of verb was not statistically significant, $\chi^2(2)$ = 1.93, p = .38. Critically, we found a significant verb x group interaction, $\chi^2(2)$ = 7.62, p = .02. That is, participants who did not see scrambling were more likely to produce local alternations with the verb that they saw them with (47.5% for "give") than with verbs they did not (28.6% for "hand" and 37.8% for "show"). Pairwise comparisons showed that the 18.8% difference between "give" and "hand" was significant (z = 2.87, p = 0.01), whereas the 9.7% difference between "give and "show" (z = 1.70, p = .20) and -9.1% between "hand" and "show" were not significant (z = -0.16, p = .99). In contrast, participants who saw scrambling produced local alternations about equally across all three verbs (18.4% for "give", 20.0% for "hand", and 27.5% for "show"). Pairwise comparisons confirmed that there was no significant difference between the verb pairs for this group: -1.6% difference in the production of non-scrambled alternations between "give" and "hand" (z = -0.55, p = .85); -9.1% between "give" and "show" (z = -0.84, p = .68); and -7.6% between "hand" and "show" (z = -0.78, p = .71). Taken together, participants who never saw scrambling were relatively conservative in their generalization and thus produced the local alternations more with the verb they learned the alternations with ("give"), whereas participants who saw scrambling generalized their production of local alternations about equally across all verbs. A supplementary analysis confirmed that the between-group difference in generalization patterns remained significant regardless of which prompt (goal or theme) the participants received (see Appendix E for the relevant figure and statistics). In our design, prompts were merely used as a proxy to encourage alternations. We thus do not further discuss prompt effects.

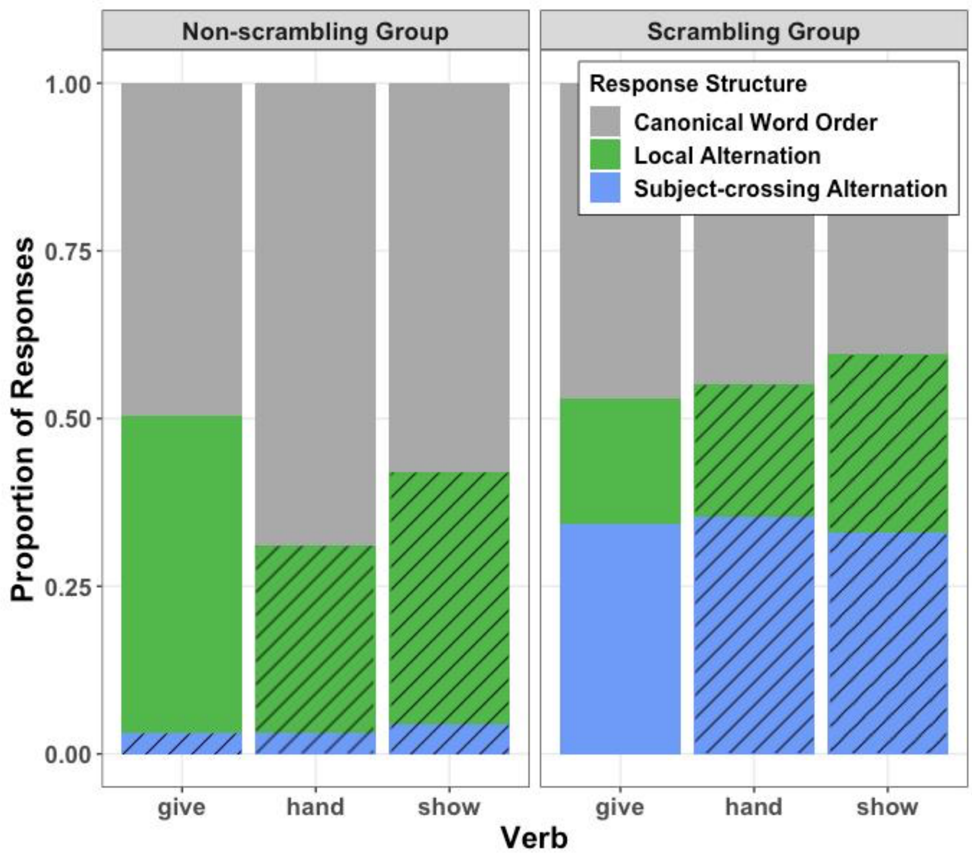

Figure 1.4. Proportion of responses produced by each exposure group with the three verbs in the picture description task, color-coded by response structure. Solid bars represent structure-verb pairings that participants saw in the exposure phase and hashed bars the pairings that participants did not see.

Acceptability Judgment Task

In this comprehension task, we examined whether the type of exposure participants received affected acceptability ratings for subject-crossing and local alternations, and whether participants generalized acceptability to verbs they never experienced alternations with.

In addition to subject-crossing and local alternations, we included the canonical structure and an ungrammatical structure (verb-subject-object) as attention checks and reference points for ratings. As expected, participants rated the canonical structure 2.0 z-score units higher (0.82) than the ungrammatical structure (-1.18). Since the canonical and the ungrammatical structures were not of interest, we did not analyze them further.

Figure 1.5 shows each exposure group's ratings in z-score units for subject-crossing and local alternations with the three verbs. We consider first ratings of subject-crossing alternations. Participants who saw scrambling rated subject-crossing alternations higher on average (0.27) than did participants who did not see scrambling (-0.55); the main effect of group was significant, $\chi^2(1) = 46.20$, $p < .001$. Notably, participants who saw scrambling rated subject-crossing alternations about equally across the three verbs (0.28 for "give", 0.39 for "hand", and 0.12 for "show"), as did the group that never saw scrambling (-0.54 for "give", -0.56 for "hand", and -0.54 for "show"); there was no main effect of verb ($\chi^2(2) = 1.58$, $p = .45$) nor a verb x group interaction ($\chi^2(2) = 2.59$, $p = .27$) for subject-crossing alternations. That is, participants who never saw scrambling consistently rated subject-crossing alternations poorly, whereas participants who saw scrambling generalized their high ratings for subject-crossing alternations across all verbs, even verbs with which they did not see such subject-crossing alternations.

Next we turn to ratings of local alternations. Participants who did not see scrambling rated the local alternations higher on average (0.64) than did participants who saw scrambling

(0.08); the main effect of group was significant, $\chi^2(1) = 5.43$, $p = .02$. Moreover, participants rated local alternations better with “give” (0.67) than with “hand” (0.38) and “show” (0.45); the main effect of verb was significant, $\chi^2(2) = 13.73$, $p = .001$. However, the verb effect was mainly driven by the non-scrambling group, as the results also showed that there was a between-group difference in generalization patterns; the verb x group interaction was significant, $\chi^2(2) = 11.29$, $p = .004$. Specifically, participants who never saw scrambling rated local alternations higher with “give” (0.97) than with “hand” (0.42) and “show” (0.52). Pairwise comparisons showed that the difference between “give” and “hand” was significant ($t = 4.88$, $p < .001$), as was the difference between “give” and “show” ($t = 3.60$, $p = .002$), but not between “hand” and “show” ($t = -1.39$, $p = .36$). In contrast, participants who saw scrambling rated local alternations about equally with “give” (0.36), “hand” (0.34), and “show” (0.38). Pairwise comparisons confirmed that there was no statistical difference between “give” and “hand”($t = 0.18$, $p = .98$), “give” and “show” ($t = -0.17$, $p = .98$), nor “hand” and “show” ($t = -0.59$, $p = .83$). In other words, participants who did not see scrambling showed lexical dependence in their generalization of local alternations, which was not observed in participants who saw scrambling.

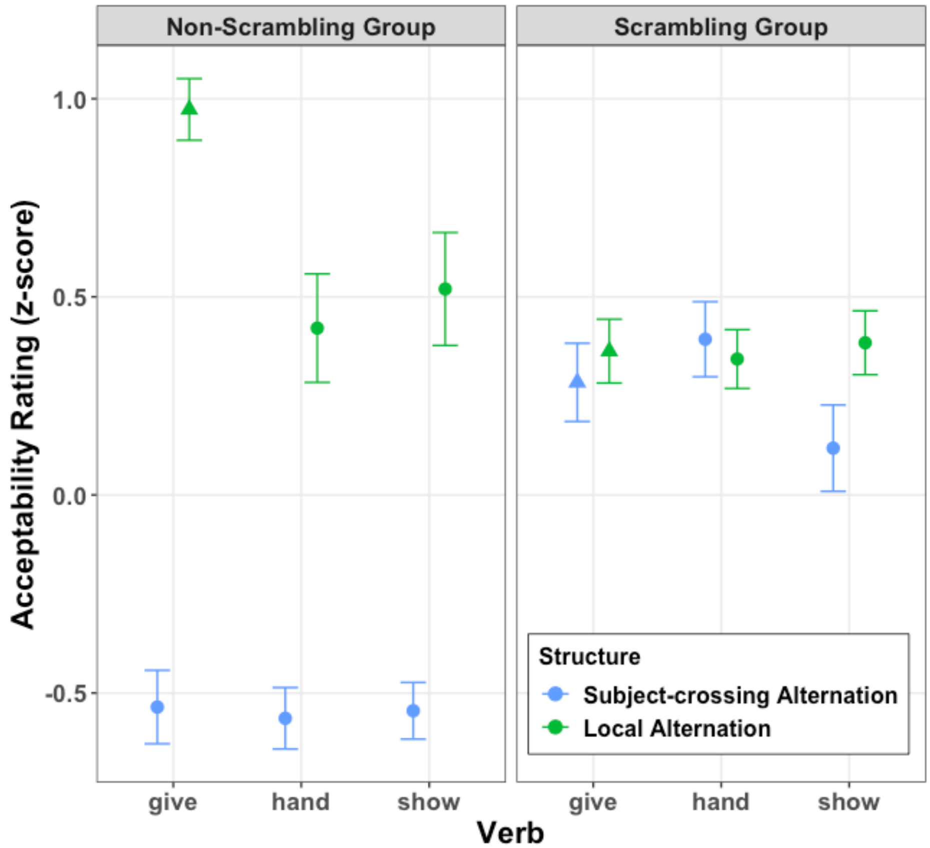

Figure 1.5. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task, color coded by structure. Triangles represent structure-verb pairings that participants saw in the exposure phase and circles the pairings that participants did not see. Error bars represent standard errors.

To summarize, Experiment 1 demonstrated two main points: First, and expectedly, participants who saw scrambling produced and accepted subject-crossing alternations, whereas participants who did not see scrambling rarely produced or accepted subject-crossing alternations. Second, and more notably, even though all participants saw the local alternation, the group that saw scrambling more readily generalized the production of local alternations with verbs not experienced with local alternations than the group that did not see scrambling. That is, the non-scrambling group restricted their production and acceptability of local alternations more to the verb that they learned these structures with, whereas the scrambling group generalized liberally across all verbs. This qualitative difference in generalization patterns is consistent with cross-linguistic patterns and so also with the internal bias account, but not with the usage-based or statistical accounts, as nothing about the usage of the verbs "hand" and "show" differed between the two exposure groups.

It is worth noting that as shown in Table 1.1, a potential confound of Experiment 1 is that participants from the different exposure groups saw different numbers of distinct structures in the learning phase, even though the total number of instances seen was the same. To be specific, the non-scrambling group only learned two distinct structures with "give" (16 instances of the canonical word order NOM-DAT-ACC and 16 instances of the local alternation NOM-ACC-DAT), whereas the scrambling group learned three distinct structures with "give" (16 instances of the canonical word order NOM-DAT-ACC, 8 instances of the local alternation NOM-ACC-DAT, and 8 instances of the subject-crossing alternation DAT-NOM-ACC). It may be that the liberal generalization behavior displayed in the group that saw three structures was due to confusion or a failure to keep track of verb-specific statistics because of this broader exposure regimen. Therefore, we conducted Experiment 2, in which both groups of participants learned two distinct structures with "give". The difficulty level of keeping track of verb-specific statistics should then be equalized, and any qualitative between-group difference in generalization thus cannot be attributed to confusion.

Experiment 2

Method

Participants

A separate group of 53 undergraduates from the University of California, San Diego participated in the experiment in exchange for course credit. Five of them were excluded from our analyses because they did not produce more than 50% usable data. All participants indicated that they were native English speakers with little to no exposure to other languages.

Materials and Procedure

The materials and design were very similar to those in Experiment 1 with one minimal change: Participants in the scrambling group saw two distinct alternations with "give" (the canonical word order NOM-DAT-ACC and the subject-crossing alternation DAT-NOM-ACC) instead of three, effectively eliminating the aforementioned confound while maintaining the critical qualitative difference in exposure for the two groups (exposure to scrambling versus the lack thereof). Table 1.2 shows the details of the learning phase in Experiment 2, with the design changes in bold. The procedure was exactly the same as in Experiment 1.

Table 1.2

Summary and Example Sentences of the Exposure in the Learning Phase by Group

Verb	Structure		Example Sentence	Number of Instances Seen by Group	
				Non-scrambling	Scrambling
give	NOM-DAT-ACC	Canonical	Boxer-ga sailor-aegae balloon-eul joo-utt-dah.	16	16
	NOM-ACC-DAT	Local Alternation	Boxer-ga balloon-eul sailor-aegae joo-utt-dah.	16	**0**
	DAT-NOM-ACC	Subject-crossing Alternation	Sailor-aegae boxer-ga balloon-eul joo-utt-dah.	0	**16**
hand	NOM-DAT-ACC	Canonical	Dancer-ga painter-aegae ball-eul gun-neo-joo-utt-dah.	16	16
show	NOM-DAT-ACC	Canonical	Cowboy-ga robber-aegae table-eul bo-yeo-joo-utt-dah.	16	16

Coding and Analyses

The coding criteria and the mixed-effects models used in the analyses were the same as those in Experiment 1. After (as noted) excluding five participants due to low accuracy, we further excluded 20 spoken responses from the remaining participants (see Appendix B for the distribution of exclusions across conditions), leaving 2,284 analyzable responses for the picture description task. There were no missing data or response exclusions in the acceptability judgment task (2,304 analyzable responses). All acceptability ratings reported below are in z-score units calculated based on each participant's mean and standard deviation prior to analysis.

Results and Discussion

Picture Description Task

Figure 1.6 shows the proportions of responses produced by each group with each verb, color coded by structure. As can be seen, the non-scrambling group produced a very small percentage of subject-crossing alternations (6.3%), whereas the scrambling group produced no local alternations at all, providing few to no observations for between-group comparisons for each structure. Thus, we report below the statistical analyses of the likelihood of producing the learned alternation (i.e., local alternations for the non-scrambling group and subject-crossing alternations for the scrambling group) as a function of group, verb, and their interaction.

First, we found that the non-scrambling group and the scrambling group were about equally likely to produce the type of alternation they were exposed to (14.7% of the responses were local alternations for the non-scrambling group and 24.9% of the responses were subject-crossing alternations for the scrambling group); the main effect of group was not statistically significant, $\chi(1) = 0.002$, $p = .96$.

Second, participants were equally likely to produce learned alternations across the three verbs. For the scrambling group, subject-crossing alternations accounted for 15% of the production responses for "give", 17.1% for "hand", and 14.8% for "show"; For the non-scrambling group, local alternations accounted for 19.4% of the production responses for "give", 14.2% for “hand”, and 10.5% for “show”. There was no significant main effect of verb ($\chi(2)$ = 0.91, p = 0.64).

Third, due to the low number of observations of alternations as reported above, we did not detect any group x verb interaction in Experiment 2; $\chi(2)$ = 1.31, p = .52.

Taken together, participants were more likely to produce the alternations that they were exposed to, and they rarely or never produced alternations that they were not exposed to. Contrary to Experiment 1, both groups used alternations about equally across all verbs.

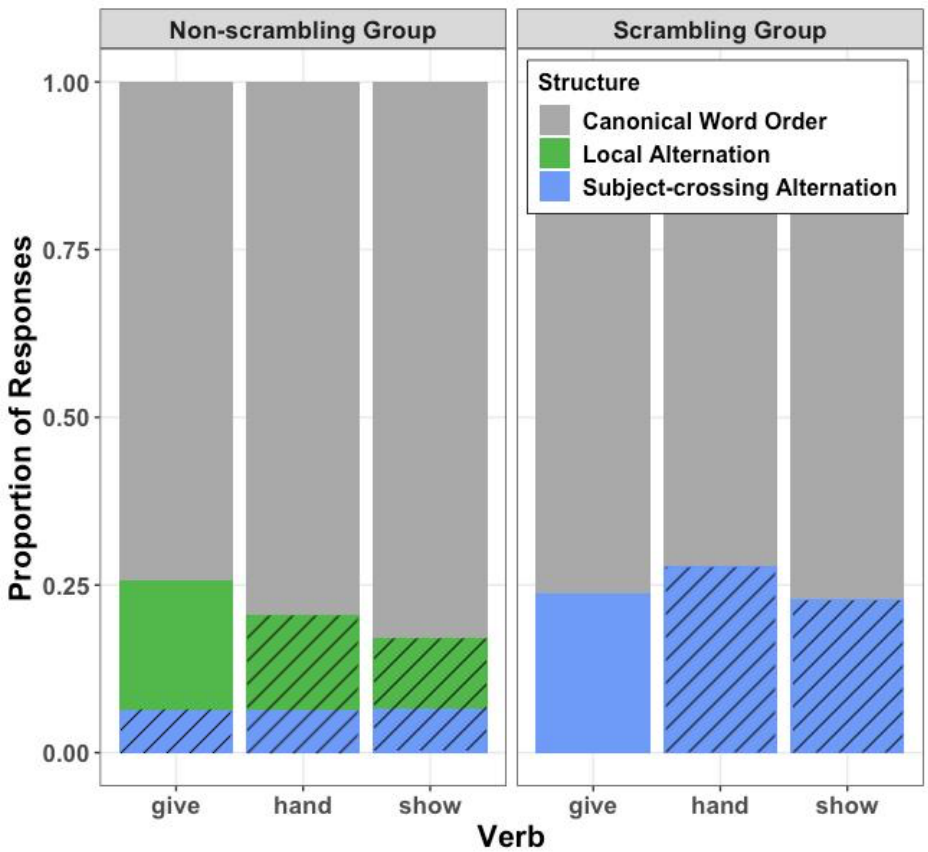

Figure 1.6. Proportion of responses produced by each exposure group with the three verbs in the picture description task, color coded by response structure. Solid bars represent structure-verb pairings that participants saw in the exposure phase and hashed bars the pairings that participants did not see.

Acceptability Judgment Task

Like Experiment 1, participants rated the canonical word order as the most acceptable (1.04) and the ungrammatical structure as the least acceptable (-1.08). Since these two were not structures of interest, we did not further analyze them. Below we report the acceptability ratings for local and subject-crossing alternations. Figure 1.7 shows the acceptability ratings (in z-score units) by each group with each verb, color coded by structure.

First, we found that participants rated the alternation that they were not exposed to significantly more poorly than the alternation that they learned. The non-scrambling group rated local alternations as significantly more acceptable (0.4) than did the scrambling group (-0.5); the main effect of group for local alternations was significant, $\chi(1) = 63.14$, $p < .001$. Likewise, the scrambling group rated subject-crossing alternations as significantly more acceptable (0.56) than did the non-scrambling group (-0.38); the main effect of group for subject-crossing alternations was significant, $\chi(1) = 47.02$, $p < .001$.

Second, the degree to which participants accepted a given structure was generally similar across all verbs, regardless of whether or not they had seen that structure used with that particular verb. The average ratings for local alternations were 0.002 for "give", -0.05 for "hand", and -0.1 for "show" ($\chi(2) = 2.53$, $p = .28$); whereas the average ratings for the subject-crossing alternations were 0.09 for "give", 0.14 for "hand", and 0.04 for "show" ($\chi(2) = 0.88$, $p = .64$).

Critically, we found evidence that the two groups differed in generalization patterns, a pattern analogous to the results in Experiment 1. In particular, participants who did not see scrambling once again showed lexical dependence for local alternations, whereas participants who saw scrambling did not; the exposure group x verb interaction for local alternations was significant, $\chi(2) = 7.72$, $p = .02$. The non-scrambling group accepted local alternations more with "give" (0.53) than with "hand" (0.41) and "show" (0.27). Pairwise comparisons revealed that the difference between "give" and "show" was significant (0.26; $t = 2.95$, $p = .02$), while the difference between "give" and "hand" (0.12; $t = 0.99$, $p = .59$) and "hand" and "show" (0.14; $t = 1.64$, $p = .25$) did not reach statistical significance. Meanwhile, the scrambling group rated local alternations about equally poorly across all verbs (-0.52 for "give", -0.50 for "hand", and -0.47 for "show"; no pairwise comparisons reached statistical significance). Also consistent with Experiment 1, this group x verb interaction was not observed for subject-crossing alternations ($\chi(2) = 0.43$, $p = .81$). As the averages reported in the first point showed, the non-scrambling group rated subject-crossing alternations about equally poorly and the scrambling group about equally highly.

To summarize, exposure to a structure seemed to be necessary for it to be judged as acceptable. Nonetheless, we still found evidence that exposure type influenced generalization patterns of learned structures.

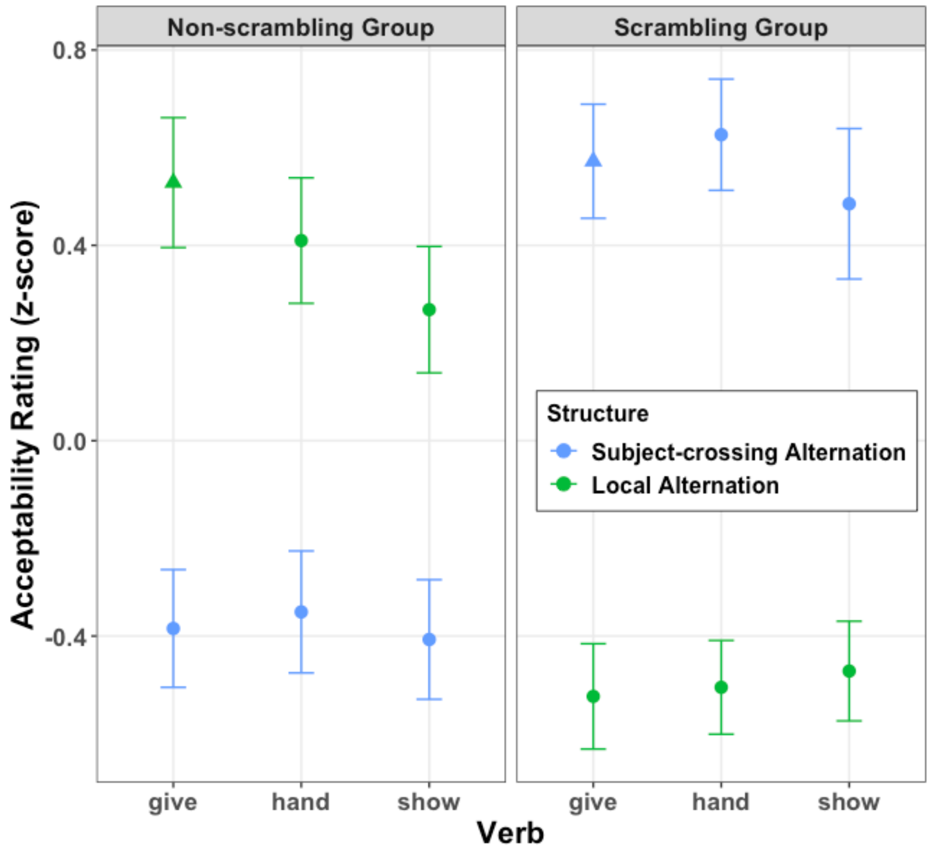

Figure 1.7. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task, color coded by structure. Triangles represent structure-verb pairings that participants saw in the exposure phase and circles the pairings that participants did not see. Error bars represent standard errors.

Note that we found a statistically significant between-group difference in generalization pattern in comprehension (like in Experiment 1) but not in production (unlike in Experiment 1), though the findings generally fit with the predictions from the internal bias account. Hence, we conducted exploratory post-hoc analyses to investigate the discrepancies in generalization strategies. A post-hoc comparison of the data from the two experiments revealed that many fewer participants produced any alternations at all in Experiment 2 than in Experiment 1: We discovered that 36 out of 48 participants in Experiment 1 produced alternations at all (17 in the non-scrambling group and 19 in the scrambling group), whereas only 25 out of 48 participants did so in Experiment 2 (12 in the non-scrambling group and 13 in the scrambling group). The low number of observations of alternations in Experiment 2 likely led to insufficient power in detecting any across-verb differences, if there were any. In a follow-up experiment, we attempted to enhance the relevance and maximize the accessibility of the production question prompt in the picture description task, in order to encourage the production of even more alternations.

Experiment 3

Method

Participants

Fifty-two undergraduates from the University of California, San Diego participated in the experiment in exchange for course credit. Four of them were excluded from our analyses because they either did not produce more than 50% usable data, or had extensive prior knowledge of languages other than English. All remaining participants indicated that they were native English speakers with little to no exposure to other languages.

Materials and Procedure

The materials in Experiment 3 were exactly the same as Experiment 2. There were two major differences between Experiment 2 and 3: First, we eliminated the quiz before the test phase in Experiment 3. That is, participants in Experiments 1 and 2 were told that they had to reach 80% accuracy before moving onto the test phase. While the majority of participants learned the hybrid language on their own and did not require any correction during the quiz, it might have inadvertently reinforced structures that were implicitly confirmed to be correct and led to a least-effort strategy. Second, in Experiments 1 and 2, when participants were asked about the theme or goal ("What's happening with the spoon/lady?"), as a way to elicit more alternation, there was only one picture to describe; this may have permitted participants to ignore the prompt.

So, in Experiment 3, participants saw two pictures on the screen in the picture description task, with only one of the two containing the prompted element in the audio. This change was an attempt to make the audio prompt relevant to the production task by requiring participants to encode the theme or goal, boosting the accessibility of that theme or goal. Everything else was identical to Experiment 2.

Coding and Analyses

The coding criteria and the mixed-effects models used in the data analysis were the same as those in Experiment 1 and 2. After excluding (as noted) four participants due to low accuracy, we further excluded 91 spoken responses from the remaining participants (see Appendix B for the distribution of exclusions across conditions), leaving 2,211 analyzable responses for the picture description task. There were no missing data or response exclusions in the acceptability judgment task (2,304 analyzable responses). All acceptability ratings reported below are in z-score units calculated based on each participant's mean and standard deviation prior to analysis.

Results and Discussion

Picture Description Task

Figure 1.8 shows the proportions of responses produced by each group with each verb, color coded by structure. Unlike Experiments 1 and 2, neither group of participants produced any alternations that they were not exposed to. Therefore, we simplified the analysis to whether verb and exposure type affected the likelihood of producing alternations other than the canonical word order.

Both groups of participants produced the type of alternation that they saw about equally often (27.2% for the non-scrambling group and 29.6% for the scrambling group). As a result, no main effect of exposure group was observed; $\chi(1) = 0.01$, $p = .92$. On average, participants produced more alternations with "give" (i.e., the verb that they learned the structure with; 34.7%) than with "hand" (26.5%) and "show" (23.9%). A significant main effect of verb was observed, $\chi(2) = 9.22$, $p = .01$.

Although the interaction between verb and exposure group was not statistically significant ($\chi(1) = 1.08$, $p = .58$), pairwise comparisons revealed suggestive evidence of group differences in generalization patterns. The verb main effect was mainly driven by the non-scrambling group, who produced more alternations with "give" (37.3%) than with "hand" (24.5%) and "show" (19.8%). There were significant differences between "give" and "hand" ($z = 2.84$, $p = .01$) and "give" and "show" ($z = 3.24$, $p = .003$), but not "hand" and "show" ($z = -0.39$, $p = .92$), which reflects a conservative generalization pattern in line with our observations in Experiment 1 and the numerical trend in Experiment 2. In contrast, the scrambling group produced alternations about equally for "give" (32%), "hand" (28.7%), and "show" (28.1%). Thus, no pairwise comparisons were significant in the scrambling group ("give" and "hand": $z = 1.54$, $p = .27$; "give" and "show": $z = 0.89$, $p = .65$; "hand" and "show": $z = -0.85$, $p = .68$).

In sum, no participant produced structures that they were not exposed to. Though there was a numerical trend of more conservative generalization in the non-scrambling group compared to the scrambling group in production, the difference was not significant.

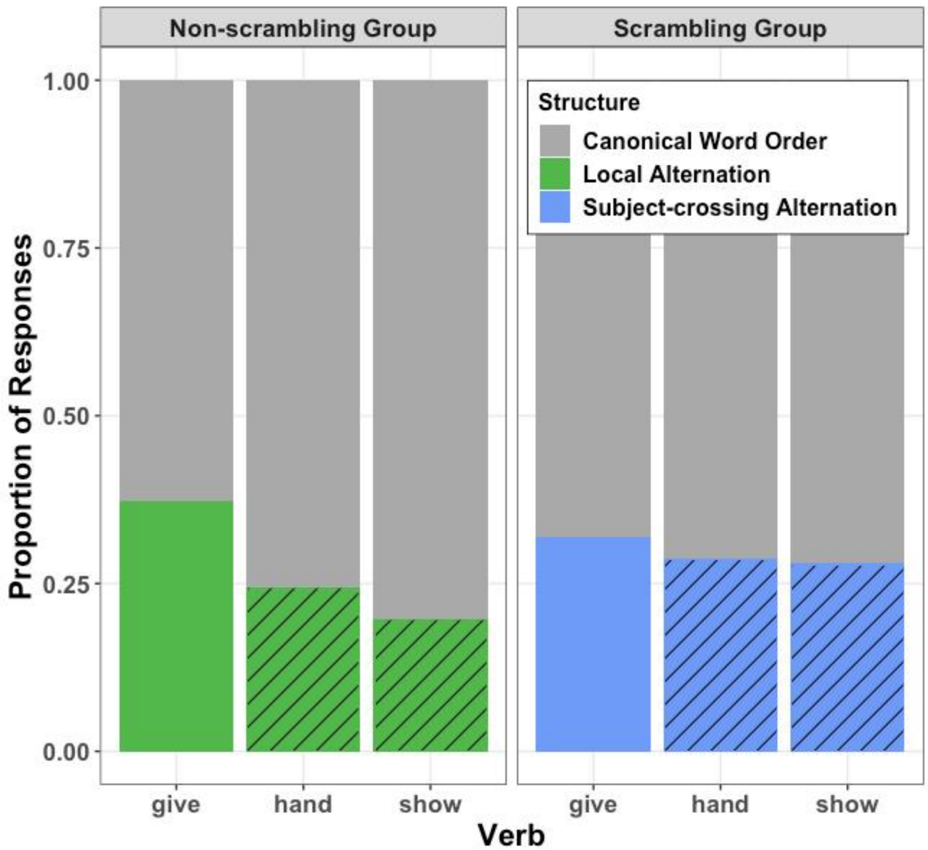

Figure 1.8. Proportion of responses produced by each exposure group with the three verbs in the picture description task, color coded by response structure, color coded by response structure. Solid bars represent structure-verb pairings that participants saw in the exposure phase and hashed bars the pairings that participants did not see.

Acceptability Judgment Task

As expected, participants rated the canonical word order as the most acceptable (0.99) and the ungrammatical structure as the least acceptable (-1.07). We therefore focus on reporting the structures of interest below. Figure 1.9 shows the acceptability ratings (in z-score units) by each group with each verb, color coded by structure.

As with Experiment 1 and 2, we found that participants rated the alternation that they were not exposed to significantly more poorly than the alternation that they learned. The non-scrambling group rated local alternations as significantly more acceptable (0.6) than did the scrambling group (-0.43); the main effect of group for local alternations was significant, $\chi(1)$ =

87.85, $p < .001$. Likewise, the scrambling group rated the subject-crossing alternations as significantly more acceptable (0.57) than did the non-scrambling group (-0.58); the main effect of group for subject-crossing alternations was significant, $\chi(1) = 103.14$, $p < .001$.

Also comparable with the previous experiments, the degree to which participants accepted a given structure was generally similar across all verbs, the main effect of verb was not significant for either structure. The average ratings for local alternations were 0.2 for "give", 0.03 for "hand", and 0.02 for "show" ($\chi(2) = 4.60$, $p = .10$); whereas the average ratings for subject-crossing alternations were 0.04 for "give", -0.007 for "hand", and -0.04 for "show" ($\chi(2) = 0.62$, $p = .73$).

Critically, we found evidence consistent with our previous findings that the two groups differed in generalization patterns. For local alternations, participants who did not see scrambling displayed lexical dependence in their generalization again, whereas participants who saw scrambling did not, leading to a significant group x verb interaction ($\chi(2) = 7.89$, $p = .02$). To be specific, the non-scrambling group accepted local alternations more with "give" (0.84) than with "hand" (0.52) and "show" (0.44). Pairwise comparisons revealed that there was a significant difference between "give" and "show" ($t = 3.40$, $p = .004$), a marginal difference between "give" and "hand" ($t = 2.32$, $p = .06$), whereas the difference between "hand" and "show" ($t = 0.90$, $p = .64$) did not reach statistical significance. Meanwhile, the scrambling group rated the local alternations about equally poorly across all verbs (-0.44 for "give", -0.46 for "hand", and -0.39 for "show"; no pairwise comparisons reached statistical significance). Also in line with our previous findings, this group x verb interaction was not observed for subject-crossing alternations. The averages reported above revealed that the non-scrambling group rated subject-crossing alternations about equally poorly and the scrambling group about equally highly (no exposure group x verb interaction, $\chi(2) = 2.50$, $p = .29$).

Once again, we found that the non-scrambling group showed more conservative generalization patterns than did the scrambling group in comprehension.

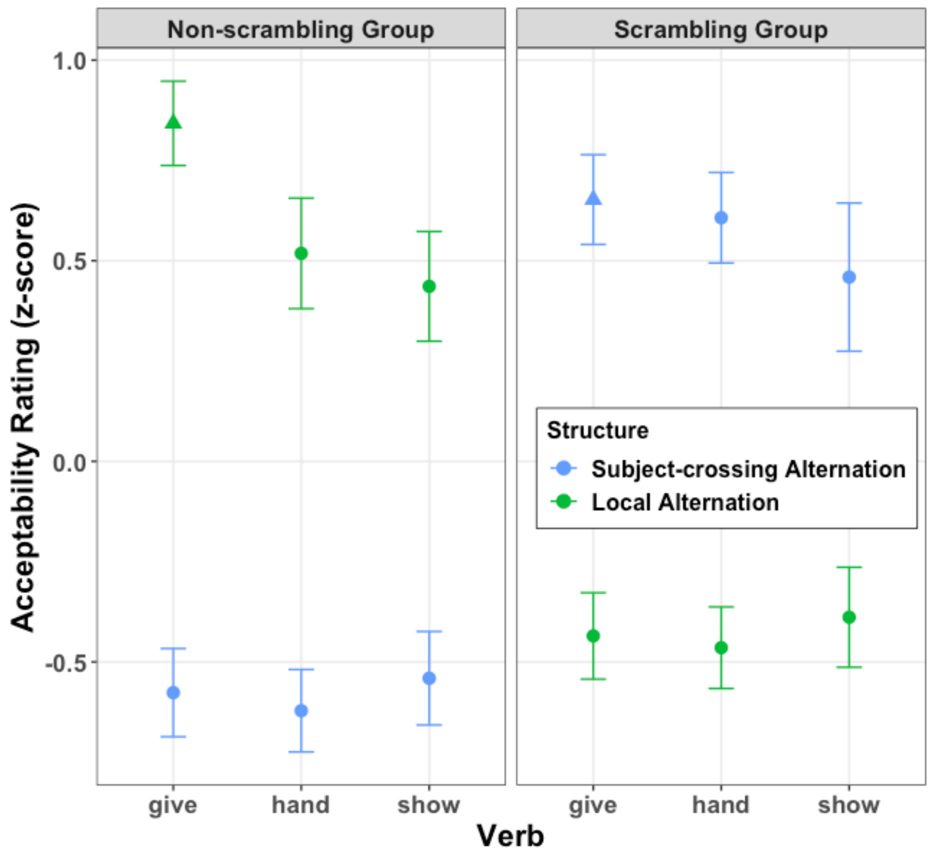

Figure 1.9. Ratings (z-scores) for subject-crossing and local alternations by each exposure group with the three verbs in the acceptability judgment task, color coded by structure. Triangles represent structure-verb pairings that participants saw in the exposure phase and circles the pairings that participants did not see. Error bars represent standard errors.

General Discussion

Across three experiments, we taught English monolinguals an artificial Korean-English hybrid language. All participants learned alternations only with one verb ("give"), but only the canonical word order with other verbs ("hand" and "show"). We investigated whether or not participants would generalize alternations to verbs that they had never experienced alternations with. Additionally, half of the participants were exposed to a grammar with no evidence of scrambling (the non-scrambling group), whereas the other half saw a grammar with scrambling (the scrambling group). The critical question here is whether or not there would be any between-group differences in generalization patterns, which separate different theoretical accounts of grammar learning.

Experiment 1 showed that participants who only saw local alternations with "give" were verb-wise conservative in their generalizations. That is, they produced and accepted local alternations more with "give" than with "hand" and "show". In contrast, participants who were exposed to both scrambling (i.e., subject-crossing alternations) and local alternations generalized all types of alternations across all verbs. This group difference in generalization was found in both production and comprehension. We suggest that these results may support an internal bias account but could also be explained by a failure to keep track of input statistics due to the greater structural variability in the input that the scrambling group received. Experiment 2 eliminated this confound by equating the number of distinct structures each group saw and yielded results that were generally in line with those in Experiment 1, with the exception that both groups (rather than only the non-scrambling group) rarely or never produced or accepted structures that they had never seen. Nevertheless, the non-scrambling group was once again verb-wise conservative with their acceptability judgments of learned structures, whereas the scrambling group was not. Though the numerical trend in production was compatible with the conclusion of Experiment 1 and the comprehension data of Experiment 1 and 2, the verb differences in production were not statistically significant. We attributed this to a lack of power, due to the majority of the production responses being the canonical word order. Experiment 3 implemented several design changes that successfully encouraged the production of alternations and showed that there were reliable group differences in generalization patterns in comprehension (and pairwise differences between verbs in the same direction in production), corroborating Experiment 1 even after a design confound was addressed.

Note that the consistency of the comprehension data was higher than that of production data. This could be due to three factors: First, there were no missing observations at all in the comprehension data, in contrast with our production data. Second, the comprehension task by design elicited learners' acceptability judgments for all four structures of interest across trials, whereas extra effort was needed to encourage alternations in a free production task. Third, and more speculatively, learners of a novel language may have felt more uncertain when constructing sentences using the necessary components of the events by themselves, compared to when all the components and a complete sentence were already laid out in front of them, and all that was required was just a relatively intuitive judgment. Nevertheless, we obtained reliable comprehension data and suggestive production data of between-group generalization differences in support of the internal bias account.

On the surface, different languages vary in many ways in terms of structural flexibility. In the domain of structural alternations, there are two covarying features among the world's languages, namely whether word order is relatively free and whether alternations are lexically-dependent. Alternations in relatively fixed word order languages, like English, tend to be local and lexically dependent. On the other hand, relatively free word order languages, such as Korean, allow all meaning-preserving alternations with all verbs, including subject-crossing alternations (i.e., dative alternations are not lexically dependent). In order to master the grammar of a language so that they can avoid over- or under-generating, learners need to know the structural flexibilities and constraints of their language. How do learners know which type of language they are operating in?

The current work pitted two classes of accounts against one another. Usage-based and purely statistical accounts suggest that learning is largely experience-based and can be reduced to some form of input-statistics tracking (Brooks & Tomasello, 1999; Tomasello & Brooks, 1998; Wonnacott et al., 2008, 2017). This class of accounts suggests that learners should be somewhat conservative in their generalization and restrict their use of alternations to verbs that they learned these structures with. One important point to note is that most accounts of this kind do not explicitly posit any internal inductive bias that guides how the input is analyzed. In contrast, internal bias accounts (Berent et al., 2011; Chomsky, 1965, 1980; Fedzechkina et al., 2012) argue that learners may have tacit knowledge about how languages of the world operate, including languages that they do not know. The relevant knowledge in learning structural alternations would be the covarying features of scrambling and lexical independence from the verb. If learners indeed implicitly represent the link between these features, then the breadth of their use of structural alternations should depend on whether or not they see the critical evidence of scrambling. Only if learners see scrambling should they generalize alternations across all verbs, but they should otherwise be verb-wise conservative.

The results across experiments generally showed that participants who did not see any evidence suggesting that the alternations they encountered were due to scrambling were verb-wise conservative, whereas participants who saw some evidence suggesting that the alternations they encountered were due to scrambling were not conservative. That is, the non-scrambling group consistently accepted and sometimes produced alternations more with "give" than with "hand" and "show", but the scrambling group produced and accepted alternations equally with all verbs. Our findings support the internal bias account.

An alternative account for our results concerns the difference in input complexity across the exposure groups. In particular, Hudson Kam and Newport (2005, 2009) showed that when asked to keep track of grammatical forms (such as determiners) that are used probabilistically and inconsistently, learners are more likely to resort to the dominant response across the board when the input complexity is high, as opposed to observing and reproducing the inconsistencies when the input complexity is relatively low. Admittedly, we cannot rule out the possibility that the scrambling group "gave up" learning the input complexities in Experiment 1, as learners of that group had to learn one more distinct structure than did the non-scrambling group. However, this explanation has more difficulty explaining the scrambling group's generalization performance in Experiment 2 and 3. After eliminating a design confound in Experiment 1, the input complexity did not differ substantially between groups in these two experiments. Yet, the scrambling group still generalized across the board. Moreover, learners in our experiments only had to keep track of two distinct structures with three verbs, which included far fewer complexities to keep track of compared to other work using similar paradigms. Also, it would require the least effort for participants to resort to the dominant response for all verbs, which was the canonical word order, and not alternate at all, but the scrambling group produced and accepted alternations across the board in all experiments. Hence, the possibility of the scrambling group consistently giving up seems unlikely.

One limitation of the current study is that we cannot prohibit the transfer of grammatical knowledge from English to the hybrid language. Note, however, that language transfer from English alone cannot fully explain our results. All participants were English monolinguals, so both groups should have constrained their generalization of local alternations to a specific verb, in the way that they should for prepositional dative and double object dative alternations in English. In fact, the word orders of the canonical word order and the local alternation in the hybrid language are identical to the double object dative and the prepositional dative in English, respectively, with the exception of the verb position. That is, the direct object switches positions with the indirect object, but the subject remains in the first position across these structures. If participants indeed used their English knowledge to guide their generalizations, they should show lexical dependence for local alternations (i.e., using local alternations more with "give" than with "hand" and "show"), regardless of whether or not they saw subject-crossing alternations in addition to local alternations. Again, only the non-scrambling group, but not the scrambling group, showed such conservativeness. Meanwhile, the scrambling group in Experiment 1 produced and accepted both local and subject-crossing alternations about equally with all verbs (i.e., no lexical dependence was observed for any structures). This suggests that the alternative explanation of English knowledge transfer does not entirely fit with the results.

Though it is unclear to us how transfer from English could lead to the differential generalization for the two exposure groups, speakers' native language could have had more subtle effects involving higher-level reasoning. For instance, as we noted above, participants may have noticed the word order similarities between the canonical word order and the local alternation in our experiment and the dative constructions in English. With the prior expectation that datives in English are lexically-constrained, participants in the non-scrambling group may have assumed the hybrid language to be an English-like language and thus generalized conservatively. In addition, it is generally unacceptable for the object(s) to precede the subject in English, so the scrambling group (who saw subject-crossing alternations) may have reasoned that the hybrid language was quite different from English, which may have weakened their prior expectations that datives could be lexically-constrained, resulting in liberal generalization. Though this type of high-level reasoning is plausible, as shown by prior work that learners can observe high-level constraints in both natural and artificial languages (e.g., Perfors et al., 2010; Thothathiri & Braiuca, 2020; Thothathiri & Rattinger, 2016; Wonnacott et al., 2008, 2017), we must also point out one important detail that this account has more difficulty explaining. The structure with the highest frequency for both groups was the canonical word order, which was also the only order that both groups experienced with all verbs. Importantly, this structure has a similar word order as the double object dative in English, which generally only alternating verbs allow. With the prior knowledge of English that double object datives are associated with alternating verbs, this account raises the question of why the non-scrambling group did not assume that the hybrid language was not lexicalist at all, despite the abundance of evidence of the more restricted structure in English being used with all verbs in the novel language.

Relatedly, across different languages, the agent of an event (the subject in our case) tends to be the most salient and is often produced earlier than other less salient thematic roles (e.g., Osgood & Bock, 1977). Given that subject-crossing alternations violate this norm and are unusual, the scrambling group may have used some form of high-level pragmatic reasoning to assume that all bets are off, also resulting in liberal generalization. As we noted in the introduction, pragmatic accounts of statistical learning postulate that discourse functions (Perek & Goldberg, 2015, 2017) can influence the usage and generalization of constructions. However, it is important to note that our version of the internal bias account complements rather than argues against pragmatic accounts, by specifically linking pragmatics with an "internal bias" in conceptual organization. We will return to this point after elaborating on our speculations on what the "internal bias" in learning structural alternations may be.

So far, we have argued for the internal bias account with the support of our data, but what exactly is this "internal bias"? How is it that learners are able to represent features that are not necessarily in their own native language (such as a subject-crossing alternation for a speaker of English), let alone make inferences based on information that is not encoded explicitly in their input? Our speculation is that conceptual structure may be organized such that the elements within the predicate (action and patient, along with any other elements of the predicate) form a more cohesive conceptual "chunk" than the subject (agent) does with any predicate element, and language users prefer linguistic structures that are isomorphic to conceptual structure (i.e., structures with a clear subject-predicate boundaries), unless there are extra-linguistic motivations (e.g., pragmatics) that warrant the violation of the preference for isomorphism. Although we cannot rule out the possibility of the subject-predicate conceptual structure being innate, the cognitive biases for isomorphism shown in our results may not require a strong nativist claim. One possibility is that it arises from the architectural constraints on our cognition, such as memory constraints. If elements within the predicate are conceptually more cohesive than any of them is with the subject, then it may be more memory-efficient to encode and process information using word orders that maintain contiguity among elements of the predicate and so reflect that organization. In other words, the internal bias is a general cognitive bias to be efficient, rather than domain-specific knowledge as in the strong nativist sense.

Distributional patterns of linguistic features are not completely arbitrary and random. In fact, learners show sensitivity to the ways that linguistic patterns reflect communicative goals or processing and articulatory constraints, even in a language they do not know. This could explain why new language learners behave in a native-like manner in the optional use of grammatical elements (Fedzechkina et al., 2012) and the recognition of phonological preferences (Berent et al., 2011). In terms of word order preferences, according to the World Atlas of Language Structures Online (Dryer, 2013), more than 90% of the world's languages have a canonical word order that shows clear subject-predicate distinctions by grouping elements in the predicate (i.e., the verb and the object) together. In other words, across the world's languages' canonical orders, the subject rarely intervenes between the verb and the object. In line with our hypothesis, these patterns are the potential consequences of our speculation of conceptual structure and the preference for isomorphism. Because scrambling violates the proposed structural preference of a natural cognitive structure, sentences that violate this isomorphism (i.e., subject-crossing sentences) may appear distinctive to naive learners in such a way that points to the inference that a greater variety of surface order variants may be motivated by pragmatic factors and are thus not constrained by verbs. That is, participants in the scrambling group might have detected this structural preference violation that came from subject-crossing sentences and used it to infer that this novel language uses a greater variety of alternations for pragmatic purposes instead. If the word order alternation is primarily the result of pragmatic forces rather than the result of alternation in verbs' argument structure, then learners should not constrain generalization to a certain verb. On this account, this may be why the scrambling group generalized alternations across all verbs.

To be clear, in response to pragmatic accounts, we speculate that when the subject-predicate distinction is disturbed (e.g., OSV structures), it is a salient cue to learners that pragmatic rather than structural factors are of relatively high importance. In cue validity terms, we argue that the lack of a clear subject-predicate boundary is a highly predictive cue (across different languages) of pragmatic forces driving word order variation, rather than verb argument structure. That said, our internal bias account is not incompatible with pragmatic accounts, though it has a different focus. Our main interest lies not in whether discourse factors led to the lack of conservatism, but in what “clued learners in" on the pragmatic forces motivating alternations. The possibility that we offer and are invested in following up on in future work is the use of structures that split up the natural conceptual units of subject-predicate relations.

A possible explanation for the internal bias concerns structural locality. It may be the case that word order alternations are triggered by some structural considerations, from which learners can make reliable inferences that guide their generalization. Possible “triggers" include the verb or a complementizer at the left-periphery of a sentence (which encodes formal features that correlate with pragmatics). The general constraint of any syntactic operations, including the operations that cause word order alternations, is that they must be local (Chomsky, 1973). Therefore, when learners only see local alternations, they may infer that the “trigger” of a local alternation is the verb and as a result constrain their application of this operation to specific verbs. In comparison, when learners see subject-crossing (i.e., non-local) alternations, they are likely to infer that a subject-crossing alternation is triggered by something else to the left of the subject (i.e., the complementizer). Under this account, the “internal bias” is learners’ assumption that syntactic operations must be local.

To further investigate whether subject-crossing is the relevant trigger for generalization of structural alternations, future research may include other unusual structures that preserve the subject-predicate distinction, which should not lead to generalization. One possibility is to have English monolinguals learn a novel language with case-markers that preserve the subject-predicate distinction (i.e., a non-scrambling language). The case-marking feature is an unusual feature to English speakers and is used across all verbs, but our prediction is that this feature alone should not lead to generalization due to the lack of disturbance to the predicate structure. Another possibility is to repeat our experiments with native speakers of a language that allows subject-crossing (e.g., Korean). Our prediction is that the results would be similar to those that we obtained. That is, learners who are only exposed to local alternations in the novel language should be verb-wise conservative, whereas learners who are exposed to subject-crossing should generalize across all verbs. These two observations together would provide strong evidence that subject-crossing is the relevant trigger regardless of the participants' native language.

In conclusion, learning structural alternations is not just a matter of tracking statistics. Rather, learners use linguistically sophisticated biases that mirror typological differences to guide their generalization of input statistics.

References

Baker, C. L. (1979). Syntactic Theory and the Projection Problem. *Linguistic Inquiry* , *10*(4), 533–581.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Mäechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version* 1.7.

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. *The Crosslinguistic Study of Sentence Processing*, *3*, 73-112.

Berent, I., Harder, K., & Lennertz, T. (2011). Phonological universals in early childhood: Evidence from sonority restrictions. *Language Acquisition*, *18*(4), 281–293.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 467-484.

Braine, M. D., & Bowerman, M. (1976). Children's First Word Combinations. *Monographs of the Society for Research in Child Development*, *41*(1), 1–104.

Brooks, P. J., & Tomasello, M. (1999). How Children Constrain Their Argument Structure Constructions. *Language*, *75*(4), 720–738.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart & Winston.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*, 1–61.

Dryer, M. S. (2013). Order of Subject, Object and Verb. Retrieved May 19, 2018, from http://wals.info/chapter/81

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

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131-153.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics*, *15*(3), 3. 289-316.

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

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

Levy, R. (2018). Using R formulae to test for main effects in the presence of higher-order interactions. *arXiv preprint arXiv:1405.2094.*

Osgood, C. E., & Bock, J. K. (1977). Salience and sentencing: Some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Peirce, J. W. (2007). PsychoPy--Psychophysics Software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.

Perek, F., & Goldberg, A. E. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language*, *84*, 108-127.

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276-293.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*, 607–642.

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Ross, J.R., 1967. *Constraints on variables in syntax*. PhD Dissertation. MIT, Cambridge, MA

Speyer, A. (2010). *Topicalization and stress clash avoidance in the history of English*. Berlin: Mouton De Gruyter

Stefanowitsch, A. (2008). Negative Entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, *19*(3), 513-533.

Thothathiri, M., & Braiuca, M. C. (2020). Distributional learning in English: The effect of verb-specific biases and verb-general semantic mappings on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, *7*, 404.

Tomasello, M. (2000). First Steps Toward a Usage-Based Theory of Language Acquisition. *Cognitive Linguistics*, *11*(1/2), 61–82.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Massachusetts: Harvard University Press.

Tomasello, M., & Brooks, P. J. (1998). Young Children's Earliest Transitive and Intransitive Constructions. *Cognitive Linguistics*, *9*(4), 379–396.

Whong-Barr, M., & Schwartz, B. D. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, *24*, 579–616.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the Evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36–48.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and Processing Verb Argument Structure: Distributional Learning in a Miniature Language. *Cognitive Psychology*, *56*(3), 165–209.

Yamashita, H. (2002). Scrambled sentences in Japanese: Linguistic properties and motivations for production. *Text*, *22*, 597–633.

Supplementary Material

The data can be found at https://osf.io/qudy9/.

Appendix A

Full Lists of Exposure Phase and Test Phase Materials

Table 1.3

Exposure Phase Materials (Presented in the Specified Pseudorandomized Order)

Item	Verb	Agent	Theme	Goal
1	give	boxer	balloon	sailor
2	give	chef	camera	painter
3	show	painter	stroller	cowboy
4	give	chef	boat	sailor
5	hand	sailor	ball	boxer
6	give	painter	balloon	robber
7	show	cowboy	tent	boxer
8	give	robber	camera	dancer
9	hand	painter	apple	dancer
10	give	boxer	ring	chef
11	show	lady	badge	robber
12	give	painter	cake	chef
13	show	robber	tent	dancer
14	hand	chef	ball	lady
15	show	dancer	stroller	lady
16	give	lady	boat	dancer
17	give	boxer	crown	chef
18	show	dancer	table	painter
19	give	chef	guitar	cowboy
20	show	painter	badge	chef
21	hand	boxer	book	sailor
22	give	dancer	balloon	chef
23	give	dancer	cake	sailor
24	show	lady	tent	sailor
25	give	cowboy	camera	boxer
26	hand	cowboy	ball	robber
27	hand	chef	cup	painter
28	give	dancer	tie	boxer
29	show	robber	stroller	boxer
30	hand	boxer	apple	lady
31	give	cowboy	crown	robber
32	give	lady	crown	sailor

Item	Verb	Agent	Theme	Goal
33	hand	cowboy	cup	boxer
34	give	robber	guitar	lady
35	hand	dancer	ball	painter
36	give	lady	ring	cowboy
37	hand	sailor	apple	chef
38	show	boxer	table	chef
39	give	sailor	camera	lady
40	show	chef	stroller	sailor
41	hand	painter	book	cowboy
42	give	chef	tie	lady
43	give	boxer	boat	cowboy
44	show	sailor	badge	dancer
45	give	dancer	crown	painter
46	give	painter	guitar	boxer
47	hand	dancer	book	chef
48	show	chef	tent	painter
49	give	sailor	guitar	dancer
50	give	painter	tie	robber
51	show	sailor	table	lady
52	hand	lady	cup	sailor
53	give	robber	boat	painter
54	show	boxer	badge	cowboy
55	hand	robber	cup	dancer
56	give	sailor	ring	dancer
57	hand	lady	book	robber
58	give	cowboy	balloon	lady
59	give	lady	cake	robber
60	hand	robber	apple	cowboy
61	give	cowboy	cake	boxer
62	show	cowboy	table	robber
63	give	sailor	tie	cowboy
64	give	robber	ring	painter

Table 1.4

Test Phase Picture Description Task Materials (Presented in a Fully Randomized Order)

Item	Verb	Agent	Theme	Goal	Prompt	Item	Verb	Agent	Theme	Goal	Prompt
1	give	boxer	carrot	dancer	goal	25	hand	lady	banana	painter	goal
2	give	boxer	shirt	robber	theme	26	hand	lady	ruler	boxer	theme
3	give	chef	carrot	boxer	theme	27	hand	painter	fork	chef	theme
4	give	chef	shirt	sailor	goal	28	hand	painter	ruler	dancer	goal
5	give	cowboy	banana	chef	theme	29	hand	robber	purse	sailor	theme
6	give	cowboy	pen	boxer	goal	30	hand	robber	spoon	boxer	goal
7	give	dancer	fork	painter	theme	31	hand	sailor	shirt	cowboy	goal
8	give	dancer	spoon	cowboy	goal	32	hand	sailor	shirt	lady	theme
9	give	lady	purse	chef	goal	33	show	boxer	ruler	chef	goal
10	give	lady	ruler	sailor	theme	34	show	boxer	fork	painter	theme
11	give	painter	banana	robber	goal	35	show	chef	spoon	sailor	theme
12	give	painter	purse	lady	theme	36	show	chef	pen	robber	goal
13	give	robber	pen	painter	theme	37	show	cowboy	ruler	robber	theme
14	give	robber	ruler	lady	goal	38	show	cowboy	banana	painter	goal
15	give	sailor	fork	dancer	goal	39	show	dancer	purse	chef	theme
16	give	sailor	spoon	cowboy	theme	40	show	dancer	carrot	cowboy	goal
17	hand	boxer	pen	painter	goal	41	show	lady	purse	sailor	goal
18	hand	boxer	spoon	lady	theme	42	show	lady	shirt	dancer	goal
19	hand	chef	banana	dancer	theme	43	show	painter	carrot	lady	goal
20	hand	chef	fork	robber	goal	44	show	painter	shirt	boxer	theme
21	hand	cowboy	pen	robber	theme	45	show	robber	pen	cowboy	theme
22	hand	cowboy	purse	chef	goal	46	show	robber	spoon	dancer	theme
23	hand	dancer	carrot	cowboy	theme	47	show	sailor	banana	boxer	goal
24	hand	dancer	carrot	sailor	goal	48	show	sailor	fork	lady	theme

Table 1.5

Test Phase Acceptability Judgment Task Materials (Presented in a Fully Randomized Order)

Item	Verb	Sentence Type	Sentence
1	give	Canonical Word Order	boxer-ga dancer-aegae carrot-eul joo-utt-dah
2	give	Canonical Word Order	cowboy-ga chef-aegae banana-eul joo-utt-dah
3	give	Canonical Word Order	lady-ga sailor-aegae ruler-eul joo-utt-dah
4	give	Canonical Word Order	robber-ga lady-aegae ruler-eul joo-utt-dah
5	give	Local Alternation	boxer-ga shirt-eul robber-aegae joo-utt-dah
6	give	Local Alternation	dancer-ga spoon-eul cowboy-aegae joo-utt-dah
7	give	Local Alternation	lady-ga purse-eul chef-aegae joo-utt-dah
8	give	Local Alternation	sailor-ga fork-eul dancer-aegae joo-utt-dah
9	give	Subject-crossing Alternation	boxer-aegae cowboy-ga pen-eul joo-utt-dah
10	give	Subject-crossing Alternation	lady-aegae painter-ga purse-eul joo-utt-dah
11	give	Subject-crossing Alternation	cowboy-aegae sailor-ga spoon-eul joo-utt-dah
12	give	Subject-crossing Alternation	boxer-aegae chef-ga carrot-eul joo-utt-dah
13	give	Ungrammatical Word Order	joo-utt-dah robber-ga painter-aegae pen-eul
14	give	Ungrammatical Word Order	joo-utt-dah painter-ga robber-aegae banana-eul
15	give	Ungrammatical Word Order	joo-utt-dah dancer-ga fork-eul painter-aegae
16	give	Ungrammatical Word Order	joo-utt-dah chef-ga shirt-eul sailor-aegae
17	hand	Canonical Word Order	boxer-ga lady-aegae spoon-eul gun-neo-joo-utt-dah
18	hand	Canonical Word Order	cowboy-ga chef-aegae purse-eul gun-neo-joo-utt-dah
19	hand	Canonical Word Order	painter-ga chef-aegae fork-eul gun-neo-joo-utt-dah
20	hand	Canonical Word Order	robber-ga sailor-aegae purse-eul gun-neo-joo-utt-dah
21	hand	Local Alternation	chef-ga banana-eul dancer-aegae gun-neo-joo-utt-dah
22	hand	Local Alternation	cowboy-ga pen-eul robber-aegae gun-neo-joo-utt-dah
23	hand	Local Alternation	lady-ga ruler-eul boxer-aegae gun-neo-joo-utt-dah
24	hand	Local Alternation	robber-ga spoon-eul boxer-aegae gun-neo-joo-utt-dah
25	hand	Subject-crossing Alternation	robber-aegae chef-ga fork-eul gun-neo-joo-utt-dah
26	hand	Subject-crossing Alternation	sailor-aegae dancer-ga carrot-eul gun-neo-joo-utt-dah
27	hand	Subject-crossing Alternation	dancer-aegae painter-ga ruler-eul gun-neo-joo-utt-dah
28	hand	Subject-crossing Alternation	lady-aegae sailor-ga shirt-eul gun-neo-joo-utt-dah

Table 1.5 (Continued)

Test Phase Acceptability Judgment Task Materials (Presented in a Fully Randomized Order)

Item	Verb	Sentence Type	Sentence
29	hand	Ungrammatical Word Order	gun-neo-joo-utt-dah boxer-ga pen-eul painter-aegae
30	hand	Ungrammatical Word Order	gun-neo-joo-utt-dah dancer-ga carrot-eul cowboy-aegae
31	hand	Ungrammatical Word Order	gun-neo-joo-utt-dah lady-ga banana-eul painter-aegae
32	hand	Ungrammatical Word Order	gun-neo-joo-utt-dah sailor-ga cowboy-aegae shirt-eul
33	show	Canonical Word Order	chef-ga robber-aegae pen-eul bo-yeo-joo-utt-dah
34	show	Canonical Word Order	cowboy-ga robber-aegae ruler-eul bo-yeo-joo-utt-dah
35	show	Canonical Word Order	lady-ga dancer-aegae shirt-eul bo-yeo-joo-utt-dah
36	show	Canonical Word Order	sailor-ga lady-aegae fork-eul bo-yeo-joo-utt-dah
37	show	Local Alternation	boxer-ga ruler-eul chef-aegae bo-yeo-joo-utt-dah
38	show	Local Alternation	cowboy-ga banana-eul painter-aegae bo-yeo-joo-utt-dah
39	show	Local Alternation	lady-ga purse-eul sailor-aegae bo-yeo-joo-utt-dah
40	show	Local Alternation	sailor-ga banana-eul boxer-aegae bo-yeo-joo-utt-dah
41	show	Subject-crossing Alternation	sailor-aegae chef-ga spoon-eul bo-yeo-joo-utt-dah
42	show	Subject-crossing Alternation	chef-aegae dancer-ga carrot-eul bo-yeo-joo-utt-dah
43	show	Subject-crossing Alternation	boxer-aegae painter-ga shirt-eul bo-yeo-joo-utt-dah
44	show	Subject-crossing Alternation	dancer-aegae robber-ga spoon-eul bo-yeo-joo-utt-dah
45	show	Ungrammatical Word Order	bo-yeo-joo-utt-dah boxer-ga fork-eul painter-aegae
46	show	Ungrammatical Word Order	bo-yeo-joo-utt-dah dancer-ga carrot-eul cowboy-aegae
47	show	Ungrammatical Word Order	bo-yeo-joo-utt-dah painter-ga carrot-eul lady-aegae
48	show	Ungrammatical Word Order	bo-yeo-joo-utt-dah robber-ga pen-eul cowboy-aegae

Appendix B

Table 1.6

Number of excluded production responses in each condition

Experiment	Non-scrambling Group			Scrambling Group		
	give	hand	show	give	hand	show
Experiment 1	5	8	6	19	21	20
Experiment 2	3	3	3	6	4	1
Experiment 3	9	8	9	21	29	17

Appendix C

Table 1.7

Detailed Summary of Production Responses Across Experiments

			Number of Instances Produced by Group for Each Verb					
Experiment	Response Category	Word Order	Non-scrambling Group			Scrambling Group		
			give	hand	show	give	hand	show
Experiment 1	Canonical Word Order	NOM-DAT-ACC	188	259	219	172	163	147
	Local Alternation	NOM-ACC-DAT	179	105	142	68	72	97
	Subject-crossing Alternation	DAT-NOM-ACC	5	7	8	73	67	67
		DAT-ACC-NOM	1	1	3	20	29	20
		ACC-NOM-DAT	5	2	6	22	22	19
		ACC-DAT-NOM	1	2	0	10	10	14
Experiment 2	Canonical Word Order	NOM-DAT-ACC	283	303	316	288	274	295
	Local Alternation	NOM-ACC-DAT	74	54	40	0	0	0
	Subject-crossing Alternation	DAT-NOM-ACC	9	10	9	75	91	75
		DAT-ACC-NOM	7	6	8	0	0	0
		ACC-NOM-DAT	5	2	5	13	11	12
		ACC-DAT-NOM	3	6	3	2	4	1
Experiment 3	Canonical Word Order	NOM-DAT-ACC	235	284	301	247	253	264
	Local Alternation	NOM-ACC-DAT	140	92	74	0	0	0
	Subject-crossing Alternation	DAT-NOM-ACC	0	0	0	116	102	103
		DAT-ACC-NOM	0	0	0	0	0	0
		ACC-NOM-DAT	0	0	0	0	0	0
		ACC-DAT-NOM	0	0	0	0	0	0

Appendix D

Full Output of Models Across Experiments

Table 1.8

Full Output of Models for the Log-odds of Producing a Given Structure in the Picture Description Task

Experiment	Structure	Predictor	Estimate	*SE*	*z-value*	*p-value*
Experiment 1	Subject-crossing Alternation	(Intercept)	7.138	1.457	4.900	<.001 ***
		Verb1	-0.282	1.601	-0.176	.860
		Verb2	-0.930	2.003	-0.464	.642
		Group	-5.758	2.070	-2.782	.005 **
		Verb1:Group	0.608	1.491	0.408	.683
		Verb2:Group	1.188	1.930	0.616	.538
	Local Alternation	(Intercept)	-2.857	0.659	-4.334	<.001 ***
		Verb1	-0.795	0.984	-0.808	.419
		Verb2	0.729	0.543	1.342	.180
		Group	-1.899	1.225	-1.550	.121
		Verb1:Group	3.633	1.593	2.280	.023 *
		Verb2:Group	-1.367	0.567	-2.411	.016 *
Experiment 2	Learned Alternation	(Intercept)	-10.745	1.467	-7.324	<.001 ***
		Verb1	-1.652	2.350	-0.703	.482
	(Local for the Non-scrambling Group and Subject-crossing for the Scrambling Group)	Verb2	2.749	2.886	0.953	.341
		Group	0.094	1.970	0.048	.962
		Verb1:Group	2.804	2.658	1.055	.291
		Verb2:Group	-3.706	3.244	-1.142	.253

Table 1.8 (Continued)

Full Output of Models for the Log-odds of Producing a Given Structure in the Picture Description Task

Experiment	Structure	Predictor	Estimate	*SE*	*z-value*	*p-value*
Experiment 3	Learned Alternation	(Intercept)	-2.789	0.523	-5.329	<.001 ***
		Verb1	-2.592	0.855	-3.032	.002 **
	(Local for the Non-scrambling Group and Subject-crossing for the Scrambling Group)	Verb2	1.900	0.909	2.089	.037 *
		Group	-0.108	1.011	-0.107	.915
		Verb1:Group	1.351	1.658	0.815	.415
		Verb2:Group	-0.045	1.565	-0.029	.977

Note. Verb1 contrasts "give" and "show"; Verb2 contrasts "hand" and "show". *$p < .05$. **$p < .01$. ***$p < .001$.

Table 1.9

Full Output of Models for Acceptability of a Given Structure in the Acceptability Judgment Task

Experiment	Structure	Predictor	Estimate	*SE*	*t-value*	*p-value*
Experiment 1	Subject-crossing Alternation	(Intercept)	-0.141	0.063	-2.234	.032 *
		Verb1	-0.032	0.124	-0.255	.803
		Verb2	-0.112	0.1234	-0.907	.382
		Group	-0.813	0.120	6.797	<.001 ***
		Verb1:Group	-0.013	0.219	-0.059	.954
		Verb2:Group	-0.287	0.218	-1.321	.208
	Local Alternation	(Intercept)	0.500	0.059	8.477	<.001 ***
		Verb1	-0.335	0.108	-3.108	.003 **
		Verb2	0.237	0.066	3.579	.001 **
		Group	-0.275	0.118	-2.329	.024 *
		Verb1:Group	0.671	0.213	3.132	.003 **
		Verb2:Group	-0.393	0.131	-3.009	.004 **
Experiment 2	Subject-crossing Alternation	(Intercept)	0.090	0.076	1.187	.241
		Verb1	-0.007	0.116	-0.062	.952
		Verb2	-0.095	0.121	-0.791	.445
		Group	0.942	0.137	6.875	<.001 ***
		Verb1:Group	-0.029	0.139	-0.212	.836
		Verb2:Group	-0.070	0.155	-0.454	.656
	Local Alternation	(Intercept)	-0.049	0.058	-0.842	.404
		Verb1	-0.103	0.098	-1.048	.303
		Verb2	-0.003	0.097	-0.028	.978
		Group	-0.902	0.114	-7.946	<.001 ***
		Verb1:Group	0.300	0.183	1.637	.111
		Verb2:Group	0.025	0.180	0.138	.891

Table 1.9 (Continued)

Full Output of Models for Acceptability of a Given Structure in the Acceptability Judgment Task

Experiment	Structure	Predictor	Estimate	*SE*	*t-value*	*p-value*	
Experiment 3	Subject-crossing Alternation	(Intercept)	-0.003	0.062	-0.055	.957	
		Verb1	-0.083	0.113	-0.728	.484	
		Verb2	0.008	0.115	0.067	.948	
		Group	1.153	0.113	10.156	<.001	***
		Verb1:Group	-0.153	0.180	-0.849	.414	
		Verb2:Group	-0.152	0.184	-0.828	.424	
	Local Alternation	(Intercept)	0.085	0.055	1.544	.129	
		Verb1	-0.238	0.115	-2.073	.044	*
		Verb2	0.115	0.096	1.208	.233	
		Group	-1.028	0.110	-9.373	<.001	***
		Verb1:Group	0.498	0.229	2.173	.035	***
		Verb2:Group	-0.091	0.192	-0.475	.637	

Note. Verb1 contrasts “give” and “show”; Verb2 contrasts “hand” and “show”. *$p < .05$. **$p < .01$. ***$p < .001$.

Appendix E

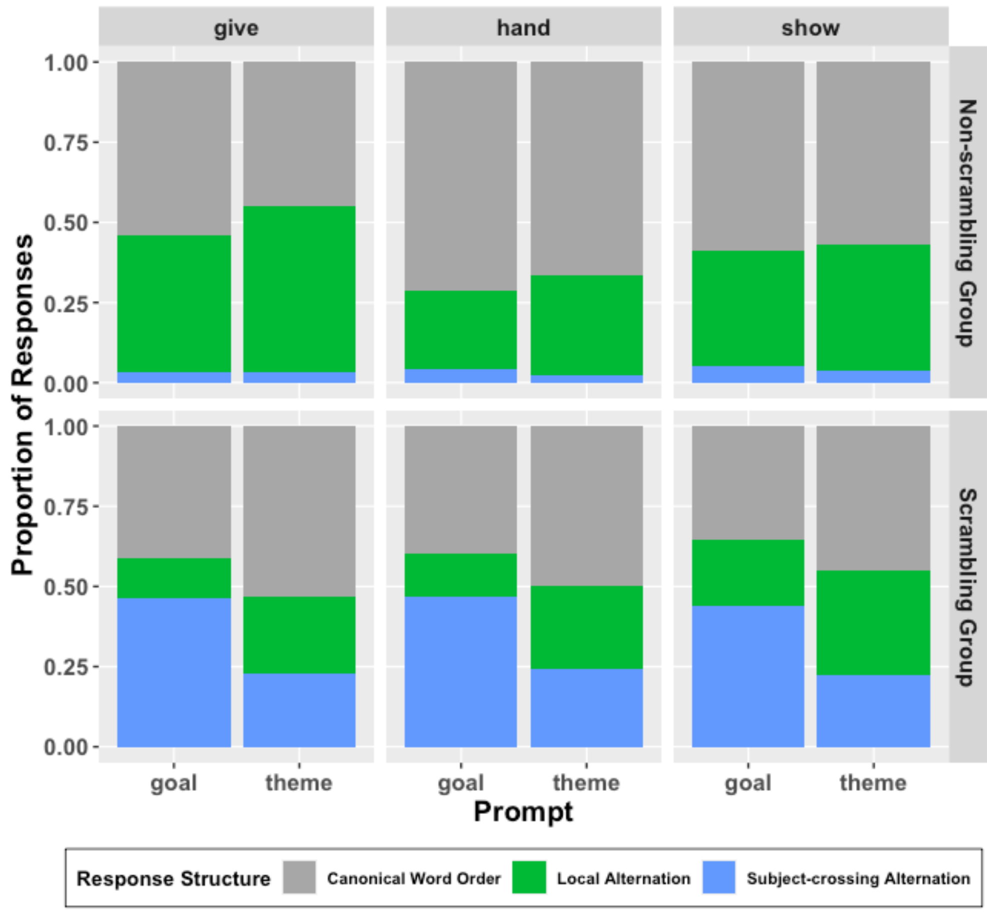

Figure 1.10. Experiment 1 Results Broken Down by Prompt

On average, participants were more likely to produce local alternations when the theme was prompted (34%), compared to when the goal was prompted (25%); the prompt main effect was statistically significant, $\chi^2(1) = 8.47$, $p = .004$. Critically, the verb-wise conservatism in the non-scrambling group and the lack thereof in the scrambling group was still observed, regardless of prompt. That is, the Verb x Group interaction remained significant, $\chi^2(2) = 7.39$, $p = .02$; and there was no Verb x Group x Prompt interaction, $\chi^2(2) = 0.48$, $p = .79$. The scrambling group produced 12.1% more local alternations when the theme rather than the goal was prompted, whereas the non-scrambling group only showed a 6.5% difference; though the Group x Prompt interaction did not reach statistical significance, $\chi^2(1) = 0.28$, $p = .60$.

For subject-crossing alternations, the non-scrambling group barely produced this structure regardless of prompt (4.25% when the goal was prompted and 2.99% when the theme was prompted; $z = -1.36$, $p = .18$). In contrast, the scrambling group was much more likely to produce subject-crossing alternations when the goal was prompted (45.7%), compared to when the theme was prompted (23.1% ; $z = -10.2$, $p < .001$); the Group x Prompt interaction was statistically significant, $\chi^2(1) = 11.39$, $p < .001$.

Altogether, this analysis showed that our main findings of the qualitative difference between groups in generalization patterns could not be reduced to prompt effects.

Acknowledgements

Chapter 1, in full, is a reprint of the material as it appears in Cognition, 2021, Lau, Sin Hang; Momma, Shota; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

CHAPTER 2

Do people structure words and sentences using shared mechanisms?

Sin Hang Lau and Victor S. Ferreira

Author Note

Sin Hang Lau, Department of Psychology, University of California San Diego; Victor S. Ferreira, Department of Psychology, University of California San Diego

Correspondence concerning this article should be addressed to Sin Hang Lau, Department of Psychology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: aubreylau@ucsd.edu

Abstract

The same sequence of linguistic units can be structured differently: Relative clauses in sentences such as “I met the students of the teacher who played the violin” can either have a high attachment (HA; i.e., the students played the violin) or low attachment reading (LA; i.e., the teacher played the violin). Similarly, morphological attachment in noun phrases such as “social psychologist” can have either an HA (i.e., someone who studies social psychology; [[social psycholog(y)][ist]]) or LA reading (i.e., a psychologist who is social; [[social][[psycholog(y)][ist]]]). Thus, abstractly at least, sentences and words have similar internal hierarchical structures. Using a structural priming paradigm, we investigated in three experiments whether shared mechanisms process the internal structures of both sentences and words, despite the difference in grain size. Overall, we only observed priming effects when the primes and targets were of the same grain size: Participants produced more HA sentences or rated the HA readings of ambiguous sentences as more probable following HA sentence primes (compared to following LA sentence primes), while the attachment structure of morphological primes did not affect subsequent sentence production and comprehension. Participants also rated the HA readings of ambiguous noun phrases as more probable following HA morphological primes (compared to following LA morphological primes), while the attachment of sentence primes did not affect subsequent morphological comprehension. We suggest that at least as reflected by structural priming, structural operations are not shared across morphological and syntactic levels.

Keywords: structural priming, attachment ambiguity, domain generality

Do people structure words and sentences using shared mechanisms?

Producing language involves generating temporally or spatially linear sequences of sounds or symbols, which are the products of piecing together smaller units to form bigger units; from morphemes to words, from words to phrases, and from phrases to sentences. These linguistic units are organized in a hierarchical manner, as the functions and meanings of the units are determined by the process of integrating units that are not necessarily adjacent to one another. As such, the same linear sequence of linguistic units can be structured in different ways, and the message or interpretation could differ depending on the specific structure one commits to. For example, in sentences such as “I met the students of the teacher who played the violin”, the relative clause “who played the violin” can either have a high attachment (HA) which means “the students” played the violin, or a low attachment (LA) which means “the teacher” played the violin. Similarly, in noun phrases such as “social psychologist”, the morpheme “-ist” can either attach high to “social psycholog(y)” to mean someone who studies the discipline of social psychology, or attach low to “psycholog(y)” only to mean someone who studies psychology and enjoys the companionship of others. Despite the difference in the grain size of the linguistic units involved, these examples illustrate that words and sentences are at least abstractly similar in their internal structures, in that processing words and sentences both require combining hierarchically organized units. That leads to the question: Are the cognitive mechanisms that process the internal structures of words and sentences shared? Here we first introduce the structural priming paradigm, which has been commonly used to test whether two superficially distinct forms of structures share the same underlying representation or operation. Then, we review previous work on cross-domain and linguistic priming.

Speakers can often choose between multiple structures to convey a similar message, but studies have shown that they are more likely to repeat a sentence structure that they have recently been exposed to (say, in the preceding sentence) than to produce an alternative structure that conveys a similar meaning, even if the previous and the current sentence are not conceptually related. This phenomenon of abstract structure reuse is referred to as *structural priming*. In a picture description task, Bock (1986) found that after participants heard and repeated out loud a passive voice prime sentence (e.g., "The referee was punched by one of the fans"), as opposed to an active voice prime sentence (e.g., "One of the fans punched the referee"), they were more likely to subsequently describe an unrelated target picture in passive voice (e.g., " The church is being struck by lightning"). Similar patterns were also found for other structures, such as between producing a prepositional dative (e.g., "The rock star sold some cocaine to the undercover agent") or a double object dative sentence (e.g., "A rock star sold the undercover agent some cocaine"; see Bock, 1986, 1989, and Bock & Loebell, 1990); and between completing the relative clause sentence fragment "The pensioner railed about the author of the fliers that…" with an HA (describing the author) versus LA continuation (describing the fliers; Scheepers, 2003).

Patterns of structural priming have been replicated regardless of whether participants were exposed to the primes through writing (Pickering & Branigan, 1998), typing (Corley & Scheepers, 2002), or listening to another interlocutor in dialogue (Branigan et al., 2000). That is, structuring operations in language appear to be modality-independent. Generally, the aforementioned studies suggest that structural priming occurs because the representation of the abstract sentence structure in use remains accessible for some time, making that template more likely to be reused compared to its alternatives (see Pickering & Ferreira, 2008 for a review on structural priming and its implications for language models). Thus, when structural priming occurs between two linguistic units that are seemingly different on the surface, the units are argued to likely have a shared underlying structural representation in processing.

Even though structural priming has been found with various linguistic structures, some argue that priming effects on active/passive voice and dative constructions may not be purely structural because the lexical entries of verbs include information about argument structure of the sentence. On the contrary, because relative clauses are modifiers that are not commonly viewed as part of the core argument structure of a lexical item, they are likely not encoded in the lexical entries of the nouns that they modify and thus in some ways provide a more purely structural test case for priming (e.g., Desmet & Declercq, 2006, Scheepers, 2003). However, it is important to note that relative clause attachment priming is qualitatively different from other forms of structural priming. In particular, the two attachment alternatives convey different propositional meanings of "who did what to whom", whereas the message remains somewhat comparable in active/passive and dative alternatives. Because the choice of attachment structure is conflated with the choice of propositional meaning, Pickering and Ferreira (2008) suggest that relative clause attachment priming is at least in part message level or comprehension priming. From a production perspective, attachment manipulations may affect speakers' subsequent choice of meaning because the typical sentence continuation task does not require speakers to commit to expressing one meaning versus the other, making message priming possible. From a comprehension perspective, attachment manipulations may be priming the speakers to comprehend the target fragment with HA or LA before continuing with their production, making the phenomenon a comprehension (rather than production) priming effect. Nevertheless, relative clause attachment priming has been widely used as a test for shared structuring mechanisms in linguistic and non-linguistic domains alike.

Within the language domain, relative clause attachment priming has been used to study whether bilinguals have shared structural representations for the two languages that they speak. For example, Desmet and Declercq (2006) demonstrated cross-linguistic relative clause attachment priming from Dutch to English in Dutch-English bilinguals. Participants first completed a sentence fragment such as “Gabriel scratched on the cover of the magazine that…” in Dutch, where “that” was manipulated to induce an HA or LA attachment with grammatical gender features agreeing with either “the cover” or “the magazine”, respectively. Then, participants completed another fragment in English that did not contain attachment-biasing grammatical features (e.g., “The farmer fed the calves of the cow that…”). The results replicated the structural priming effect: Participants were more likely to produce HA continuations in English following HA Dutch primes, compared to following LA Dutch primes. Similarly, Hartsuiker et al. (2016) reported cross-linguistic priming in Dutch-French-English trilinguals, regardless of whether priming was between the first and second language or two different second languages. The authors took cross-linguistic relative clause attachment priming as evidence that bilinguals may have a single abstract representation for each of the relative clause attachment structures that is shared between the languages they speak (see Van Gompel & Arai, 2018 for a review on structural priming in bilinguals).

Outside of language, humans frequently interact with hierarchical structures in other domains of cognition, which raises the question of whether structuring mechanisms of abstract structures are domain-general. For example, parentheses are used in arithmetic to denote the order and scope of operations, and the resulting configurations of equations sometimes resemble the structure of relative clause sentences. Scheepers et al. (2011) noted these structural similarities and tested whether priming occurs across arithmetic and language. They found that participants were more likely to produce an HA sentence continuation after successfully solving an equation with a structure analogous to an HA relative clause (e.g., 80 - (9 + 1) x 5), compared to an equation analogous to LA (e.g., 80 - 9 + 1 x 5). The observation of cross-domain structural priming suggests that hierarchical structural information (such as the global configuration or the shape of the structure) may be represented at a very high level of abstraction such that it may be shared across domains.

In a later conceptual replication of the math-to-language priming effect (Scheepers & Sturt, 2014), it was shown that the effect is bidirectional (i.e., language-to-math priming also occurs) and extends beyond the sentence level to adjective-noun-noun compounds. Arithmetic equations and adjective-noun-noun compounds are similar in that they can both be organized as left- (i.e., (A B) C structure) or right-branching (i.e., A (B C) structure), such as "5 x 2 + 7" versus "5 + 2 x 7" in math, and "organic coffee dealer" (meaning "a dealer of organic coffee" rather than "a coffee dealer who is organic") versus "bankrupt coffee dealer" (meaning "a coffee dealer who is bankrupt" rather than "a dealer of bankrupt coffee") in adjective-noun-noun compounds. In a series of experiments where participants solved math equations and gave sensibility ratings (on a scale of 1 "makes no sense" to 5 "makes perfect sense") for the linguistic compounds, it was found that sensibility ratings for linguistic items were higher after successfully solving an equation of a congruent branching structure, compared to an incongruent one, replicating the math-to-language priming effect. Moreover, in a sample of participants that were predetermined to be relatively less adept in math, arithmetic equations were more likely to be successfully solved after giving sensibility ratings for linguistic components of a congruent branching structure, compared to an incongruent one, demonstrating a language-to-math priming effect. Once again, these results corroborate the claim that structural representation may be domain-general.

Note, however, that both Scheepers et al. (2011) and Scheepers and Sturt (2014) pointed out that math priming effects are sensitive to many extraneous factors, such as whether the instructions of the task explicitly remind participants of the rules of arithmetic order of operations, whether there are redundant parentheses in the math equations, and the participants' baseline math abilities. Nevertheless, Van de Cavey and Hartsuiker (2016) further supported the cross-domain shared structural representation account by providing evidence for both within- and cross-domain priming among musical sequences, math, structured descriptions of events, and relative clause sentences.

If structuring mechanisms are shared across cognitive domains, as suggested by the ubiquitous cross-domain priming effects found across different tasks and linguistic structures, it seems reasonable to assume that the mechanisms may also be shared within the language domain across units of different grain sizes (i.e., across words and sentences). There is suggestive evidence from priming between math and language that this assumption may be true. Recall that Scheepers et al. (2011) found that math primed relative clause sentences, whereas Scheepers and Sturt (2014) found bidirectional priming between math and compounds of words. When combined together, these two studies suggest that relative clause sentences and compounds of words may prime each other, despite the difference in grain size. The current work aims to test that exact prediction and confirm whether words and sentences indeed have shared structuring mechanisms.

Experiment 1 tested whether attachment in relative clause sentence production and free recall of noun phrases affected subsequent relative clause sentence production; Experiment 2 tested whether attachment in sentence production and noun phrase interpretation affected subsequent noun phrase interpretation; and lastly, Experiment 3 simultaneously tested whether attachment in sentence and noun phrase interpretation affected subsequent sentence and noun phrase interpretation. If structuring mechanisms across different linguistic levels are shared, we should expect preferences in relative clause attachment in sentences and morphological attachment in words to affect subsequent attachment preferences both within- and across-grain size. Conversely, if structuring mechanisms are not shared, we should expect sentence attachment to only affect subsequent sentence attachment but not morphological attachment, and vice versa.

Experiment 1

Data Availability

The data and analysis scripts can be found at https://osf.io/jrmkw/.

Method

Participants

Sixty undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Twelve participants were excluded from our analyses because they either did not produce more than 50% usable data (exclusion criteria explained below in Coding and Data Analysis) or did not complete the experiment due to technical issues. All 48 remaining participants indicated that they were native English speakers.

Materials

The experiment was hosted on Qualtrics. We used a within-subject 2 (Prime Attachment: High or Low) x 2 (Prime Type: Sentence or Morphological) design. Each trial contained one prime and one target. That is, for each trial, participants either completed a sentence fragment prime (Sentence condition) or memorized a morphological prime (Morphological condition), which either has a predominantly high (HA) or low attachment (LA) meaning, before completing the target sentence continuation task (see Procedure section below). Both the prime and target sentence fragments were adopted from Desmet and Declercq (2006) and Scheepers et al. (2011), whereas the morphological primes were designed for the current work. There were 24 sentence primes and 24 morphological primes, with an equal split of HA and LA in each prime type. There were 48 target items in total (i.e., each participant completed 48 critical trials in this experiment). Participants were randomly assigned to one of two experimental lists, which counterbalanced whether a given target item was preceded by an HA or LA prime.

The sentence primes were sentence fragments with agreement features that coerced the production of a high or low attachment continuation, such as "The firemen saved the residents of the penthouse *who were*… (HA; coercing a continuation about the residents)/*that was*… (LA; coercing a continuation about the penthouse)". For morphological primes, we constructed noun phrases of professions that either had predominantly high or low attachment meanings. For example, "primate researcher" was considered an HA item, with the morpheme "-er" attaching to a higher and larger unit of "primate research" to mean someone who studies primates (i.e., [[primate research][er]]) rather than a primate that conducts academic studies (i.e., [[primate][[research][er]]]); whereas "diligent researcher" was considered an LA item, with "-er" attaching low to "research" to mean a hardworking person who conducts academic studies (i.e., [[diligent][[research][er]]]) rather than someone who studies a discipline called "diligent research" (i.e., [[diligent research][er]]).

Similar to the sentence primes, the target items were also sentence fragments, except that they did not contain agreement features that biased attachment preferences. For example, in the fragment "the florist befriended the servant of the princesses who…", participants could freely choose to describe the servant (HA) or the princesses (LA).

In addition to the critical trials, participants also completed 24 filler trials, which were interleaved with the target trials in a pseudorandomized order such that participants did not complete more than two trials of the same type in a row. The filler primes were mathematical primes adapted from the original study that reported the math priming effect (Scheepers et al., 2011), with half of them being arithmetic equations with HA structures (e.g., "31 + (8 – 5) × 2") and the other half with LA structures (e.g., "31 + 8 – 5 × 2"). Similar to critical trials, the target items were sentence fragments that allowed participants to freely produce a high or low attachment continuation. An analysis of the filler items revealed that we failed to replicate the math priming effect. That is, participants were about equally likely to produce a HA continuation regardless of whether they successfully solved an HA or LA arithmetic problem beforehand (42.4% vs. 42.2%; $\beta = -0.01$, $SE = 0.16$, $z = -0.06$, $p = .95$). Given the main goal of the current work to establish whether there are shared cognitive mechanisms between linguistic units of different grain size rather than across domains, we did not continue to include the mathematical fillers in later experiments and will focus our discussion below on any potential linguistic priming effects.

The full list of prime and target materials for Experiment 1 can be found in Appendix A.

Procedure

Participants completed the experiment hosted on Qualtrics in the laboratory under the supervision of an experimenter. All responses were produced by typing.

Each trial consisted of a prime and a target task. The prime was either a sentence or a morphological prime, which was manipulated to have either a high (HA) or low (LA) attachment structural organization. Participants were given instructions to perform a sentence continuation task and a noun phrase free recall task, but they were not informed of the prime-target manipulation. For sentence continuation, participants were instructed to type a response that formed a complete sentence whenever they were presented with a sentence fragment. For free recall, participants were told to expect to see a noun phrase displayed on the screen for three seconds and to memorize the given noun phrase in preparation for free recall at a later time. Free recall for each morphological prime was always prompted immediately after the associated target item. In other words, when the participant completed a trial with a sentence prime, they first completed a sentence with attachment-biasing agreement features, then another sentence without such features. If the participant completed a trial with a morphological prime, they first saw a noun phrase and attempted to keep it in their working memory while completing a sentence without attachment-biasing agreement features, then recalled the noun phrase afterwards.

Coding and Analyses

Every prime response that was incompatible with our intended prime attachment manipulation (i.e., when participants ignored the sentence agreement features or inaccurately recalled the memory items) was deemed invalid and left out of our analyses. Additionally, target responses in which it was unclear whether the relative clause modified the high or low attachment noun phrase were also excluded. There were 1,782 (out of 2,304; 77%) analyzable trials in total. Valid target responses were coded in terms of whether the freely produced relative clause attached high (HA) or low (LA).

We built a generalized linear mixed effects model with the dependent variable being whether the target response belonged to HA (1) or LA (0), in order to analyze the likelihood of HA sentence continuation as a function of Prime Type (sentence or morphological), Prime Attachment (HA or LA), and their interaction. We used the maximal random effects structure for both models, which included intercepts for participants and items, as well as the by-subject random slopes for the main effect of Prime Type and Prime Type x Prime Attachment interaction and a by-item random slope for counterbalancing lists.

Results and Discussion

Figure 2.1 shows the average proportion of HA responses following HA or LA sentence or morphological primes. On average, participants were more likely to produce a HA sentence following a HA prime, compared to a LA prime (the main effect of Prime Attachment was significant; $\beta = -0.22$, $SE = 0.11$, $z = -2.05$, $p = .04$). However, a closer look at the data revealed that this was likely driven by the sentence primes; there was a marginal Prime Type x Prime Attachment interaction ($\beta = -0.42$, $SE = 0.23$, $z = -1.84$, $p = .07$). Participants were no more likely to produce HA sentences following HA morphological primes ($M = 49.2\%$, $SE = 2.47\%$), compared to following LA morphological primes ($M = 49\%$, $SE = 2.94\%$); the pairwise comparison was not significant ($z = 0.10$, $p = .92$). However, participants were 9.3% more likely to produce HA sentences following HA sentence primes ($M = 46.2\%$, $SE = 3.38\%$), compared to following LA sentence primes ($M = 36.9\%$, $SE = 3.27\%$), and this pairwise comparison was significant ($z = 2.72$, $p = .006$).

In sum, our results from Experiment 1 suggest that only sentence (but not morphological) attachment preference affected later sentence production. This suggests that the structuring mechanisms between the two levels may not be shared. However, the lack of morphological effect on later sentence production may be attributed to one limitation of the design: The free recall task did not require the participants to think about the possible meanings of the morphological primes. If the hierarchical structures of morphemes that give rise to different meanings of noun phrases were not activated by the free recall task, the absence of priming effects in the morphological conditions may reflect this limitation, rather than a distinct structuring mechanism for words. To further test the robustness and directionality of the priming effects (or the lack thereof), we changed how the morphological primes were presented and switched the target task to a morphological task in Experiment 2.

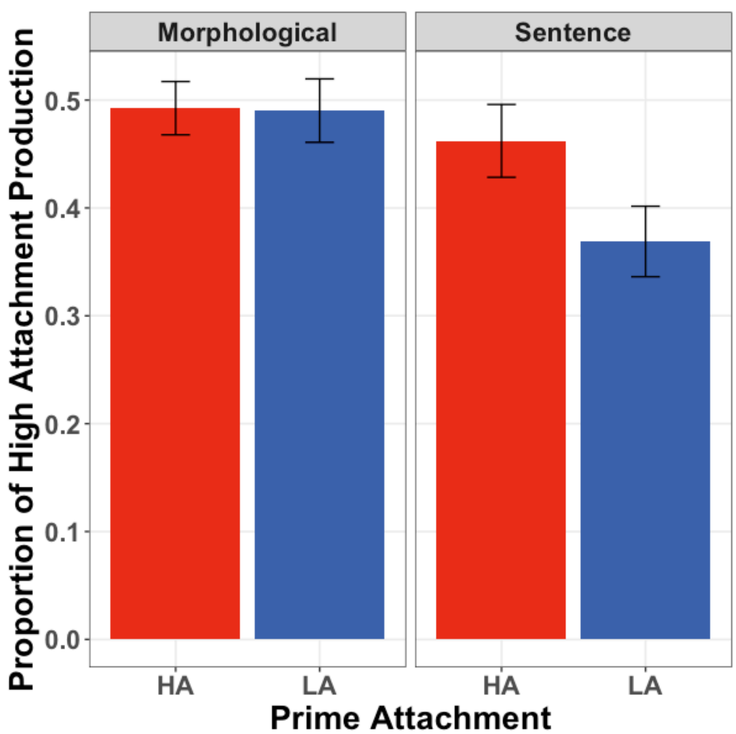

Figure 2.1. Proportion of high attachment sentence continuations following high- or low-attachment morphological or sentence primes, color-coded by prime attachment and separated into two panels by prime type. Error bars represent standard errors.

Experiment 2

Method

Participants

Two hundred and fifty-two undergraduates from the University of California San Diego participated in exchange for course credit. All participants indicated that they were native English speakers. Experiment 2 contained fewer trials than did Experiment 1, so a power analysis was conducted using data from a pilot experiment to determine the target sample size. The analysis indicated that 250 participants would be sufficient to achieve 80% power in this experiment. All participants produced more than 50% valid data for analysis (see Coding and Analyses for criteria) and thus no one was excluded.

Materials

Similar to Experiment 1, we used a within-subject 2 (Prime Attachment: High or Low) x 2 (Prime Type: Sentence or Morphological) design, with each trial containing a prime and target pair. The sentence primes were sentence fragments with agreement features that coerced either a high (HA) or low attachment (LA) continuation, whereas the morphological primes were noun phrases of professions that had predominantly HA or LA interpretations. The prime materials in Experiment 2 were a subset of the ones used in Experiment 1.

The target items were noun phrases that were judged to be relatively ambiguous in a pilot experiment. On a 1-7 Likert scale (with 1 meaning the given noun phrase can only have the LA interpretation, and 7 only the HA interpretation), we selected eight noun phrases (out of 48 in the pilot) with mean ratings closest to the midpoint of the scale (mean ratings ranged from 3.19 to 4.65). These items also happened to be the items with the relatively high variance in ratings (standard deviations ranged from 1.97 to 2.52). These two features of the selected items maximized the chance of the interpretations of these noun phrases being influenced depending on the preceding prime, if there was indeed any priming effect.

Due to the difficulty in finding relatively ambiguous noun phrases, there were only eight trials in Experiment 2. That is, there were four sentence and four morphological primes (equal split of HA and LA), paired with eight morphological target items. Participants were randomly assigned to one of four counterbalancing lists, in which the Prime Type and Prime Attachment for each target item were counterbalanced across lists. The full list of prime and target materials for Experiment 2 can be found in Appendix B.

Procedure

The experiment was a typed experiment hosted on Qualtrics. Participants completed the experiment online without supervision.

Each trial consisted of a prime and a target task. The prime was either a sentence or a morphological prime, which was manipulated to have either a high (HA) or low (LA) attachment structural organization. Participants were given instructions to perform a sentence continuation task and a semantic judgment task, but they were not informed of the prime-target manipulation. For sentence continuation, participants were instructed to type a response that formed a complete sentence whenever they were presented with a sentence fragment. This task served as the sentence prime. For semantic judgment, participants were instructed to rate the likelihood of the possible meanings of noun phrases on a 1-7 Likert scale (with 1 meaning the given noun phrase can only have the LA interpretation, and 7 only the HA interpretation). When participants were asked to rate noun phrases with predominant meanings, this task served as the morphological prime; when participants were asked to rate relatively ambiguous noun phrases, this task served as the target task.

To summarize, when the participant completed a trial with a sentence prime, they first completed a sentence with attachment-biasing agreement features, then gave a semantic judgment for an ambiguous noun phrase. If the participant completed a trial with a morphological prime, they first rated a noun phrase with a predominant meaning, then rated an ambiguous noun phrase. Because the semantic judgment task specifically required participants to think about the possible meanings of the morphological primes and targets, it resolved the limitation presented by Experiment 1.

Coding and Analyses

Prime responses that were incompatible with our intended prime attachment manipulation (i.e., when participants ignored the sentence agreement features, rated LA primes as 4 or above, or HA primes as 4 or below) were deemed invalid and left out of our analyses, leaving 1,661 (out of 2016; 82%) analyzable trials in total. The ratings for valid target responses (1-7) were then transformed into z-score units using by-subject means and standard deviations to account for individual differences in how participants use scales and susceptibility to priming.

We built a linear mixed effects model with the dependent variable being the rating for each trial in z-score units, in order to analyze the attachment preference for ambiguous noun phrases as a function of Prime Type (sentence or morphological), Prime Attachment (HA or LA), and their interaction. We used the maximal random effects structure for both models, which included intercepts for participants and items, as well as the by-subject random slopes for the Prime Type x Prime Attachment interaction and a by-item random slope for counterbalancing lists.

Results and Discussion

Figure 2.2 shows the mean rating for ambiguous noun phrases (in z-score units) following HA or LA sentence or morphological primes. On average, participants rated ambiguous noun phrases following HA primes 0.09 z-score units higher than following LA primes (the main effect of Prime Attachment was significant; $\beta = -0.09$, $SE = 0.05$, $z = -1.96$, $p = .05$). However, this was likely driven by the morphological primes; there was a significant Prime Type x Prime Attachment interaction ($\beta = 0.24$, $SE = 0.10$, $z = 2.54$, $p = .01$). Participants rated ambiguous noun phrases similarly following HA sentence primes ($M = -0.02$, $SE = 0.19$), compared to following LA sentence primes ($M = 0.01$, $SE = 0.19$); the pairwise comparison was not significant ($t = -0.03$, $p = .78$). In contrast, participants rated ambiguous noun phrases 0.21 z-score units higher following HA morphological primes ($M = 0.16$, $SE = 0.19$), compared to LA sentence primes ($M = -0.05$, $SE = 0.19$), and this pairwise comparison was significant ($t = 2.30$, $p = .04$).

To summarize, Experiment 2 showed that only morphological (but not sentence) attachment preference affected later ambiguous noun phrase interpretation. Along with the results in Experiment 1, our results suggest that structuring mechanisms are not shared across grain size.

One alternative interpretation for the results of Experiment 1 and 2 is that the patterns could be explained by task priming (i.e., production priming production; comprehension priming comprehension) rather than priming within linguistic units of the same grain size. Specifically, Experiment 1 showed that sentence production (but not morphological recall) primed sentence production, and Experiment 2 showed that morphological comprehension (but not sentence production) primed morphological comprehension. In other words, task type and grain size were confounded. To control for the possibility of task priming, we ensured that all tasks were comprehension tasks in Experiment 3. Additionally, the target task was always sentence production in Experiment 1 and noun phrase semantic judgment in Experiment 2 regardless of the prime type, which led to participants experiencing more trials of a particular grain size in each experiment. In Experiment 3, we assessed semantic judgment of both sentences and noun phrases in the target task.

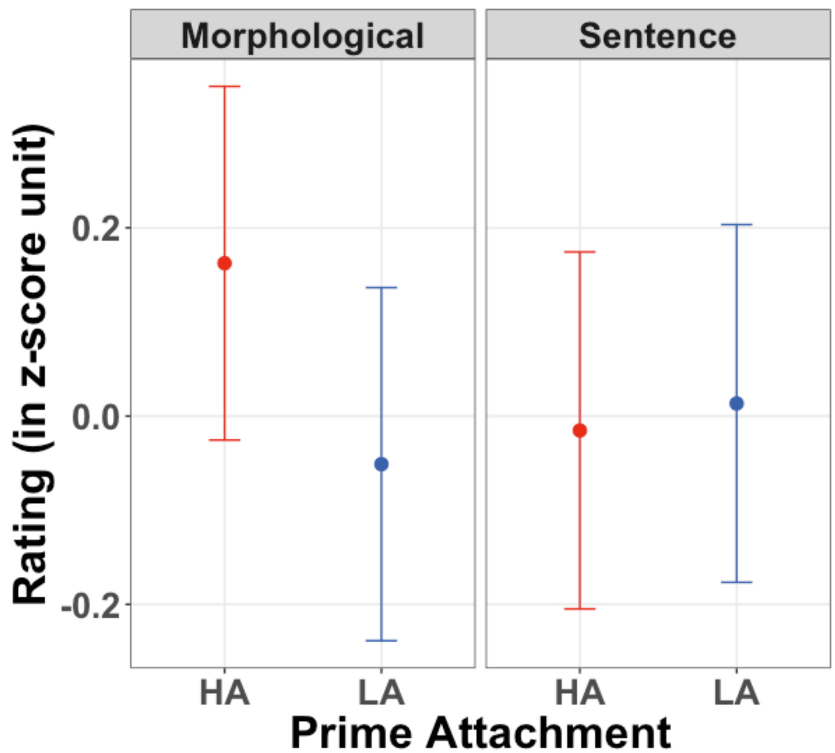

Figure 2.2. Mean rating (in z-score unit) for ambiguous noun phrases following high- or low-attachment morphological or sentence primes, color-coded by prime attachment and separated into two panels by prime type. Error bars represent standard errors.

Experiment 3

Method

Participants

Two hundred and sixty-two undergraduates from the University of California San Diego participated in exchange for course credit. Ten participants were excluded who either did not produce more than 50% valid data (see Coding and Analyses for criteria) or otherwise did not complete the experiment due to technical issues. The remaining two hundred and fifty-two participants indicated that they were native English speakers.

Materials

To address concerns of task priming and imbalanced numbers of trials for different grain sizes in previous experiments, we used a within-subject 2 (Prime Attachment: High or Low) x 2 (Prime Type: Sentence or Morphological) x 2 (Prime-Target Pairing: Same or Different grain size) design. Consistent with previous experiments, each trial contained a prime and target pair, and the prime was either a sentence or a noun phrase of an HA or LA organization. Unlike previous experiments, the grain size of the linguistic unit in the target was not held constant. Instead, it was an equal split between sentence and noun phrase targets. That is, the Same Prime-Target Pairing condition included trials in which the prime and target were both sentences, or both noun phrases; the Different Prime-Target Pairing condition included trials where a noun phrase target was primed by a sentence, or where a sentence target was primed by a noun phrase.

The sentence primes were similar to the materials in Experiment 1 and 2. Instead of sentence fragments, we presented participants with complete sentences that utilized agreement features to coerce HA or LA meanings in Experiment 3, such as "The boy teased the hamsters of the girl that were running around (HA)/who was an animal advocate (LA)". The morphological primes were identical to the materials of Experiment 2 (i.e., noun phrases with predominantly HA or LA meanings).

The sentence targets were complete sentences with ambiguous agreement features, such that both HA and LA meanings were plausible. For instance, in "someone shot the servant of the actress who was on the balcony", it is unclear who was on the balcony (i.e., both "the servant" and “the actress” are plausible). The morphological targets were relatively ambiguous noun phrases identical to the materials in Experiment 2.

There were eight sentence primes and eight morphological primes, with an equal split between HA and LA structures. They were paired with eight sentence and eight morphological targets. That is, each participant completed 16 trials in total. Participants were randomly assigned to one of four counterbalancing lists, in which the prime type and attachment that preceded each target were counterbalanced. The full list of prime and target materials for Experiment 3 can be found in Appendix C.

Procedure

The experiment was a typed experiment hosted on Qualtrics. Participants completed the experiment online without supervision.

The task in Experiment 3 was a semantic judgment task, regardless of whether participants were responding to a sentence or a noun phrase, and whether the item served as a prime or a target. Participants were instructed to first memorize the item on the screen, then to perform a semantic judgment task with a 1-7 Likert scale (with 1 meaning the LA interpretation was possible and 7 meaning only the HA interpretation was possible) on the next page. For example, when presented with the HA sentence prime “the boy teased the hamsters of the girl that were running around”, participants first attempted to memorize the sentence and advanced to the next page when ready. Then, when prompted with the question “Who was/were running around?”, participants rated on a scale of 1-7 the likelihood of the interpretation being “the girl” (LA; 1 on the scale) or “the hamsters” (HA; 7 on the scale). In other words, in the Same Prime-Target Pairing condition, participants rated an unambiguous sentence prime (HA or LA meaning) followed by an ambiguous sentence target, or rated an unambiguous noun phrase prime

(HA or LA) followed by an ambiguous noun phrase target. In the Different Prime-Target Pairing condition, participants rated an unambiguous sentence prime followed by an ambiguous noun phrase target, or rated an unambiguous noun phrase prime followed by an ambiguous sentence target. Because both prime types operated under a comprehension task, this design helped us rule out the possibility of task-priming effects. If the structuring mechanisms were indeed not shared across grain sizes, we should only observe priming effects in the Same but not Different Prime-Target Pairing conditions.

Coding and Analyses

Prime responses that were incompatible with our intended prime attachment manipulation (i.e., when participants rated LA primes as 4 or above, or HA primes as 4 or below) were deemed invalid and left out of our analyses, leaving 3653 (out of 4032 trials; 90%) analyzable trials in total. The ratings for valid target responses (1-7) were then transformed into z-score units using by-subject means and standard deviations to account for individual differences in how participants use scales and susceptibility to priming.

In the main analysis, we built a linear mixed effects model with the dependent variable being the rating for each trial in z-score units, in order to analyze the attachment preference of ambiguous target items as a function of Prime Attachment (HA or LA), Prime-Target Pairing (Same or Different), and their interaction. The random effects structure included intercepts for participants and items, as well as the by-subject random slopes for the Prime Attachment x Prime-Target Pairing interaction and a by-item random slope for counterbalancing lists. This analysis allowed us to confirm whether there was a baseline attachment priming effect, and whether such an effect was unique to linguistic units of the same grain size.

In a secondary analysis, we added to the original linear mixed effects model. The dependent variable was still the rating for each trial in z-score units, whereas the independent variables were the main effects of Prime Type (Sentence or Morphological), Prime Attachment (HA or LA), Prime-Target Pairing (Same or Different), as well as all two-way and three-way interactions. The random effects included intercepts for participants and items. No random slopes were included due to convergence issues. This analysis lent insights into whether the priming effects (if any) differed in strength for sentence versus morphological items.

Results and Discussion

Figure 2.3 shows the mean ratings for ambiguous target items (in z-score units) following HA or LA primes of same or different prime-target pairing. On average, when the prime and target were of the same grain size, the ratings were 0.17 z-score units higher than when the prime and target were of different grain sizes (the main effect of Prime-Target Pairing was significant; $\beta = 0.17$, $SE = 0.03$, $t = 5.16$, $p < .001$).

To address whether attachment priming was specific to linguistic units of the same grain size only, we turned to the main effect of Prime Attachment and the Prime Attachment x Prime-Target Pairing interaction. The average ratings for target items following HA and LA primes were comparable (i.e., no main effect of Prime Attachment; $\beta = -0.03$, $SE = 0.03$, $t = -0.85$, $p = .40$). However, when grain size was taken into account, we found that priming effect was present among linguistic units of the same grain size but not among units of different grain sizes (i.e., a significant Prime Attachment x Prime Target Pairing interaction; $\beta = -0.20$, $SE = 0.07$, $t = -2.89$, $p = .01$). Specifically, when the prime and the target were of the same grain size, there was a 0.12 z-score units difference in ratings for target items following HA primes ($M = 0.14$, $SE = 0.09$) versus LA primes ($M = 0.02$, $SE = 0.09$); the pairwise comparison was significant ($z = 2.73$, $p =$

.01). However, when the prime and the target were of different grain sizes, the ratings following HA (*M* = -0.12, *SE* = 0.09) and LA primes (*M* = -0.05, *SE* = 0.09) were comparable; the pairwise comparison was not significant (z = -1.62, p =.10).

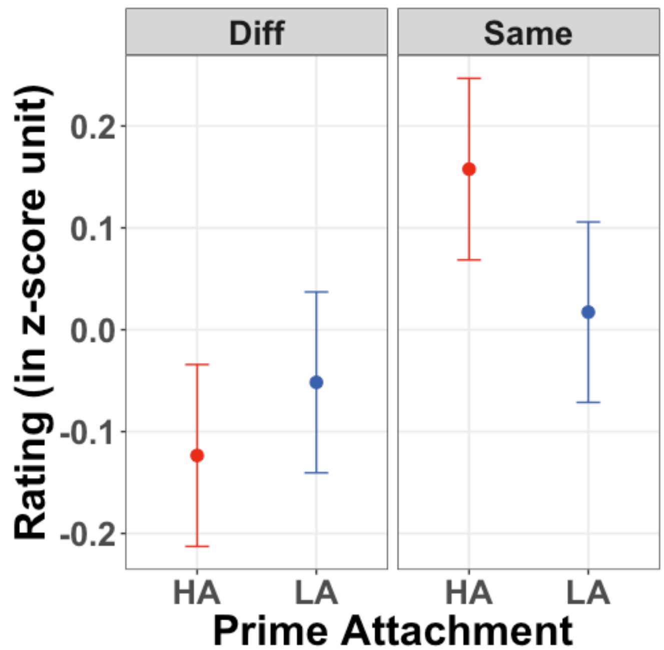

Figure 2.3. Mean rating (in z-score unit) for ambiguous target items following high- or low-attachment primes, color-coded by prime attachment and separated into two panels by prime-target pairing. Error bars represent standard errors.

In our secondary analysis, we added the variable of Prime Type into the model to examine whether the priming effect was modulated by whether the prime was a sentence or a noun phrase. Figure 2.4 further breaks down Figure 2.3 into the two prime types. We only found a marginal Prime Type x Prime-Target Pairing interaction (β = -0.70, *SE* = 0.36, t = -1.94, p = .07), such that the average ratings for morphological targets were 0.54 z-score units higher when the primes were also morphological, compared to when the primes were sentences (z = -2.01, p = .003). The ratings for sentence targets were comparable when primes were of the same or different grain size (-1.70 z-score units difference; z = 0.92, p =.36).

Critically, there was no main effect of Prime Type, Prime Attachment x Prime Type interaction, nor Prime Attachment x Prime-Target Pairing x Prime Type interaction. In other words, there was no evidence to suggest that the priming effects we reported were dependent on whether the prime was a sentence or a noun phrase. Taken together, we found evidence suggesting that attachment priming most likely only occurs when the prime and target were linguistic units of the same grain size. That is, structuring mechanisms are likely not shared between the sentence and the morphological level.

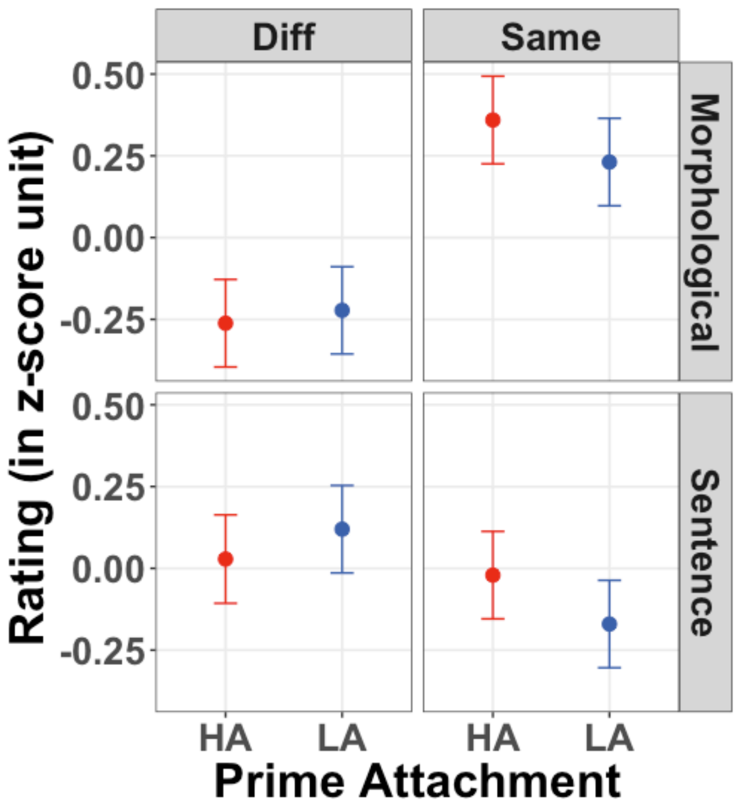

Figure 2.4. Mean rating (in z-score unit) for ambiguous target items following high- or low-attachment primes, color-coded by prime attachment and separated into four panels by prime-target pairing and prime type. Error bars represent standard errors.

General Discussion

In three experiments, the current work investigated whether the structuring mechanisms for linguistic units of different grain sizes are shared. Specifically, we tested whether relative clause attachment in sentences and morphological attachment in noun phrases shared the same structuring mechanisms and thus had priming effects on each other. We used a variety of tasks throughout the experiments, but priming was generally operationalized throughout as whether participants were more likely to produce or interpret a relatively attachment-neutral target item with high attachment (HA) bias following the production or comprehension in response to an HA prime, compared to following producing or interpreting a low attachment (LA) prime. Based on previous work on structural priming in relative clause sentences (e.g., Desmet & Declercq, 2006; Scheepers, 2003), we expected to replicate this priming effect within the same grain size at least on the sentence level. At the time of the completion of our work, we were not aware of any previous work showing within-domain priming on the morphological level among ambiguous noun phrases (only cross-domain priming from math to adjective-noun-noun compounds in Scheepers & Sturt, 2014). However, we expected to see the priming effect from noun phrases to other noun phrases if such an effect extends beyond the sentence level (i.e., priming should be observed in any linguistic units that indeed have similar underlying structures despite differences in surface features and meanings). Critically, if the structuring mechanisms were shared across grain sizes, we should see attachment priming from sentences to noun phrases, and vice versa. If the mechanisms were distinct, then we should see within-level priming as described above only.

Experiment 1 examined whether attachment preferences in relative clause sentence production and in free recall of noun phrases affected subsequent relative clause sentence production. The results showed that participants were more likely to produce HA relative clause sentences following HA primes (compared to following LA primes), but a marginal interaction provided suggestive evidence that the effect was most likely driven by the sentence primes. That is, we replicated the relative clause priming effect reported in the literature but found no strong evidence that noun phrases could prime sentence production. However, to establish that there is no cross-grain size priming, we needed to establish a double dissociation. Moreover, it was unclear at that point whether noun phrases could produce priming effects at all, even within-level. If there is no cross-grain size priming, the result should be bi-directional.

Experiment 2 examined whether attachment preferences in relative clause sentence production and in noun phrase comprehension affected subsequent noun phrase comprehension. The results demonstrated that participants rated the HA interpretation of an ambiguous noun phrase as more probable following an HA prime (compared to following an LA prime), but the patterns were mostly driven by the morphological primes. These patterns are important in two different ways: First, we presented novel evidence showing that structural priming effects within the linguistic domain extend beyond the sentence level to ambiguous noun phrase comprehension. Second, morphological (but not sentence) attachment preferences affected subsequent morphological attachment preferences. Crucial to the question of the sharedness of structuring mechanisms across linguistic units of different grain sizes, the combination of the results in Experiment 1 and 2 resembled a double dissociation, suggesting that the mechanisms are distinct.

However, Experiment 1 and 2 may have suffered from the potential confound of task priming. Because the modalities of the tasks for the two linguistic levels were different, our results could have been explained by preferences in production priming production but not comprehension, and vice versa, regardless of grain size. Note, though, that relative clause priming could be in part due to comprehension or message level priming (Pickering & Ferreira, 2008), and the two possibilities have different implications on the concerns of modality or task priming. Even though sentence continuation is a production task, it could be the comprehension of the fragment before production that is being primed. If all our tasks were essentially comprehension in nature (i.e., no difference in the modality being primed between tasks), modality priming is not a likely explanation of observing within- but not across-level priming. Another possibility is that the nature of the effect is message level priming in sentence continuation but comprehension priming in semantic judgment. Because sentence continuation requires a low commitment to conveying one meaning versus the other while semantic judgment requires a direct comparison between two meanings against each other, the difference in task requirements may explain the lack of across-level priming in Experiment 1 and 2 (i.e., task priming). As such, Experiment 3 sought to confirm the double dissociation and to eliminate the possibility of task priming by holding the modality of the tasks across grain sizes constant to more similar comprehension tasks.

Experiment 3 compared whether attachment preferences in sentence and morphological comprehension affected subsequent comprehension preferences in the same versus different linguistic level. The results showed that attachment priming was indeed only present in linguistic units of the same grain size, but not different. Further analysis found no evidence suggesting any difference in priming magnitude for different grain sizes, meaning that the results are unlikely to be driven by one particular grain size. Additionally, all tasks were comprehension in nature, so the results could not be reduced to modality or task priming. Altogether, the results from three experiments suggested that structuring mechanisms for linguistic units of different grain sizes are not shared.

Given the abundance of evidence in cross-domain priming in the literature (e.g., Scheepers et al., 2011; Scheepers & Sturt, 2014; Van de Cavey & Hartsuiker, 2016), it is somewhat surprising that we failed to observe cross-grain size priming within the linguistic domain. Perhaps, this discrepancy highlights the possibility that structuring mechanisms are only shared across domains in very abstract and general terms (i.e., global configuration and shape). Moreover, structural priming may also be highly sensitive to nuances in processing. For example, Scheepers et al. (2011) did not observe math priming in participants who were explicitly reminded of the operator-precedence rules before the experiment, which may have encouraged them to strategically direct their attention towards searching for the multiplication and division operators. The authors speculated that explicit instruction dampened priming effects due to these effects being relatively dependent on implicit processing. In terms of the online processing dynamics of words versus sentences, it is possible that high frequency noun phrases and words have “larger chunks” of stored meanings, whose semantics are not computed by assembling each morpheme on-the-fly in ways comparable to how comprehenders identifies what entity a relative clause modifies in real-time. Thus, low frequency morphological items or a task that requires participants to create new words and phrases may offer a promising avenue into furthering our understanding of the processing of morphological structures.

The current work ultimately demonstrated within-level linguistic attachment priming in comprehension, but a few modality-specific research questions remain unanswered due to the difficulty in designing a morphological production task. For instance, there is evidence that distributional patterns in relative clause production are linked to comprehension performance (Gennari & MacDonald, 2009), but it is unclear whether noun phrase production affects subsequent noun phrase production or comprehension.

Aside from innovating ways to measure morphological production, future work may also specify how the structuring mechanisms for different linguistic units differ, and further explore how these non-shared mechanisms are integrated in online production and comprehension. In the interest of keeping the results and the experimental logic comparable to previous work in the generality of structuring mechanisms, the current work chose to use specific test cases of linguistic ambiguity coupled with the structural priming paradigm, which allowed for clear predictions of behaviors. Although ambiguity is ubiquitous in everyday language use, speakers and comprehenders do not always detect it because their prior knowledge and linguistic biases often help them convey or derive the intended meaning effectively. For instance, in the case of relative clause attachment, English speakers tend to have a low attachment preference in production (as shown in our data and previous work, e.g., Van de Cavey & Hartsuiker, 2016), and other semantic cues in discourse may aid in comprehension. Additionally, our pilot data indicated that comprehenders often have strong interpretation preferences for noun phrases that could have more than one plausible meanings due to attachment ambiguity (which led to the small number of morphological target items in our experiments). It is unclear whether there is a clear morphological attachment preference in noun phrases and multimorphemic words in the English language as a whole, and whether that preference aligns with the low attachment preference in sentences. And if so, how do cognitive factors (such as memory and attentional constraints) guide those statistics? In Scheepers and Sturt (2014)'s data, sensibility ratings tended to be higher for right-branching than for left-branching adjective-noun-noun compounds, though it is unclear in what the preference would be in noun phrases like our materials where both HA and LA alternatives are right-branching. Corpus studies with attachment statistics on the morphological level and comparisons with statistics on the sentence level will lend important insights into the universality of cognitive biases in language use.

To conclude, the current study used attachment ambiguity and priming to investigate the sharedness of structuring mechanisms for linguistic units of different grain sizes, namely relative clause sentences and noun phrases. Based on results across three experiments showing that priming was only observed between linguistic units of the same grain size, we concluded that the structuring mechanisms on the sentence level and the morphological level are likely not shared.

References

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology, 18*(3), 355-387.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition, 31*(2), 163-186.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*(1), 1-39.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13-B25.

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an Internet-based study. *Psychonomic Bulletin & Review, 9*(1), 126-131.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language, 54*(4), 610-632.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition, 111*(1), 1-23.

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language, 90*, 14-30.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*(4), 633-651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin, 134*(3), 427.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition, 89*(3), 179-205.

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science, 22*(10), 1319-1326.

Scheepers, C., & Sturt, P. (2014). Bidirectional syntactic priming across cognitive domains: From arithmetic to language and back. *Quarterly Journal of Experimental Psychology, 67*(8), 1643-1654.

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition, 146*, 172-184.

Van Gompel, R. P., & Arai, M. (2018). Structural priming in bilinguals. *Bilingualism: Language and Cognition, 21*(3), 448-455.

Appendix A

Full Lists of Prime and Target Materials in Experiment 1

Table 2.A1

Prime Materials (Presented in the Specified Counterbalanced and Pseudorandomized Order)

Item	HA Prime	LA Prime
1	3 + (6 − 2) / 2 =	3 + 6 − 2 / 2 =
2	10 + (7 − 5) × 3 =	10 + 7 − 5 × 3 =
3	41 − (8 + 3) × 3 =	41 − 8 + 3 × 3 =
4	20 + (32 – 6) / 2 =	20 + 32 – 6 / 2 =
5	56 – (5 + 3) × 4 =	56 – 5 + 3 × 4 =
6	31 + (8 – 5) × 2 =	31 + 8 – 5 × 2 =
7	43 – (27 – 9) / 3 =	43 – 27 – 9 / 3 =
8	19 + (24 – 8) / 4 =	19 + 24 – 8 / 4 =
9	90 – (5 + 15) / 5 =	90 – 5 + 15 / 5 =
10	78 – (9 + 6) × 2 =	78 – 9 + 6 × 2 =
11	45 – (10 + 5) × 3 =	45 – 10 + 5 × 3 =
12	70 – (25 + 5) / 5 =	70 – 25 + 5 / 5 =
13	80 − (9 + 1) × 5 =	80 − 9 + 1 × 5 =
14	67 − (24 − 12) / 3 =	67 − 24 − 12 / 3 =
15	7 + (28 − 4) × 2 =	7 + 28 – 4 × 2 =
16	9 + (20 + 10) / 5 =	9 + 20 + 10 / 5 =
17	15 – (12 – 4) / 2 =	15 – 12 – 4 / 2 =
18	2 + (8 + 4) × 3 =	2 + 8 + 4 × 3 =
19	85 – (14 + 21) / 7 =	85 – 14 + 21 / 7 =
20	10 + (6 + 3) × 2 =	10 + 6 + 3 × 2 =
21	56 + (6 + 6) / 2 =	56 + 6 + 6 / 2 =
22	4 + (22 – 4) / 2 =	4 + 22 – 4 / 2 =
23	98 – (50 – 30) / 10 =	98 – 50 – 30 / 10 =
24	12 + (26 – 1) × 4 =	12 + 26 – 1 × 4 =
25	social psychologist	careless psychologist

Table 2.A1 (Continued)

Prime Materials (Presented in the Specified Counterbalanced and Pseudorandomized Order)

Item	HA Prime	LA Prime
26	political scientist	poor scientist
27	criminal lawyer	stubborn lawyer
28	quantitative analyst	wealthy analyst
29	pure mathematician	talkative mathematician
30	standup comedian	boring comedian
31	classical musician	creative musician
32	primate researcher	diligent researcher
33	marine biologist	adventurous biologist
34	organic chemist	angry chemist
35	electrical engineer	lazy engineer
36	nuclear physicist	meticulous physicist
37	Gabriel cut the tags of the shirt that were (HA) / that was (LA) ____ .	
38	Everyone stared at the mansion of the millionaire that was (HA) / who was (LA) ____ .	
39	The witness recognized the driver of the vehicle who was (HA) / that was (LA) ____ .	
40	The recruiters discussed the performance of the candidates that was (HA) / who were (LA) ____ .	
41	We consulted the accountants of the bank who were (HA) / that was (LA) ____ .	
42	Martin hugged the pets of the school friend that were (HA) / who was (LA) ____ .	
43	The storm destroyed the stairs of the house that were (HA) / that was (LA) ____ .	
44	Hans cleaned the windows of the pharmacy that were (HA) / that was (LA) ____ .	
45	The boy teased the hamsters of the girl that were (HA) / who was (LA) ____ .	
46	Frank thought of the brothers of the friend who were (HA) / who was (LA) ____ .	
47	The voters supported the policies of the politician that were (HA) / who was (LA) ____ .	
48	The firefighters saved the occupants of the penthouse who were (HA) / that was (LA) ____ .	
49	Maria consoled the friends of the roommate who were (HA) / who was (LA) ____ .	
50	Peter heard the birds of the girl that were (HA) / who was (LA) ____ .	
51	The police interrogated the suspect of the crimes who was (HA) / that were (LA) ____ .	
52	Judith prosecuted the owner of the animals who was (HA) / that was (LA) ____ .	

Table 2.A1 (Continued)

Prime Materials (Presented in the Specified Counterbalanced and Pseudorandomized Order)

Item	HA Prime LA Prime
53	Peter used the printers of the department that were (HA) / that was (LA) ____ .
54	Someone shot the cousins of the actress who were (HA) / who was (LA) ____ .
55	Frida complained to the butchers of the supermarket who were (HA) / that was (LA) ____ .
56	The gardener mowed the lawns of the park that were (HA) / that was (LA) ____ .
57	The patient contacted the head physician of the neurologists who was (HA) / who were (LA) ____ .
58	Leo pointed to the drawings of the old man that were (HA) / who was (LA) ____ .
59	The portraitist painted the grandparents of the king who were (HA) / who was (LA) ____ .
60	The officer searched the house of the criminals that was (HA) / who were (LA) ____ .

Table 2.A2

Target Materials

Item	Target Fragment
1	The philanthropist drove the limo of the charities that ____ .
2	They were shocked by the paragraphs of the essay that ____ .
3	The hairdresser helped the stylist of the celebrities who ____ .
4	John met the supervisor of the employees who ____ .
5	The bus driver talked to the leader of the boy scouts who ____ .
6	The writer deleted the lines of the poem that ____ .
7	The hacker attacked the websites of the service provider that ____ .
8	Klara interviewed the mentees of the mentor who ____ .
9	The commission referred to the source of the donations that ____ .
10	The broker communicated with the agent of the buyers who ____ .
11	The thief stole the documents of the organization that ____ .
12	The volunteer bathed the kittens of the cat that ____ .
13	The barista broke the parts of the machine that ____ .
14	Kurt distributed the tickets of the show that ____ .
15	The frost ruined the harvest of the fruit farms that ____ .
16	The tutor advised the students of the lecturer who ____ .
17	The personal trainer adjusted the settings of the treadmill that ____ .
18	The security guard comforted the visitor of the tenants who ____ .
19	The mover called the landlord of the customers who ____ .
20	The homeowner kept the letters of the office that ____ .
21	Ben attacked the boss of the workers who ____ .
22	The expert praised the investor of the young entrepreneurs who ____ .
23	The manager waited for the musicians of the pop star who ____ .
24	The train conductor criticized the kid of the passengers who ____ .
25	The mover insured the furniture of the apartments that ____ .
26	The superintendent checked the earnings of the company that ____ .
27	The florist befriended the servant of the princesses who ____ .
28	We were amused at the articles of the newspaper that ____ .

Table 2.A2 (Continued)

Target Materials

Item	Target Fragment
29	The reader insulted the partners of the editor who ____ .
30	The journalist stalked the dancers of the singer who ____ .
31	The scholar studied the language of the tribes that ____ .
32	The professor taught the daughters of the president who ____ .
33	The astronomer observed the stars of the spiral galaxy that ____ .
34	The farmer fed the calves of the cow that ____ .
35	Donna laughed at the apprentices of the designer who ____ .
36	The historians documented the stories of the city that ____ .
37	The dentist scraped the surface of the teeth that ____ .
38	The restaurant owner fired the helper of the chefs who ____ .
39	The researcher reviewed the debates in the field that ____ .
40	The marketing officer advertised the promotions of the month that ____ .
41	The businessman thanked the newcomer of the workers who ____ .
42	The knight slayed the dragons of the cavern that ____ .
43	The minister saw the bodyguard of the diplomats who ____ .
44	The lifeguard saved the toddler of the parents who ____ .
45	The tour guide mentioned the bells of the church that ____ .
46	The witch cursed the ancestors of the villager who ____ .
47	The pensioner complained about the content of the fliers that ____.
48	The flower girl waved at the relatives of the bride who ____ .
49	The mermaid polished the diamonds of the ring that ____ .
50	Francesca corrected the supporters of the author who ____ .
51	The social worker greeted the nurse of the senior-citizens who ____ .
52	The botanist examined the roses of the garden that ____ .
53	The scientist criticized the method of the studies that ____ .
54	The woman stared at the decorations of the box that ____ .
55	The data scientists analyzed the strategies of the company that ____ .

Table 2.A2 (Continued)

Target Materials

Item	Target Fragment
56	The producer described the plot of the episodes that ____ .
57	The frost destroyed the products of the farm that ____ .
58	The public admired the doctor of the patients who ____ .
59	The protesters disagreed with the general of the soldiers who ____ .
60	The mechanic threatened the driver of the performers who ____ .
61	Nora visited the students of the piano teacher who ____ .
62	A stranger blackmailed the butler of the royals who ____ .
63	The housekeeper replaced the remote of the lights that ____ .
64	The secret service confiscated all files of the organization that ____ .
65	The programmer improved the software of the games that ____ .
66	The economist questioned the report of the businesses that ____ .
67	The reporter spoke to the captain of the players who ____ .
68	The neurologist operated on the mother of the twins who ____ .
69	The chauffeur met the representative of the state guests who ____ .
70	The concierge escorted the negotiator of the union members who ____ .
71	The pilot nodded at the head of the flight attendants who ____ .
72	The porter smiled at the children of the hotel resident who ____ .

Appendix B

Full Lists of Prime and Target Materials in Experiment 2

Table 2.B1

Prime Materials (Presented in the Specified Counterbalanced and Pseudorandomized Order)

Item	Prime Material
1	political scientist (HA)
2	marine biologist (HA)
3	skillful musician (LA)
4	careless psychologist (LA)
5	Frank thought of the brothers of the friend who were (HA) / who was (LA) ____ .
6	The recruiters discussed the performance of the candidates that was (HA) /who were (LA) ____ .
7	Hans cleaned the windows of the pharmacy that were (HA) / that was (LA) ____ .
8	The boy teased the hamsters of the girl that were (HA) / who was (LA) ____ .

Table 2.B2

Target Materials

Item	Target Material
1	baroque painter
2	fine artist
3	revolutionary scholar
4	Latin dancer
5	creative writer
6	canine detective
7	artistic gymnast
8	Russian teacher

Appendix C

Full Lists of Prime and Target Materials in Experiment 3

Table 2.C1

Prime Materials (Presented in the Specified Counterbalanced and Pseudorandomized Order)

Item	Prime Material
1	political scientist (HA)
2	marine biologist (HA)
3	primate researcher (HA)
4	criminal lawyer (HA)
5	skillful musician (LA)
6	careless psychologist (LA)
7	talkative mathematician (LA)
8	angry chemist (LA)
9	Frank thought of the brothers of the friend who were twins (HA) / who was sick (LA).
10	The recruiters discussed the performance of the candidates that was surprising (HA) /who were international students (LA).
11	Hans cleaned the windows of the pharmacy that were full of fingerprints (HA) / that was broken into (LA).
12	The boy teased the hamsters of the girl that were running around (HA) / who was an animal advocate (LA).
13	The farmer fed the calves of the cow that were crying (HA) / that was big and strong (LA).
14	The witness recognized the driver of the vehicle who had a suspended license (HA) / that was stolen (LA).
15	Everyone stared at the mansion of the millionaire that was renovated recently (HA) /who was generous (LA).
16	The priest spoke to the leader of the scouts who was dishonest (HA) / who were going camping (LA).

Table 2.C2

Target Materials

Item	Target Material
1	baroque painter
2	fine artist
3	revolutionary scholar
4	Latin dancer
5	creative writer
6	canine detective
7	artistic gymnast
8	Russian teacher
9	John met the friend of the teacher who was in Germany.
10	Someone shot the servant of the actress who was on the balcony.
11	Andrew was speaking with the niece of the cleaner who was in Brazil.
12	The journalist interviewed the daughter of the colonel who had the accident.
13	The police arrested the sister of the porter who was in Melilla.
14	The boys poked fun at the son of the painter who was in the park.
15	My mother argued with the maid of the duchess who left the house.
16	Amilia exchanges letters with the cousin of the singer who was in the church.

Acknowledgements

Chapter 2, in full, is a reprint of the material currently being prepared for submission for publication. Lau, Sin Hang; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3

Lexical tone is different and special: Evidence from a speeded repeated production task

Sin Hang Lau, Chuchu Li, Victor S. Ferreira

Author Note

Sin Hang Lau, Department of Psychology, University of California San Diego; Chuchu Li, Department of Psychiatry, University of California San Diego; Victor S. Ferreira, Department of Psychology, University of California San Diego

Correspondence concerning this article should be addressed to Sin Hang Lau, Department of Psychology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: aubreylau@ucsd.edu

Abstract

A distinctive feature of tonal languages such as Mandarin Chinese is that the same consonant-vowel (CV) sequence is a different word depending on the *tone* it is spoken with. However, there is currently no strong evidence on how tone is represented in speech planning (i.e., attached to the vowel or CV as a whole, or represented independently), as well as whether tone is involved in early or late encoding in phonological encoding. We explored these questions with a word sequence repetition task and measured how quickly speakers produced sequences of CV-tone syllables. Across two experiments, we obtained unexpected results showing that the number of unique CV or tone units in the sequence did not predict speech rate, nor did their repetition pattern. Instead, speech rate was robustly faster when each CV needed to be produced with only one tone (i.e., about equal speech rate for ba2 di1 da1 bi2 and ba1 ba1 ba1 ba1), compared to when a particular CV needed to be produced with more than one tone (i.e., slower speech rate for ba1 ba2 ba1 ba2). We suggest that Mandarin speakers represent CVs as syllable "chunks," integrating each with a structural frame involving tone, so that producing the same chunk with more than one tone in a sequence is difficult. Our results support models of phonological encoding that select syllables as proximate units, represent tone separately but not independently from segments, and include a late integration stage of syllable and tone.

Keywords: tone, speech planning, Mandarin production, phonological encoding

Lexical tone is different and special:

Evidence from a speeded repeated production task

In tonal languages, producing the same sequence of segments with different pitch variations (i.e., tones) leads to different meanings. For example, in Mandarin, *ma* could mean “mother”, “hemp”, “horse”, or “to scold” depending on which of the four tones it is produced with. Despite the prevalence of tone in African (Odden, 1995) and East Asian languages (Yip, 1995), its representation in phonological encoding and the cognitive mechanisms involved in integrating tone with other phonological elements is not well-established. Many language production models are based on (atonal) Indo-European languages (e.g., Levelt, 1989; Roelofs, 1997) which leaves open the question of how tone can be incorporated into these models. Not only is it necessary to account for the additional feature of tone and its integration with other elements in production models, recent cross-linguistic comparisons have also revealed that tonal and non-tonal languages may also differ in syllable representation (see O’Seaghdha et al., 2010 for cross-linguistic comparisons), calling into question the language-generality of Indo-European focused models. The current work aims to validate previous proposals about the syllable structure of Mandarin and further investigate the role of tone in phonological encoding. Here we first review evidence showing the cross-linguistic differences in status of the syllable in speech planning, then introduce two classes of theories regarding tone representation.

Even though words in both tonal and atonal languages can generally be broken down into syllables and segments, the roles of different units in phonological encoding may differ. In particular, the implicit priming paradigm (Meyer, 1990, 1991) has been widely used to study the phonological units that are involved in speech planning, by observing the conditions that lead to form preparation benefits in various languages. The paradigm requires participants to study pairs of prompt-target words in blocks, with the target words in a block being either homogeneous or heterogeneous with respect to the proposed speech planning unit. Borrowing examples from the English experiment in O'Seaghdha et al. (2010), when the onset segment is the proposed speech planning unit, the target words in the homogeneous block would all share the same onset (*sting-**b**ee; port-**b**ay; ghost-**b**oo; leave-**b**ye*), whereas the target words in the heterogeneous block would have different onsets (*mud-**g**oo; scale-**w**eigh; lock-**k**ey; clothes-**t**ie*). Participants are typically asked to memorize prompt-target pairs in study blocks. At test, participants are presented with each prompt word and asked to retrieve and produce the corresponding target.

In the implicit priming paradigm, if participants' production times for the target words in the homogeneous condition are faster than in the heterogeneous condition, it is taken as evidence of speakers engaging in form preparation using the proposed planning unit. In turn, such a form preparation benefit implies that the homogenous unit can be selected in advance in production and buffered. Cross-linguistic comparisons have revealed that onset segments lead to form preparation in Indo-European languages such as English and Dutch (Meyer, 1990, 1991; Roelofs & Meyer, 1998), but not in Mandarin (Chen et al., 2002; O'Seaghdha et al., 2010). Rather, in Mandarin, whole syllables lead to form preparation benefits. Importantly, the benefits were observed even when the syllables are not homogenous in tones.

To unify models of phonological encoding and accommodate cross-linguistic differences in planning units, O'Seaghdha et al. (2010) refer to the first selectable phonological units below the word as *proximate units*. To summarize the implicit priming results under this framework, the proximate units appear to be segments in Indo-European languages but atonal syllables in Mandarin. They proposed a model of Mandarin phonological encoding that first selects atonal syllable chunks as a proximate step following the word level, then retrieves segments to sequence into a syllable frame, which is in turn assigned a tone value as a penultimate pre-articulatory step to phonetic encoding. Note, however, that the role of tone was only proposed based on the presence of syllable form preparation in the absence of tone overlap, and tone was not an active manipulation or a proposed planning unit in the experiments. Hence, further testing is required to validate the idea of tone assignment to syllable frame.

While there is converging evidence for syllables as proximate units in Mandarin phonological encoding, the investigation of tone representation is still on-going and contentious. A sizable portion of the existing literature on tone representation has relied on analyzing speech errors, which developed into two main classes of theories: On the one hand, early encoding accounts propose that tone functions similarly as segments and participates in early selection processes, making it as susceptible to selection errors as consonants and vowels (e.g., Alderete et al., 2019; Wan & Jaeger 1998; Moser, 1991; Shen, 1993). On the other hand, late encoding accounts posit that tone is only involved in the pre-articulatory step of integrating segments with the metrical frame and thus is relatively immune to selection errors (e.g., Chen, 1999; Kember et al., 2015; Roelofs, 2015).

The two classes of theories make different specific predictions in regards to the types and relative frequencies of speech errors one should observe in tonal languages. The early encoding accounts predict that tone errors should be relatively common and qualitatively similar to segmental errors. In particular, errors are expected to be contextual, meaning that it is more likely than chance that there is some segmental or tonal overlap between the intended and error words (see Wan & Jaeger, 1998 for Mandarin, and Garrett, 1984 for English). In contrast, the late encoding accounts predict that tone errors should be rare (relative to segmental errors), because tone is attached to the abstract structural syllable frame and thus not subjected to active selection competition that its contents (i.e., segments) engage in. Overall, there is not much consensus on whether early or late encoding of tone is more likely.

Earlier linguistic theories diverged on whether tone has an independent phonological representation separate from segments. For example, Halle and Steven (1971) posited that tones are encoded as phonological features of vowels, whereas Leben (1978) argued that tones are representationally independent of vowels but realized as phonetic features of vowels (i.e., suprasegmental features). Various speech error studies in Mandarin (Jaeger & Wan, 1998; Moser, 1991; Shen, 1993) and Thai (Gandour, 1977) have observed contextual misorderings of tone (i.e., perseveration, anticipation, and exchange), which are also commonly found with consonants and vowels in English (Fromkin, 1971). Importantly, the error patterns suggest that tone not only has its own phonological representation that is similar to but independent of the representation of segment, but that it is also actively involved in selection as early as segments are.

For example, Wan and Jaeger (1998) analyzed 788 slips of the tongue from Mandarin naturalistic conversations and found that lexical substitutions and blends have the same tones twice as often as expected by chance, illustrating that tone influences early lexical selection. In tone errors, such as the perseveration error *tui1 jian4* ***han4*** (ungrammatical) "letter of recommendation" (intended as *tui1 jian4* ***han2***), a tone slipped while the segmental content remained intact (i.e., Tone 4 from "jian" perseverated and substituted Tone 2 while "han" remained as intended), which suggests that tone has its own phonological representation that is separate from segments but as sensitive to context as segments. However, the small number of errors and the recording procedure (tape-recorded or handwritten by the authors) in Wan and Jaeger (1998) limited the scope of the results. Nevertheless, Alderete et al.'s (2019) study analyzing 2462 Cantonese spontaneous speech errors from audio recordings of natural conversations corroborated the previous findings. With a larger and more inclusive sample, this study was able to shed light on the relative frequencies of different types of errors, with a particular interest on how often contextual tone errors occur. In addition to replicating that tones can be mis-selected while segments are not, Alderete et al. reported that the rate of tone slips that were contextual (76.37%) was comparable to the proportion of contextual errors out of all observed speech errors (62.13%), providing support to the early encoding account.

In contrast, based on 987 Mandarin speech errors from tape-recorded audio call-in programs, Chen (1999) reported that tone errors were rare, compared to segment errors. Specifically, 136 segment errors were observed, but only 24 suspected tone errors were recorded, of which 19 could have been attributed to an alternative error process. The majority of the five remaining tone errors were perseverations, whereas segmental errors were primarily anticipations. That is, tones were found to be both quantitatively and qualitatively different from segments, which is incompatible with the idea of tone engaging in segment-like early encoding. Chen proposed a production model in which tone is initially represented as part of the phonological frame similar to lexical stress in English (i.e., suprasegmental) and later translated into the phonetic configuration of the vowel in phonetic encoding, which is referred to as the late encoding account.

Traditional methods of collecting and analyzing spontaneous speech errors are unfortunately susceptible to perceptual biases (Chen, 1999; Alderete & Davies, 2018). Further investigation using experimental methods is required to elucidate insights made from measures that are especially vulnerable to collection methods, such as relative frequencies of types of errors. One such method is a tongue twister task that carefully manipulates the patterns of segments and tones. Kember et al. (2015) designed 120 four-character Mandarin tongue twisters that rotated pairs of initial segments or tones, or both, across ABAB or ABBA format, across positions. For example, "突哭突哭" (*tu1 ku1 tu1 ku1*) comprises ABAB initial segments and constant tone; "突土土突" (*tu1 tu3 tu3 tu1*) constant initial segment and ABBA tones; and "突苦土哭" (*tu1 ku3 tu3 ku1*) ABAB initial segments and ABBA tones. Participants produced the tongue twisters six times in a row, and speech errors were recorded. The experiment yielded 3503 segment errors and 1372 tone errors, which validates Chen's (1999) claim that segments are more prone to error than tones, even with a task that supposedly affords both elements the same opportunities for error. Additional detailed analyses on the error positions are included in Kember et al. (2015), but the findings are more relevant to articulatory planning rather than our focus of phonological encoding (we thus do not review these findings in detail). The brief summary is that the distribution of tone and segment errors across the positions of the tongue twister differed. For example, in tone-alternating conditions, segment errors were more likely at positions 2 and 4, whereas tone errors were more likely at positions 3 and 4. Altogether, the evidence from the tongue twister study supports the late encoding account, suggesting that tone is part of a metrical frame in phonological encoding.

Though speech error studies have lent important insights to the topic of tone representation, they inevitably suffer from lack of power because the vast majority of speech is error-free regardless of whether it is collected naturally or in the laboratory. Additionally, even though the materials Kember et al. (2015) were carefully designed, the repetition patterns were not fully crossed in the design (e.g., no condition with both initial segment and tone held constant, nor both initial segment and tone alternating in the same pattern).

In light of these limitations, the current work employed a paradigm from Sevald and Dell (1994) that more directly measures speech planning and retains more data. The original study examined how English speakers plan and produce monosyllabic CVC words by asking participants to repeat four word sequences as many times as possible in eight seconds. It was similar to Kember et al. (2015), except speech rate was measured, rather than errors. Importantly, Sevald and Dell (1994) independently varied the repetition pattern for the onset C, V, and final C, such that they either followed an AAAA, ABBA, or ABAB pattern (i.e., three different levels of repetition frequency, in descending order), constituting a fully crossed design with 27 possible conditions. In our adapted version, we substituted the final C position with tone (the materials are described below in Experiment 1 method). The general logic of the paradigm is that repetition benefits should be observed in conditions where a planning unit is repeated more frequently, compared to less. That is, reusing a plan should be easier than alternating between different plans, leading to faster speech rate. By observing which repetitions lead to repetition benefits, we can then infer which candidate units indeed function as speech planning units. Since Mandarin proximate units and tone representation have not been studied using speech rate, we felt that as a first pass it was important to use a design that entertained all possibilities. Experiment 1 was thus exploratory in nature, exploring possibilities that included but were not limited to C, V, and tone functioning as independent planning units, CV functioning together as one proximate unit, or Tone functioning as a phonological feature of V. Experiment 2 is a confirmatory test that further examines how the planning units are integrated based on the results of Experiment 1.

Experiment 1

Method

Participants

Fifty-eight undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Fourteen participants were excluded from our analyses because they either did not produce more than 50% usable data (exclusion criteria explained below in Coding and Data Analysis), or had extensive prior knowledge of tonal languages other than Mandarin, including Cantonese. All 44 remaining participants indicated that they were native Mandarin Chinese speakers with little to no exposure to other tonal languages, and that they moved to the US after the age of 15.

Materials and Design

We used a word sequence repetition task (adapted from Sevald & Dell, 1994) to assess the effects of the repetition frequency of each position (Consonant, Vowel, or Tone) and their interaction on speech rate. Materials in this task were four-syllable sequences. All four-syllable sequences comprised semantically-unrelated monosyllabic words (i.e., all sequences were nonwords) with no alternate pronunciation. Additionally, we avoided Tone 3 (due to tone sandhi) and polyphones. These were controlled with a within-subjects, 3 (C Pattern) x 3 (V Pattern) x 3 (Tone Pattern) factorial design, such that each of the three Positions of each monosyllabic word (C, V, and Tone; generated from three unique phoneme sets) repeated independently in one of three repetition patterns (AAAA, ABBA, or ABAB), creating 27 possible combinations for each phoneme set. That is, there were 81 trials in the experiment. Each phoneme set consisted of two sounds in each position, and three counterbalancing lists were constructed to randomize which sound in each position was arbitrarily assigned as Sound A or Sound B in each repetition pattern across participants. Table 3.1 shows the pinyin of the sounds in the three phoneme sets. Taking Phoneme Set 1 as an example, if the C repeated in a AAAA pattern (*b*__ *b*__ *b*__ *b*__), V in a

ABBA pattern (_*a*_ _*i*_ _*i*_ _*a*_), and Tone in a ABAB pattern (__*2* __*1* __*2* __*1*), the resulting four word sequence would be "拔逼鼻巴" (*ba2 bi1 bi2 ba1*). Table 3.2 includes 27 example trials from one Phoneme Set.

Table 3.1

Sounds in Each Phoneme Set

Phoneme Set 1			Phoneme Set 2			Phoneme Set 3		
C	V	Tone	C	V	Tone	C	V	Tone
b	a	2	n	i	2	l	a	1
d	i	1	t	u	4	p	u	4

Table 3.2

Example trials of the 27 Conditions from Phoneme Set 1

Pinyin	Characters	C Pattern	V Pattern	Tone Pattern
ba2 ba2 ba2 ba2	拔拔拔拔	AAAA	AAAA	AAAA
ba2 ba1 ba1 ba2	拔巴巴拔	AAAA	AAAA	ABBA
ba2 ba1 ba2 ba1	拔巴拔巴	AAAA	AAAA	ABAB
ba2 bi2 bi2 ba2	拔鼻鼻拔	AAAA	ABBA	AAAA
ba2 bi1 bi1 ba2	拔逼逼拔	AAAA	ABBA	ABBA
ba2 bi1 bi2 ba1	拔逼鼻巴	AAAA	ABBA	ABAB
ba2 bi2 ba2 bi2	拔鼻拔鼻	AAAA	ABAB	AAAA
ba2 bi1 ba1 bi2	拔逼巴鼻	AAAA	ABAB	ABBA
ba2 bi1 ba2 bi1	拔逼拔逼	AAAA	ABAB	ABAB
ba2 da2 da2 ba2	拔达达拔	ABBA	AAAA	AAAA
ba2 da1 da1 ba2	拔搭搭拔	ABBA	AAAA	ABBA
ba2 da1 da2 ba1	拔搭达巴	ABBA	AAAA	ABAB
ba2 di2 di2 ba2	拔笛笛拔	ABBA	ABBA	AAAA
ba2 di1 di1 ba2	拔低低拔	ABBA	ABBA	ABBA
ba2 di1 di2 ba1	拔低笛巴	ABBA	ABBA	ABAB
ba2 di2 da2 bi2	拔笛达鼻	ABBA	ABAB	AAAA
ba2 di1 da1 bi2	拔低搭鼻	ABBA	ABAB	ABBA
ba2 di1 da2 bi1	拔低达逼	ABBA	ABAB	ABAB
ba2 da2 ba2 da2	拔达拔达	ABAB	AAAA	AAAA
ba2 da1 ba1 da2	拔搭巴达	ABAB	AAAA	ABBA
ba2 da1 ba2 da1	拔搭拔搭	ABAB	AAAA	ABAB
ba2 di2 bi2 da2	拔笛鼻达	ABAB	ABBA	AAAA
ba2 di1 bi1 da2	拔低逼达	ABAB	ABBA	ABBA
ba2 di1 bi2 da1	拔低鼻搭	ABAB	ABBA	ABAB
ba2 di2 ba2 di2	拔笛拔笛	ABAB	ABAB	AAAA
ba2 di1 ba1 di2	拔低巴笛	ABAB	ABAB	ABBA
ba2 di1 ba2 di1	拔低拔低	ABAB	ABAB	ABAB

Procedure

Participants first completed a language history questionnaire, in which they indicated whether they identified as native speakers of Mandarin Chinese, whether they spoke any other tonal languages, and the age at which they moved to the US.

Then, an experimenter explained the instructions of the experimental task (modeled closely after Sevld & Dell, 1994) and went through five practice trials with each participant. The beginning of each trial was signaled by a fixation cross (presented for 200 ms) and a high-pitched tone. Then, participants saw a four-syllable sequence written in Simplified Chinese characters centered on the screen, with the word "准备" (prepare) printed underneath for eight seconds. They were instructed to silently rehearse the sequence during that period in preparation for the upcoming production phase. At the end of the eight seconds, participants heard three low-pitched tones and one high-pitched tone as a countdown signal for the production phase. The high-pitched tone and the disappearance of the word "准备" (prepare) signaled the beginning of the eight second long production phase, during which participants were instructed to produce the four-word sequence aloud repeatedly as quickly and as accurately as possible. At the end of the production phase, a high-pitched tone signaled the end of the trial. Participants pressed the spacebar to advance to the next trial whenever they were ready. The experimenter provided feedback to the participants for five practice trials before the participants completed the 81 experimental trials independently.

Coding and Data Analysis

The spoken response for each trial was audio-recorded and later coded into the dependent variable (speech rate) with the aid of Audacity, an audio processing program. Specifically, we manually counted the number of syllables produced within the eight second production period, excluding any production that overlapped with the beginning or terminal beep signal of each trial. Based on our goal to indirectly infer the planning units of speech using speech rate when a sequence was correctly produced, we excluded 523 (14.67%) trials which included speech errors, significant disfluencies (other than breathing, such as the inclusion of fillers or any other non-speech sounds or restarting a sequence), abnormal reaction times (below 150 ms or above 800 ms in response to the beginning signal) or other technical errors, resulting in 3,041 (85.33%) analyzable trials. Two counters (native Mandarin speakers) were each responsible for counting half of the trials and cross-checking the other half that the other counter initially counted. Of the analyzable trials, only 29 trials led to inter-counter disagreements, which were all resolved with a third round of counting. The syllable counts were then transformed into average production times per syllable by dividing 8,000 ms by the syllable count.

The statistical analyses were conducted using R (R Core Team, 2014) and the lme4 package (Bates et al., 2014). In our main analysis, we used a linear mixed-effects model to analyze the effects of C Pattern, V Pattern, Tone Pattern, and their interactions on average production times. We attempted to use the maximal random effects structure for both models. The model that converged included random intercepts for participants and items.

Results

Our planned analysis revealed significant main effects of repetition pattern for all three positions: C, V, and Tone. However, unlike English, more frequent repetition did not necessarily lead to faster speech rate. Table 3.3 shows the average production times for all 27 conditions, as well as the marginal means for C and V repetition patterns. As can be seen, speech rate was the fastest when the C pattern was ABBA (280 ms), followed by ABAB (285 ms). Surprisingly, AAAA was the slowest (290 ms). The main effect of the C repetition pattern was statistically significant, $\chi^2(2) = 12.08$, $p = .002$. Similarly for the V pattern, ABBA was the fastest (279 ms), followed by ABAB (283 ms), and finally AAAA (292 ms); the main effect of V repetition pattern was statistically significant, $\chi^2(2) = 21.07$, $p < .001$. Meanwhile, the Tone repetition pattern followed a different trend, with AAAA being the fastest (280 ms), then ABAB (286 ms), and ABBA (290 ms); the main effect of Tone repetition pattern was statistically significant, $\chi^2(2) = 12.57$, $p = .002$.

Our results thus far have shown that speech rates in Mandarin production cannot be reliably predicted by repetition of each of the phonological elements alone. In addition, we obtained significant interactions between C and Tone, $\chi^2(4) = 23.64$, $p < .001$; V and Tone, $\chi^2(4) = 27.12$, $p < .001$; as well as a three-way interaction between C, V, and Tone, $\chi^2(8) = 23.04$, $p = .003$. As described above, the repetition patterns in these three positions trended in different directions, rendering these interactions difficult to interpret in terms of repetition frequency alone. Moreover, these interactions suggest that perhaps the three phonological elements should not be treated as completely independent of each other. Thus, we conducted additional post-hoc analyses to explore whether the current data set could help to inform what the planning units are in Mandarin production, by analyzing what groupings of C, V, and Tone could better predict speech rate.

Table 3.3

Average Production Times of the 27 Conditions and Marginal Means for Each C, V, and Tone Repetition Patterns

V Repetition Pattern	Tone Repetition Pattern	C Repetition Pattern						*M*
		AAAA	*M*	ABBA	*M*	ABAB	*M*	
	AAAA	278		282		279		
AAAA	ABBA	315	299	283	292	315	290	292
	ABAB	304		311		277		
	AAAA	283		270		285		
ABBA	ABBA	269	284	275	274	286	283	279
	ABAB	299		276		279		
	AAAA	281		284		281		
ABAB	ABBA	297	288	281	279	297	284	283
	ABAB	287		273		273		
M		290		280		285		

Post-hoc Analyses

In light of the surprising findings from our planned analysis, particularly that the repetition frequency of the phonological elements (when they were treated as independent units) did not reliably predict speech rate, we explored the potential dependence between C, V, and Tone in three post-hoc analyses.

First, we explored the possibility of a speech planning advantage when there are covarying repetition patterns. That is, perhaps Mandarin speech planning is more affected by whether the planning units share the same repetition pattern or not, more so than what the exact pattern is. At this point, we have not yet established what the planning units in Mandarin speech are, other than suggesting that (given the results of the planned analyses) it is highly unlikely that C, V, and Tone function as independent units. Thus, in order to explore if and how Mandarin phonological elements may be dependent on each other, we recoded our data by analyzing two of the elements at a time (i.e., C and V, V and Tone, and C and Tone) and comparing production times in trials where those two elements either shared the same repetition pattern or not. The results showed that whether C and V shared the same pattern made no difference to speech rate (285.51 ms when C and V shared the same pattern and 285.17 ms they did not; 0.35 ms difference, z = -0.09, p = .93), whereas significantly faster speech rates were observed when C and Tone shared the same pattern (277.73 ms for same and 289.06 for different; 11.33 ms difference, z = 3.05, p = .002), and when V and Tone shared the same pattern (276.93 ms for same and 289.46 ms for different; 12.54 ms difference, z = 3.25, p = .001). Because C and V seemed to behave similarly in relation to Tone, we additionally explored the notion that CV may together function as a syllabic unit. We once again recoded the data and compared trials in which Tone shared the same repetition pattern with the syllables (i.e., CV treated as one combined unit) with other trials. We found that speech rates were also faster when CV shared the same pattern with Tone (274.30 ms for same and 286.66 ms for different; 12.36 ms difference, z = 2.15, p = .03).

The set of comparisons above point towards two points: One, Tone seemed to have a privileged status in speech planning, such that speech production was easier when Tone was "attached" to another element (i.e., shared the same repetition pattern with another element). Two, given that the speech advantages we observed above were around the same magnitude regardless of whether Tone was attached to C or V alone or CV as a unit, we suspected that C and V may not be represented independently in Mandarin, unlike in English. However, these results remained preliminary, as they were inferences that were not obtained from direct comparisons. Our second post-hoc analysis attempted to address this limitation by using a single model. Specifically, we re-classified the trials into five categories: trials in which only one tone was involved (Tone pattern was AAAA), trials in which there were two tones (i.e., Tone pattern was either ABBA or ABAB) that covaried with (i.e., shared the same repetition pattern with) C alone, V alone, both C and V, and neither C nor V.

We compared the production times across the five covariation categories, in order to assess whether trials involving two tones were slower than those involving only one, and whether covariation between Tone and other elements produced any speech rate advantage. The results are shown in Figure 3.1. Our model revealed that there was a statistically significant difference in speech rate among these five categories ($\chi^2(4) = 41.80$, $p < .001$). Specifically, the effect was solely driven by the slow production times in the “neither” condition. Pairwise comparisons showed that there was no difference between trials that only contained one tone (279 ms) and trials that contained two tones sharing the same repetition pattern as C alone (279 ms), V alone (279 ms), or both C and V (272 ms); pairwise comparisons *ps* > .05. Importantly, the trials involving two tones that did not share repetition pattern with any other element (302 ms) were significantly slower than all other conditions: 22.39 ms slower than one tone ($z = -5.22$, $p < .001$), 22.24 ms slower than sharing with C alone ($z = -4.98$, $p < .001$), 23.06 ms slower than sharing with V alone ($z = -4.95$, $p < .001$), and 29.7 ms slower than sharing with both C and V ($z = -4.89$, $p < .001$).

These comparisons confirmed that speech production was significantly more difficult when Tone did not covary with other elements. Additionally, based on the lack of additive speech rate advantage when Tone shared the repetition with both C and V, compared to with C or V alone, C and V likely form a single planning unit together in Mandarin. Notably, trials involving one tone were not significantly different from trials involving two tones, with the exception of the “neither” condition, suggesting that it may not be the number of unique phonological units but rather the CV-Tone pairing that affects Mandarin production. In particular, we speculated that perhaps trials in the "neither" condition (e.g., *ba2 bi1 ba1 bi2*: AAAA for C, ABAB for V, and ABBA for Tone) were particularly difficult because each CV was paired with two different tones in the sequence, such that speakers had to repeatedly "detach" a CV from one tone and reattach it to another tone in preparation for the next time it appears in the sequence, slowing production. We refer to this speculation as the *reattachment hypothesis* and explored the validity of it in the third post-hoc analysis.

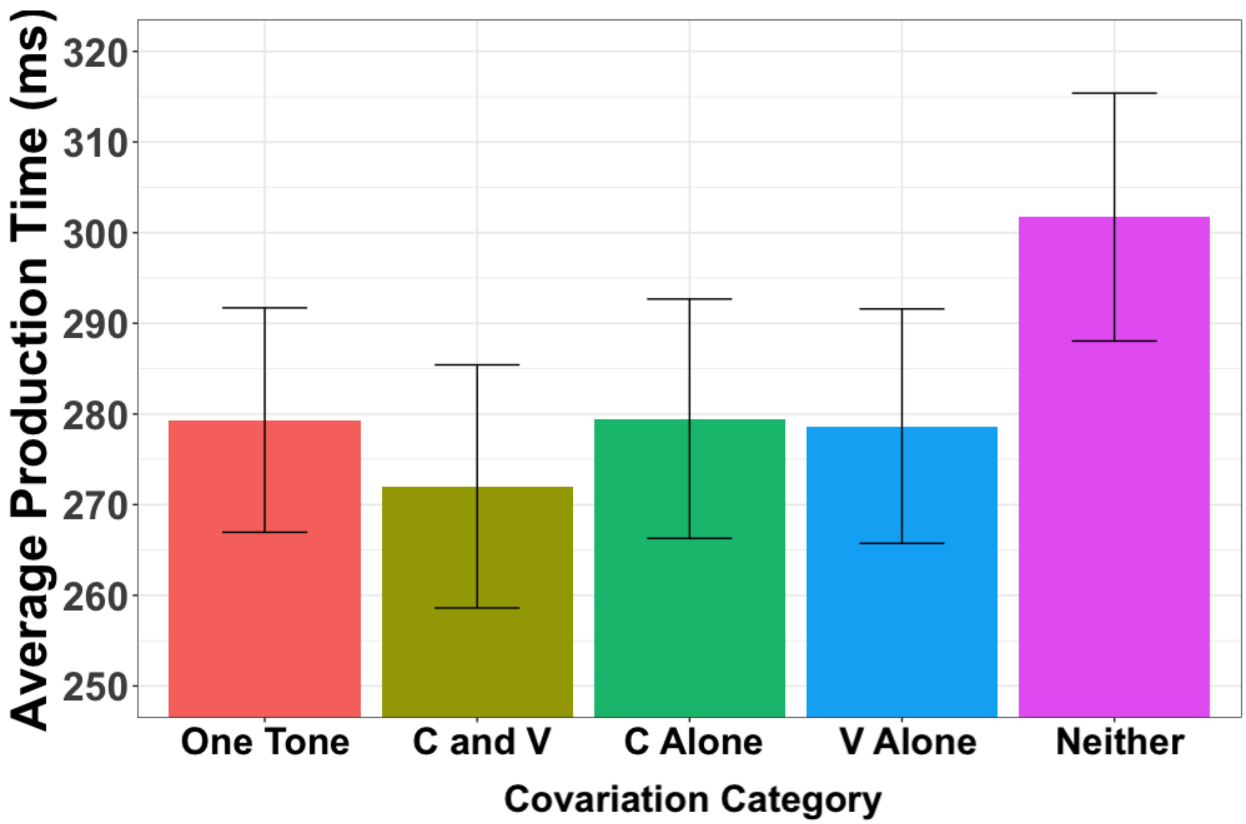

Figure 3.1. Average production times for trials with one tone only (e.g., ba2 ba2 ba2 ba2), two tones that covaried with both C and V (e.g., ba2 di1 di1 ba2), C alone (e.g., ba2 da1 da1 ba2), V alone (e.g., ba2 bi1 bi1 ba2), and neither (e.g., ba1 bi2 bi1 ba2). Error bars represent standard errors.

In the third and final post-hoc analysis, we redefined repetition using CV as a unit. Notably, in addition to the previous patterns (AAAA, ABBA, and ABAB), this way of coding now presents a new possibility of the ABCD pattern (e.g., *ba_ di_ da_ bi_*; previously coded as

ABBA for C and ABAB for V). More importantly, we also classified trials into conditions where "reattaching" is required versus not. Reattaching is defined by whether each unique CV is only paired with one (i.e., no reattaching) versus two tones in the sequence (and thus requiring detaching from one tone and reattaching to another repeatedly).

We assessed the effects of CV repetition pattern, reattaching, and their interaction on production times. Figure 3.2 shows the results. Consistent with what we have reported thus far, CV repetition pattern did not affect speech rate ($\chi^2(3) = 6.15$, p = .10) and there was thus no interaction with reattaching ($\chi^2(2) = 1.83$, $p = .40$). Critically, we observed a main effect of reattaching ($\chi^2(1) = 16.37$, $p < .001$), such that speech rate was on average 23 ms slower when each CV was required to reattach to a different tone within the sequence (301 ms), compared to when no reattaching was required (278 ms).

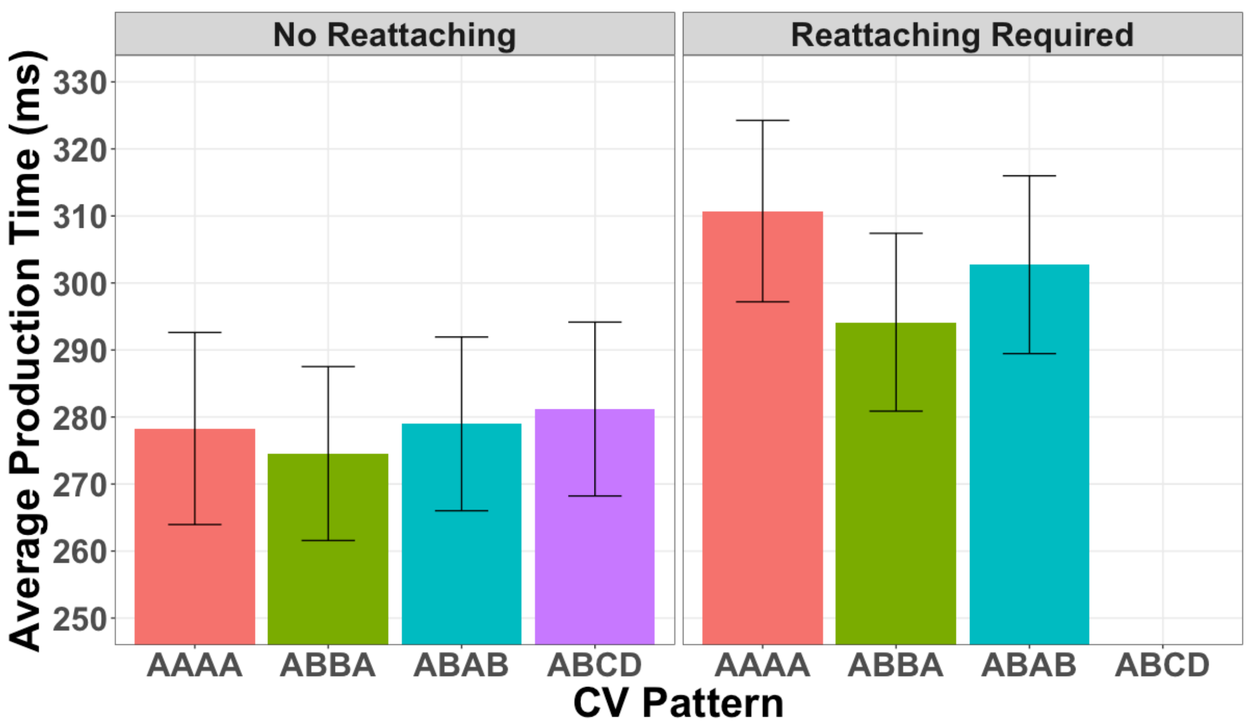

Figure 3.2. Average production times for trials for various CV repetition patterns that required reattaching versus not. Error bars represent standard errors.

To conclude, the planned and post-hoc analyses in Experiment 1 suggested that C, V, and Tone are not independent planning units in Mandarin. Instead, C and V likely function together as a single planning unit, whose production difficulty relies on the pairing with Tone. Mechanistically, we speculate that Mandarin speakers represent CVs as syllable "chunks," programming each with tone upon phonetic encoding, so that producing the same chunk with more than one tone in a sequence is difficult. We simplified the design of Experiment 1 to directly test the reattachment hypothesis in Experiment 2.

Experiment 2

Method

Participants

Our pre-registered target sample size (based on a power analysis conducted with data from Experiment 1) was 26 participants. Forty-four undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Eighteen were excluded from our analyses because they either did not produce more than 50% usable data (based on the same exclusion criteria in Experiment 1, noted above in Coding and Data Analysis), had extensive prior knowledge of tonal languages other than Mandarin, including Cantonese, or had unstable internet connections. All 26 remaining participants indicated that they were native Mandarin Chinese speakers with little to no exposure to other tonal languages, and that they moved to the US after the age of 15.

Materials and Design

The primary goal of Experiment 2 was to test the reattachment hypothesis with a simpler design. That is, we investigated whether speech rate was indeed slower when each CV was paired with more than one Tone in a sequence (i.e., requiring reattaching). We used the same word sequence repetition task as Experiment 1, with the logic that slower speech rate indicates higher speech planning and articulatory difficulty. We adopted a within-subjects, 2 (CV Number) x 2 (Tone Number) x 2 (Repetition Pattern) factorial design, such that the four-syllable sequences comprised either one or two CV units; either one or two tones; and when two CV units or two tones were involved, they either followed an ABBA or ABAB pattern. Out of the eight conditions, the two conditions that paired one CV with two tones were further recoded as "requiring reattaching", while the other six were recoded as "no reattaching required". The materials were constructed using the nine phoneme sets in Table 3.4, counterbalancing across participants which sound was assigned as Sound A or Sound B in each repetition pattern. In sum, each participant contributed to 72 trials (i.e., 8 conditions x 9 Phoneme Sets). Table 3.5 shows the eight experimental conditions from Phoneme Set 1. Taking this set as an example, "拔巴巴拔" (*ba2 ba1 ba1 ba2*) and "笛低笛低" (*di2 di1 di2 di1*) were the two conditions that required each CV to detach from a tone and reattach to another within the sequence, with the former reattaching less frequently than the latter (i.e., reattaching every other word versus every word).

Table 3.4

Sounds in Each Phoneme Set

Phoneme Set	CV	Tone
1	ba	2
	di	1
2	ni	2
	tu	4
3	la	1
	pu	4
4	ti	2
	qü	4
5	ma	2
	fu	4

Phoneme Set	CV	Tone
6	du	4
	mi	2
7	ke	2
	jü	4
8	pa	2
	zu	1
9	ji	2
	hu	4

Table 3.5

Example trials of the 8 Conditions from Phoneme Set 1

Pinyin	Characters	No. of CV	No. of Tone	Repetition Pattern	Reattaching
ba2 ba2 ba2 ba2	拔拔拔拔	One	One	N/A	No
ba2 di2 di2 ba2	拔笛笛拔	Two	One	ABBA	No
ba2 ba1 ba1 ba2	拔巴巴拔	One	Two	ABBA	Yes
ba2 di1 di1 ba2	拔低低拔	Two	Two	ABBA	No
di2 di2 di2 di2	笛笛笛笛	One	One	N/A	No
di2 ba2 di2 ba2	笛拔笛拔	Two	One	ABAB	No
di2 di1 di2 di1	笛低笛低	One	Two	ABAB	Yes
di2 ba1 di2 ba1	笛巴笛巴	Two	Two	ABAB	No

Procedure

The procedure was exactly the same as Experiment 1, except that the experiment was administered remotely over the internet, via Zoom. Instead of having the participants control the pace of the experiment in between trials, we constructed a video such that each trial was presented one after another without breaks. An experimenter showed the video to each participant using the screen share function, and the participant's spoken responses were recorded. The preparation and the production durations remained unchanged from Experiment 1 (i.e., eight seconds each).

Coding and Data Analysis

Similar to Experiment 1, we manually counted the number of syllables produced within the eight second production period, excluding any production that overlapped with the beginning or terminal beep signal of each trial. Based on our goal to indirectly infer the planning units of speech using speech rate when a sequence was correctly produced, we excluded 331 (17.68%) trials using the same exclusion criteria as Experiment 1, resulting in 1541 (82.32%) analyzable trials. Two counters (native Mandarin speakers) were each responsible for counting half of the trials and cross-checking the other half that the other counter initially counted. Of the analyzable trials, only 25 trials led to inter-counter disagreements, which were all resolved with a third round of counting. The syllable counts were then transformed into average production times per syllable by dividing 8,000 ms by the syllable count.

There were two major analyses in this experiment. First, we used a linear mixed-effects model to analyze the effects of the number of CV, Tone, Repetition Pattern, and their interactions on average production times. We used the maximal random effects structure, such that the resulting model included all the fixed effects as by-subject random slopes and the counterbalancing lists as the by-item random slope. This analysis allowed us to (as in Experiment 1) investigate whether the number of unique planning units involved and their repetition frequency affected speech rate. Second, we used another linear mixed-effects model to analyze the effects of reattaching and its interaction with the repetition pattern on speech rate. Again, we used the maximal random effects structure, which included reattaching, repetition pattern, and their interaction as by-subjects random slopes and the counterbalancing lists as the by-item random slope.

Results

The first analysis examined whether the number of unique CVs and tones in a sequence, as well as their repetition pattern, affected speech rate. We found a significant main effect of the number of distinct CVs on speech rate ($\chi^2(1) = 28.57$, $p < .001$), but not of the number of tones, and neither interacted with repetition pattern (*ps* > .05). Participants were about 20 ms faster at producing sequences with two CVs versus one, regardless of the repetition pattern. Specifically, the average production times were 322 ms for one CV versus 300 ms for two when the repetition pattern was ABAB (i.e., 22 ms difference; $t = 3.94$, $p < .001$), and 318 ms for one CV versus 293 ms for two when the repetition pattern was ABBA (i.e., 25 ms difference; $t = 4.27$, $p < .001$). In contrast, the number of tones did not seem to significantly affect speech rate. The average production times were 308 ms for one tone versus 314 ms for two tones when the repetition pattern was ABAB (i.e., -5.75 ms difference, $t = -1.12$, $p = .27$), and 303 ms for one tone versus 308 ms for two when the repetition pattern was ABBA (i.e., -4.49 ms difference, $t = -0.83$, $p = .40$).

Consistent with the results of Experiment 1, repetition pattern did not reliably predict speech rate. Additionally, we found that participants in Experiment 2 were slower at producing sequences with one CV rather than two. At first glance, this finding seems at odds with Experiment 1's post-hoc results showing similar average production times for sequences with one, two, and four unique CVs. It is important to note that all of the trials that required reattaching in Experiment 2 were sequences with one CV and two tones. If reattaching indeed reliably slowed production times, this feature of the design could have explained the slower average production times for sequences with one CV.

The second analysis sought to confirm whether reattaching reliably slowed speech rate, and whether the effect interacted with reattaching frequency. Table 3.6 and Figure 3.3 summarize the average production times for each of the eight conditions of the number of CV, tone, and repetition pattern. Recall that sequences in conditions with one CV and two Tones require reattaching, whereas the rest do not. We found a main effect of reattaching ($\chi^2(1) = 19.77$, $p < .001$), but it did not interact with repetition pattern ($\chi^2(1) = 0.01$, $p = .93$). Specifically, the average production time was 328 ms when participants had to reattach tone to CV for every word, versus 306 ms when rettaching was not required (i.e., 22.5 ms difference; $t = 3.38$, $p =$

.001). Similarly, the average production time was 322 ms when participants had to reattach for every other word, versus 300 ms when reattaching was not required (i.e., 21.8 ms difference; t = -3.00, p = .004).

Table 3.6

Example trials of the 8 Conditions from Phoneme Set 1

Pinyin	Characters	No. of CV	No. of Tone	Repetition Pattern	Reattaching	Average Production Time
ba2 ba2 ba2 ba2	拔拔拔拔	One	One	N/A	No	316 (16)
ba2 di2 di2 ba2	拔笛笛拔	Two	One	ABBA	No	291 (14)
ba2 ba1 ba1 ba2	拔巴巴拔	One	Two	ABBA	Yes	324 (17)
ba2 di1 di1 ba2	拔低低拔	Two	Two	ABBA	No	294 (14)
di2 di2 di2 di2	笛笛笛笛	One	One	N/A	No	313 (16)
di2 ba2 di2 ba2	笛拔笛拔	Two	One	ABAB	No	302 (15)
di2 di1 di2 di1	笛低笛低	One	Two	ABAB	Yes	329 (14)
di2 ba1 di2 ba1	笛巴笛巴	Two	Two	ABAB	No	301 (15)

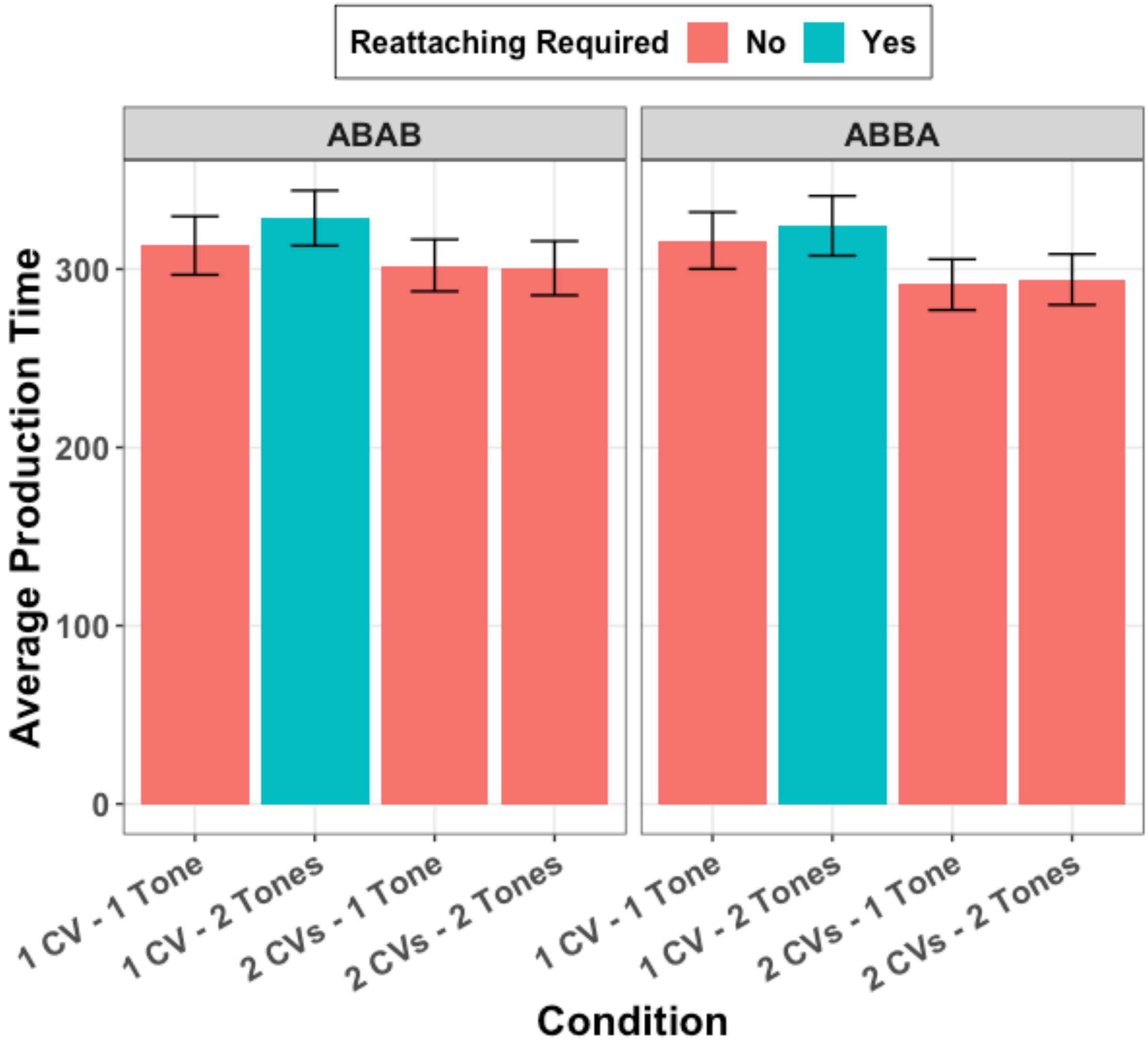

Figure 3.3. Average production times for the eight conditions of CVs, Tones, and Repetition Patterns. Error bars represent standard errors. The blue bars indicate conditions that required reattaching, whereas red bars indicate conditions that did not require reattaching.

In sum, we found evidence supporting the reattaching hypothesis. In other words, Mandarin speakers are reliably slower at producing sequences in which a CV is paired with more than one tone, potentially due to having to detach a tone and reattach another one to the CV. However, we found that engaging in reattaching for every word versus every other word made minimal difference to production times.

General Discussion

In this study, we reported two experiments that investigated Mandarin phonological representation using a word sequence repetition task. We measured the average production times per syllable in four-word sequences containing phonological units of various repetition patterns and examined what speech planning units are in Mandarin, as well as how they interacted with each other in production.

Experiment 1 explored whether consonants, vowels, and tones function as independent speech planning units in Mandarin, or if some combinations of them are selected and processed together as a unit. The planned analysis revealed a major cross-linguistic difference between Mandarin (as investigated here) and English (as reported in the literature): In Mandarin, neither the number of distinct units in a sequence nor their repetition frequency reliably predicted speech rate. In contrast, previous English experiments reported that speakers were generally faster at producing sequences with fewer distinct units and a higher repetition frequency, compared to sequences with more distinct units and a lower repetition frequency (Sevald & Dell, 1994 and Sevald et al., 1995). This suggests that Mandarin consonants, vowels, and tones are likely not represented independently, or they may not show qualitatively similar patterns of repetition benefits as in previous English data.

Given that repetition frequency did not predict speech rate, we conducted two sets of post-hoc analyses to explore the idea that it is perhaps the quality rather than the quantity of repetition across units that affects speech rate. That is, even though the frequency with which each planning unit repeats does not necessarily affect production difficulty, whether different planning units share the same or different repetition patterns might. Indeed, we found suggestive evidence that production times were faster when Tone shared the same pattern with another phonological unit (C or V alone, or both of them), compared to when it did not, suggesting that Tone likely does not function independently of segments. Additionally, when Tone shared the same pattern as both C and V, it did not lead to additive benefit compared to sharing with C or V alone, suggesting that C and V are unlikely to be represented independent of each other either.

Experiment 2 sought to confirm two key insights from Experiment 1. The first is whether CV is represented together as a planning unit in Mandarin, separate from Tone. The second is whether there is a process in phonological processing that involves the pairing of selected CVs and tones, such that changes in the set of CV-tone pairings in a sequence leads to production difficulty. Indeed, the results once again showed that the number of unique CVs or tones in a sequence did not affect speech rate; but if a given CV is associated with more than one tone (i.e., requiring the production system to detach the CV from the previous tone and attach to the new one), speech rate was slow. We coined this *the reattaching hypothesis*.

Our results add to the evidence of cross-linguistic differences in speech planning units suggested by implicit priming paradigms (see Meyer 1990, 1991; Roelofs & Meyer, 1998 for atonal Indo-European languages, Chen et al., 2002 for Mandarin, and O'Seaghdha et al., 2010 for direct cross-linguistic comparisons). Specifically, our findings demonstrate that C and V together form planning units that are separate from tones and selected at an early stage before tones are integrated, which is compatible with O'Seaghdha et al.'s (2010) notion that the proximate unit (i.e., the first selectable phonological unit below the word level) in Mandarin is the atonal syllable. We speculate that Mandarin speakers represent all possible CV combinations in the language in a syllabary, selecting the CV(s) corresponding to each word before passing the syllable(s) on for further phonological processing (such as integrating with tone(s)) and phonetic encoding.

Note, however, that we did not reliably observe form preparation benefits similar to those reported in implicit priming paradigms. Recall that it is commonly found that speakers benefit from knowing in advance the onset segment of the target response in English and Dutch (Meyer, 1990, 1991; Roelofs & Meyer, 1998), or the initial syllable (atonal CV) in Mandarin (Chen et al., 2002; O'Seaghdha et al., 2010), and thus produce the target quicker in those conditions, compared to conditions where little to no information is available for advance planning. If those form preparation benefits were directly transferable to our paradigm, we might have expected to see faster speech rates for sequences that contained fewer planning units (e.g., fewer unique CVs), as they supposedly required less advance planning. On the contrary, we found that repeatedly producing the same CV and tone was not necessarily faster than producing two or more distinct CVs in a sequence. Also, producing one CV that was paired with different tones in a sequence was much slower than producing two CVs that were paired with the same tones.

However, there are many factors that could potentially affect whether form preparation benefits are observed or not. In particular, a long preparation period may obscure any subtle preparation benefits, which may have occurred in our experiments. With the eight-second silent preparation period, speakers may have been given more than enough time to complete any advance planning possible to produce the target responses in our experiments, which is in stark contrast to no preparation period in between the presentation of the prompt and the production of the target in implicit priming. If the preparation period indeed absorbed any form preparation benefits, then we may only be able to observe significant production difficulties that advance planning may not effectively overcome, such as reattaching CVs and tones on-the-fly. Moreover, the implicit priming paradigm measures onset latency in a one-off manner per trial, which makes it sensitive to speech preparation. On the other hand, our paradigm measures average production time, which includes both preparation and execution time, potentially obscuring some preparation benefits.

Given our conclusion that Mandarin CVs and tones are represented separately but still dependent on each other during processing, we return to the discussion of early versus late encoding. The best evidence for early encoding accounts comes from the comparison of error types and frequencies between segments and tones (e.g., Wan & Jaeger, 1998; Alderete et al., 2019). Unfortunately, by design, the majority of our data came from error-free production, and the small number of errors we obtained did not offer enough power to conduct such systematic comparisons. Hence, we did not find strong and direct evidence against tones being involved in early encoding.

That said, the process of integrating CVs and tones highlighted by the reattaching hypothesis points towards tones being involved in late encoding. The crucial piece of evidence is that the production difficulty associated with reattaching is anchored on CVs and not on tones. That is, what determines whether a sequence is difficult is whether each CV is associated with multiple tones, but not whether each tone is associated with multiple CVs. In fact, a single tone being associated with multiple distinct CVs did not lead to production difficulties. These observations illustrate that CVs likely have a privileged status in planning, such that they are selected first and later integrated with tones. If tones had been selected before CVs or simultaneously involved in early encoding, we would have expected active competition in selection between tones such that producing the same tone with multiple CVs in a sequence would have led to similar production difficulties.

Our results are the most compatible with production models that assume tone to be part of a metrical frame (e.g., Chen et al., 2002 and O'Seaghdha et al., 2010), similar to stress in

Indo-European languages. These models propose that while atonal CV representations are proximal to lexical retrieval, their tonal counterparts are penultimate representations that precede actual articulation. In between those two steps lie the processes of linearization and tone value assignment, during which selected CVs are slotted into structural frames with diacritics indicating the tone values inherent to the frames. Returning to our reattaching idea, because CV is the element that is actively inserted into a structural frame (while tone has become an inherent feature of a frame at this time), the process of detaching a selected CV from a frame and reattaching it to another is costly.

A question still remains: If reattaching is costly, why is there no difference between reattaching more often versus less? Specifically, the average production time for reattaching every word (ABAB pattern, e.g., "笛低笛低" *di2 di1 di2 di1*) was 328 ms, which was comparable to that for reattaching every other word (ABBA pattern, e.g., "拔巴巴拔" *ba2 ba1 ba1 ba2*), which was 322 ms. Similar to the lack of form preparation benefits, it could be that the difference in difficulty between reattaching versus not is significantly larger than that of reattaching every other word versus every word, and that the long production period obscured more subtle effects such as the latter. Another possibility is that reattaching is an all-or-none mechanism, which leads to similar levels of difficulty regardless of frequency once it is engaged. Future research may attempt to address the limitations of our work by using a shorter preparation period and production procedure, which may lead to increased sensitivity towards relatively subtle preparation benefits and production slow-downs.

In conclusion, the current study reported corroborating evidence for the notion that atonal syllables function as proximate units in Mandarin Chinese production. Additionally, based on data showing that speakers were slower to produce sequences that required reattaching between

CVs and tones, we suggest that the integration process between CVs and tones occur relatively late in phonological processing, supporting late encoding accounts.

References

Alderete, J., & Davies, M. (2019). Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and Speech*, *62*(2), 281-317.

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, *191*, 103952.

Bates, D., Mäechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.7.

Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied psycholinguistics*, *20*(2), 289-301.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*(4), 751-781.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27-52.

Gandour, J. (1977). Counterfeit tones in the speech of Southern Thai bidialectals. *Lingua*, *41*(2), 125-143.

Garrett, M. (1984). The organization of processing structure for language production. In D. Caplan, A. R. Lecours, & A. Smith (Eds.), *Biological perspectives on language* (pp. 172193). Cambridge, MA: MIT Press.

Halle, M. & Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report, 101*. 198-212.

Kember, H., Croot, K., & Patrick, E. (2015). Phonological encoding in Mandarin Chinese: Evidence from tongue twisters. *Language and Speech*, *58*(4), 417-440.

Leben, W. (1978).The Representation of Tone. In V. Fromkin (Eds.), *Tone: A Linguistic Survey* (pp.177-220), Academic Press, New York

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(5), 524-545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69-89.

Moser, D. (1991). *Slips of the tongue and pen in Chinese*. (Sino-Platonic Papers, No. 22). Department of Oriental Studies, University of Pennsylvania, Philadelphia, PA.

Odden, D. (1995). Tone: African languages. In J. A. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 444-475). Oxford: Blackwell.

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282-302.

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research*, *57*(1), 22-37.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249-284.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 922-939.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91-127.

Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas?. *Journal of Memory and Language*, *34*(6), 807-820.

Shen, J. (1993). Slips of the tongue and the syllable structure of Mandarin Chinese. In S.-C. Yau (Ed.), *Essays on the Chinese language by contemporary Chinese scholars* (pp. 139-161). Paris: Centre de Recherche Linguistiques sur L'Asie Orientale. École des Hautes Études en Sciences Sociales.

Wan, I. P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, *15*(3), 417-461.

Yip, M. (1995). Tone in East Asian languages. In J. A. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 476-494). Oxford: Blackwell.

Acknowledgements

Chapter 3, in full, is a reprint of the material currently being prepared for submission for publication. Lau, Sin Hang; Li, Chuchu; Ferreira, Victor S. The dissertation author was the primary investigator and author of this paper.

CONCLUSION

This dissertation presents three lines of research that address the general question of whether seemingly distinct linguistic structures are linked cognitively.

Chapter 1 examined whether two typologically relevant linguistic features are linked when learning structural alternations in a novel language. Specifically, we investigated whether learners condition their generalization to be verb-wise conservative or not based on the evidence of scrambling (i.e., subject-crossing alternations) or the lack thereof in their input. Results from our artificial language learning experiments suggested that learners indeed conditioned their responses accordingly (i.e., verb-wise conservative when they saw no evidence of scrambling, and generalized across verbs when they saw scrambling), suggesting that the two linguistic features are linked. We attributed the link to a linguistic bias in favor of maintaining a clear subject-predicate boundary in word orders, and the violation of such a preference is due to motivations that are not likely to be constrained by particular verbs, such as pragmatics.

Chapter 2 investigated whether seemingly analogous attachment structures in words and sentences are linked by shared structuring mechanisms. Using structural priming, we tested whether attachment manipulations in relative clause sentences and noun phrases influenced subsequent attachment preferences within- and across-grain size. We found that sentences primed sentences but not words, and vice versa, suggesting that attachment structures constructed by linguistic units of different grain sizes are not linked by shared structuring mechanisms. We speculated that meanings of words may not be computed on-the-fly by combining each morpheme, while the entity that a relative clause modifies likely needs to be identified anew each time.

Chapter 3 reported a mostly exploratory investigation on how segments (consonants and vowels) and tone are represented and integrated in speech in tonal languages. In a speeded repeated production task, we compared the average production time for trials in which potential planning units varied in repetition patterns and confirmed that the basic planning unit in Mandarin Chinese is likely to be the whole syllable (rather than segments as in non-tonal languages like English). More importantly, we found that the number of unique syllables and tones in a sequence did not reliably predict speech rate. Instead, production was slowed by pairing a syllable with more than one tone in a sequence. We take these findings as evidence that segments and tones are represented independently but integrated at a relatively later point in phonological processing.

All in all, we conclude that there is more to linguistic structure than meets the eye – features and structures that look different could be explained by the same latent cognitive factor (Chapter 1), whereas those that appear to be similar may not be linked as predicted at all (Chapter 2 and 3). By finding structures that may be implicitly linked and testing specific predictions using various paradigms, we furthered the understanding of the cognitive mechanisms behind syntax, morphology, and phonology in different ways. Specifically, we gathered evidence in support of an account that links two seemingly unrelated typological patterns, challenged intuitive assumptions about words and sentences, and elucidated the mental representation of phonological units in languages of unique features that are not as well-understood. Investigating the relationship between structures proves to be a promising avenue for building a more comprehensive and inclusive model of language.