UNIVERSITY OF CALIFORNIA

Santa Barbara

Effects of nonlinguistic context on language production

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Brendan Barnett Barnwell

Committee in charge:

Professor Stefan Gries, chair

Professor Patricia Clancy

Professor Sandra Thompson

Professor Mary Hegarty

June 2014

The dissertation of Brendan Barnwell is approved.

Patricia Clancy

Sandra Thompson

Mary Hegarty

Stefan Gries, chair

June 2014

Acknowledgements

I am grateful to my committee members for their support and insight, and especially to my committee chair, Stefan Gries, whose class on exemplar theory was the genesis of this project. My fellow graduate students in the linguistics department also provided helpful feedback and discussion. Special thanks are due to Kay Bock for allowing me to use the pictures from her syntactic priming studies. Bob Kennedy and Laura Robinson were kind enough to offer extra credit to students in some of their classes as compensation for participation in the experiment. Finally, my sincerest thanks to my tireless research assistant, Mike LaRosa, who handled most of the logistics of scheduling and carrying out the experimental sessions.

Vita of Brendan Barnett Barnwell

June 2014

Education

Bachelor of Science in Mathematics (highest honors)

University of California, Santa Barbara, June 2006

Master of Arts in Linguistics

University of California, Santa Barbara, June 2010

Doctor of Philosophy in Linguistics

University of California, Santa Barbara, June 2014 (expected)

Professional Employment

2008-2010: Teaching assistant

Department of Linguistics, University of California, Santa Barbara

Summer 2011: Instructor

Department of Linguistics, University of California, Santa Barbara

Awards

2007-2011: BP Foundation Scholarship (Scholarship Foundation of Santa Barbara)

2010-2011: UCSB Dean's Fellowship

2011-2012: UCSB Graduate Humanities Research Assistantship

2012-2013: UCSB Dean's Fellowship

2013: UCSB Graduate Division Dissertation Fellowship

Publications

2011. Entextualized humor in the construction of scientist identities among U.S. undergraduates. *Anthropology and Education Quarterly* 42(3): pgs. 177-192 (with Mary Bucholtz, Jung-Eun Janie Lee, and Elena Skapoulli).

2012. Itineraries of Identity in Undergraduate Science. *Anthropology and Education Quarterly* 43(2): pgs. 157-172 (with Mary Bucholtz, Jung-Eun Janie Lee, and Elena Skapoulli).

2013. Perception of Prosodic Boundaries by Untrained Listeners. In Beatrice Szczepek Reed and Geoffrey Raymond (eds.), *Units of Talk – Units of Action*. Amsterdam: John Benjamins.

Fields of study

Major Field: Psycholinguistics

Studies in First Language Acquisition with Professor Patricia Clancy

Studies in Field Linguistics with Professors Matthew Gordon and Sandra Thompson

Studies in Corpus Linguistics and Statistics with Professor Stefan Gries

Abstract

Effects of nonlinguistic context on language production

Brendan Barnwell

Recent usage-based approaches to linguistic theory have claimed that linguistic processing is driven by domain-general cognitive abilities which operate on a rich memory store that retains all the details of every experience with language, extracting patterns from these experiences purely on the basis of regular patterning. It has also been claimed that these mechanisms operate similarly on all levels of linguistic structure and at all stages of the human lifespan. Taken together, these claims imply the hypothesis that any dimension of experience can influence the linguistic knowledge and behavior of any language user. This dissertation tests this hypothesis.

A series of experiments were conducted, each consisting of a prime phase and a test phase, using native English-speaking participants. In the prime phase, participants were exposed to combinations of linguistic structures (active and passive voice) and nonlinguistic contextual elements (colors, background music, sounds, or physical environments). The linguistic and nonlinguistic components of the experiences so created bore no semantic relationship to one another, but the pattern of cooccurrence between them was completely regular and reliable, such that, for each experimental participant, a particular syntactic voice always occurred in a particular nonlinguistic context. Participants then performed a picture description task, in which each picture was accompanied by one of the nonlinguistic contexts to which they had previously been exposed. The hypothesis was that, when describing each picture, participants should be more likely to use the syntactic voice which had previously been associated with the nonlinguistic context which accompanied the picture.

This hypothesis was not supported by the data. Instead, it was found that the only consistently significant factor influencing the syntactic voice of participants' responses was the syntactic voice of their own previous responses: people were more likely to keep using whatever voice they had already been using. In addition, in every experiment, it was

found that the results were most accurately characterized by an extremely simple model using only subject-specific and picture-specific baseline response rates. These results suggest that usage-based theories have been overly optimistic in asserting that regular patterns of experience alone are sufficient to explain linguistic knowledge and behavior. Instead, it is argued that more specific constraints on linguistic processing mechanisms are needed in order to provide a full-fledged, causal account of how people's experiences affect their mental representations of language.

Contents

Contents **viii**

1 Introduction **1**

1.1 Components of the usage-based theory . . . 3

1.2 Structural priming . . . 28

1.3 Context-dependent memory . . . 31

1.4 Summary . . . 37

1.5 The present study . . . 39

2 Methods **41**

2.1 Materials . . . 41

2.2 Participants . . . 42

2.3 Procedure . . . 42

3 Experiment 1a: Color **46**

3.1 Participants . . . 46

3.2 Methods . . . 46

3.3 Results . . . 46

3.4 Discussion . . . 55

4 Experiment 1b: Random-prime color **57**

4.1 Participants . . . 57

4.2 Methods . . . 57

4.3 Results 57
4.4 Discussion 60

5 Experiment 2: Sound **61**
5.1 Participants 61
5.2 Methods 61
5.3 Results 61
5.4 Discussion 67

6 Experiment 3: Music **68**
6.1 Participants 68
6.2 Methods 68
6.3 Results 70
6.4 Discussion 71

7 Aggregate analysis of Experiments 1a-3 **74**
7.1 Results 74
7.2 Discussion 78

8 Experiment 4: Room **79**
8.1 Participants 79
8.2 Methods 79
8.3 Results 81
8.4 Discussion 85

9 Power analysis **87**

10 General discussion **92**

11 Conclusion **104**

References **114**

Chapter 1

Introduction

Recent usage-based theories in linguistics have posited that mental representations of language arise from domain-general pattern-recognition processes operating on memories of specific experiences with language. There is considerable evidence for many parts of this claim: evidence that some pattern-recognition processes are domain-general, evidence that language use involves engagement of nonlinguistic cognition, evidence that detailed memories of language are stored. However, the theoretical conception of the usage-based framework involves a synthesis of these disparate observations into a more cohesive whole, and a significant gap remains between the evidence supporting the parts and evidence supporting the overall theory that integrates them. That is, results from several different lines of research appear to be converging toward a particular view of how language structure is mentally represented and constructed, but these lines of research do not quite meet in the middle to provide a unified account of how all of the pieces work together. Substantial questions remain about how claims and findings from different research areas can be integrated into a broader characterization of how language structure arises through usage.

An important consequence of this is that the boundaries of the theory remain untested in some places. Memory and pattern-recognition are extremely broad areas of cognition; when combined with claims of domain-generality, they give rise to a theory that risks being "too powerful" in the sense that it may make a vast array of predictions beyond what it was intended to explain. This does not mean that the overall theory is incorrect, but it may mean that the details need to be ironed out. For instance, what are the constraints on the patterns that can be learned according to the theory? What are the constraints on what kinds of elements may constitute the patterns? What are the constraints on what aspects of experience are recorded in memory and accessible to the pattern-recognition system? If the domain-general cognitive system can recognize a given pattern, does that automatically mean that that pattern is available for incorporation into language, or is the linguistic system selective, such that there are patterns which can be learned but which cannot be integrated into language?

Finally, despite the appeal to domain-general faculties of memory and pattern recognition, discussion of these theories within linguistics has remained almost wholly limited to linguistic memory and the recognition of linguistic patterns. Although it is obviously natural for linguistic research to be focused on language, it is also true that recruitment of domain-general processes brings with it an accountability to the total facts about those processes, as it were in all their domain-general glory. Thus, for instance, if it is claimed that memory is crucially involved in the construction of linguistic representations, it is incumbent upon the claimants to explain how these processes of memory are related to and consistent with what is known about memory in general. In other words, simply saying that a process works by means of memory does not preclude the possibility that it relies on a highly domain-specific sort of memory; "memory" as invoked in a linguistic theory is not really a domain-general process unless there is some evidence that known facts about memory in other domains also apply in the linguistic context.

In this dissertation, I address some of these questions. The general approach is to take several of the core tenets of the usage-based theory and derive some hypotheses that apparently ought to be true if all of these core tenets are taken at face value. I test these hypothesis using experimental studies that combine paradigms from linguistic work and memory research. To begin with, in the following sections, I describe in more detail the claims of the usage-based theory, as well as the evidence adduced in support of them, and

I identify the areas where the different components of the overall theory do not mesh smoothly.

1.1 Components of the usage-based theory

I use the term "the usage-based theory of language" (or UBTL) to refer to the emerging nexus of theoretical perspectives which posit that language structure is dervied from patterns of use as modulated by domain-general cognitive processing. In the strictest sense, there is of course no "the" (i.e., single) usage-based theory, since the theory is the result of the accumulation of work by many researchers over decades, and these researchers have not always been in exact agreement. Nonetheless, as described above, this accumulation has given rise to a common collection of principles relied on by many researchers pursuing variations on what I call "the usage-based theory". Other authors have used similar terminology to refer to this broad research orientation (Bybee 2007; Bybee and Beckner 2009; Croft 2001; Ibbotson 2013; Langacker 1987; Tomasello 2000). Simply put, the fundamental claim of the UBTL is that "the structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes" (Beckner et al. 2009).

Bybee (2010) is the most direct and comprehensive statement of the overall theoretical perspective, and in other work Bybee (Bybee 2001, 2007; Bybee and Hopper 2001; Bybee et al. 1994, among many others) has been a leading proponent of this view, as well as a leading provider of evidence in support of it. Bybee and Beckner (2009) is a concise statement of the fundamentals of this theoretical approach. Lines of research contributing to the usage-based theory include construction grammar (e.g. Croft 2001; Fillmore et al. 1988; Goldberg 1995, 2006), constructivist language-acquisition work (e.g. Akhtar and Tomasello 1997; Ellis 2001; Goldberg et al. 2004; Lieven et al. 1997; Pine and Lieven 1993, 1997; Tomasello 1992, 2000), cognitive corpus linguistics (e.g. Redington et al. 1998; Stefanowitsch and Gries 2003), functionalist psycholinguistics (e.g. Boyd et al. 2009; Gurevich et al. 2010; Saffran et al. 1999), and, in a broader sense, a wide array of typological/functionalist linguistic literature (e.g. Croft 2003; Givón 1979; Haiman 1985; Hopper 1987; Hopper and Thompson 1980; Thompson 1988, 1998; Tomasello 1998). Some of the earliest work within each of these lines was more isolated from work in the others, but in more recent years there has been a greater degree of integration and combining of the different components (as suggested by the appearance of some researchers' names in more than one of the citation lists above). (Beckner et al. 2009), in an overview of the UBTL position, highlight relevant work in numerous research areas.

I emphasize that, although I will of necessity draw on work by many previous researchers in summarizing the usage-based position, I do not wish to be understood as restricting "the UBTL" to any particular researcher, set of researchers, or ideological camp. Rather, by "the UBTL" I mean, more or less, "everything that the research community currently knows, understands, and hypothesizes about how language is shaped by patterns of use". Thus, in an appropriate metaphorical twist, the UBTL is itself emergent — although much of the explicit articulation of the theory's conceptual machinery has come fairly recently, that machinery was not recently invented but built up gradually over decades, and continues to evolve. Importantly, although I will have much to say about problems within the UBTL as it exists now, the proposed solution to these problems maintains the fundamental goal of explaining language via usage, and thus this dissertation is (or attempts to be) another step in the UBTL's evolution, not a move away from the UBTL or a proposal of an alternative.

Four principles of the UBTL are of central concern for present purposes: first, the theory takes domain-general cognition as a starting point for the creation of linguistic structure; second, the theory regards memory of specific experiences with language as the raw material from which structure is extracted; third, the theory essentially rejects a priori distinctions between levels of linguistic structure, arguing that all levels are represented similarly; and fourth, the theory posits that the mental representations of language arising from domain-general cognition and memory are subject to change throughout the lifespan (not only during language acquisition in childhood).

1.1.1 Domain-general cognition

Let us turn our attention first to the issue of domain-general cognition. It is important to note that the UBTL does not take it as a proven fact that domain-general cognition can explain language; nor is domain generality an "assumption" in the sense that it is taken as an axiom that is treated as true without proof. Rather, taking domain-general cognition as a starting point is a practical strategic choice made to minimize the likelihood of overlooking parsimonious and maximally explanatory theories:

> [A]re the processes that give us linguistic structure specific to language or are they processes that also apply in other cognitive domains? The best strategy for answering this question is to start first with domain-general processes and see how much of linguistic structure can be explained without postulating processes specific to language. If this quest is even partially successful, we will have narrowed down the possible processes that have to be specific to language. The opposite strategy of assuming processes specific to language will not lead to the discovery of how domain-general processes contribute to linguistic structure. (Bybee 2010: 6-7)

More succinctly, the UBTL takes up the challenge issued by Lindblom et al. (1984, emphasis in the original) to "*derive language from nonlanguage*!"

Thus, domain-generality is not really part of the content of the UBTL so much as a guiding principle of research practice. The UBTL aims to explain as much as possible in terms of domain-general processes, but is not committed to the claim that all of language, or even any particular aspect of language, can necessarily be so explained.

Since domain-generality is not part of the content of the UBTL, it is not subject to falsification as such. In principle, the UBTL could evolve into something that relies totally on language-specific mechanisms, with no role left for domain-general cogition. Because the UBTL is neutral on the actual importance of domain-general cognition, it requires affirmative demonstrations of that importance. Since the goal is to "see how much of linguistic structure can be explained without postulating processes specific to language", the UBTL puts the onus on its own practitioners to demonstrate just how much of linguistic structure can be so explained. Many such demonstrations have in fact been offered, and some will be mentioned in the following discussion.

Bybee (2010: 7-9) identifies five domain-general processes of interest to the UBTL: categorization, chunking, rich memory storage, analogy, and cross-modal association. I have already identified memory as a separate key component of the UBTL, and I defer discussion of it for the moment. Among the remaining features, the two that are of greatest relevance to this dissertation are cross-modal association and chunking.

Cross-modal association

Cross-modal association is the ability to form associative links between disparate aspects of experience. Bybee as well as Ellis (1996) link this to the Law of Contiguity. Although the concept goes back to Aristotle's *On Memory and Reminisces* and perhaps even to Plato's *Phaedo*, the more modern touchstone for associationist theory is William James (1890, chap. XIV), who gave the name "Law of Contiguity" to the principle that

> objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also, in the same order of sequence or coexistence as before.

Ellis (2001) discusses a number of wrinkles associated with applying this law. In particular, there is the question of exactly which elements of experience are available for association. According to Ellis (2001, 42, emphasis in the original), "The implicit pattern-detection processes that occur *within* these modalities of representation [i.e., sensory modalities] entail that any such *cross-modal* associations typically occur between the highest chunked level of activated node." Bybee (2010: 8) likewise says that "meaning is assigned to the largest chunk available: a word, a phrase or a construction."

There has been a good deal of work demonstrating this in various contexts. Much of it has focused on associations formed in language acquisition. Notably, Akhtar (1999) developed a "weird word order" paradigm, subsequently used also by others (e.g. Casenhiser and Goldberg 2005; Theakston et al. 2004), in which children heard a syntactic construction that used a word order that was ungrammatical for adults in their language (for instance, SOV order for English-speaking children) while also watching video scenes depicting a particular semantic situation. The construction was exemplified using several different verbs. In experiments including those just cited, it has been repeatedly found that young children will readly associate the "meaning" — that is, the common semantic thread of the visual scenes — with the construction. Crucially, the meaning is linked to the syntactic pattern, not to the individual verbs; children can generalize the syntactic knowledge thus acquired to use the construction with new verbs. A similar paradigm has been used with adult subjects (e.g. Boyd et al. 2009), who are also able to extract the generalization, although less readily. (Adult acquisition of the syntactic frame has only been shown with forced-choice recognition tasks; adults will not spontaneously produce the "weird word order" construction. Apparently, then, even when cross-modal associations can be formed, they may not be fully "available" to linguistic processing, in the sense of acting to influence both comprehension and production.)

Also relevant is work within simulation semantics (e.g. Bergen and Wheeler 2010; Glenberg and Kaschak 2002; Zwaan et al. 2002), which has demonstrated that sensory and motor cognition is activated during language processing. This work suggests that cross-modal associations are formed between, for instance, the word *push* and the physical motion of the arm away from the body, such that comprehension of the word activates motor programs associated with the motion.

Computational modeling also has shown that structure can be extracted from fairly messy input even by algorithms that make use only of general distributional information. Stevenson and collaborators (e.g. Alishahi et al. 2008; Alishahi and S. Stevenson 2008; Fazly et al. 2008; Mathe et al. 2008) have developed computational models which, using very general (i.e., not language-specific) probabilistic procedures, learn word meanings, syntactic categories, and argument-structure generalizations from input consisting of language and a very simple semantic "scene" consisting of just a set of labels of items and events. Yu and colloborators (Yu 2008; Yu and L. B. Smith 2007) have developed similar models, including a model (Yu et al. 2005) that learns word meanings from raw auditory input combined with eye-tracking data — that is, by correlating what is said with what is visually attended.[1] These models obviously do not learn as much or as well as human children do. However, such work represents a significant step in "seeing how much of linguistic structure can be explained without postulating processes specific to language".

Cross-modal association may also be seen as a form of classical (Pavlovian) conditioning (Pavlov 1927). One apparent difference is that in a relatively abstract domain such as language there is no obvious analogue of the Pavlovian "unconditioned stimulus". An unconditioned stimulus is one which produces a particular response "automatically" by tapping into some unconditioned (more basic or innate) response; the classic example of an unconditioned stimulus is the food which caused Pavlov's dogs to salivate, and in this case the unconditioned response is apparently physiological. In contrast, a linguistic stimulus such as a particular word induces no consistent unconditioned response, so it is not obvious how even a sequence of linked conditioning associations (that is, higher-order conditioning) could explain language-relevant associations.

Nonetheless, there is research applying classical conditioning theory to cognition in comparably abstract domains, such as the study of how people formulate causal inferences (Jara et al. 2006; Perales et al. 2004). More generally, Rescorla (1988: 152) argues that Pavlovian conditioning is a rich learning mechanism, far more powerful than suggested by the typical layman's view, in which conditioning is limited to simple transference of reflexive responses from one stimulus to another:

[1] The model of Yu et al. (2005) does involve some language-specific training in that it uses a phoneme identifier to transform the raw audio speech stream into a sequence of phonemes. However, there is no language-specific input at the level of the learned associations. The model segments sounds with a language-specific mechanism, but then learns sound-meaning associations without any language-specific constraints on how to find sound-meaning associations.

> Much modern thinking about conditioning instead derives largely from the associative tradition originating in philosophy. It sees conditioning as the learning that results from exposure to relations among events in the environment.

Of course, "relations among events in the environment" are exactly what the linguistic system is supposed to be sensitive to according to the UBTL. Rescorla also draws links between Pavlovian conditioning and connectionist models of psychology — exactly the types of models advocated by exemplar theorists (Bybee and McClelland 2005; McClelland and Bybee 2007).

Questions about cross-modal mapping still remain. The Law of Contiguity is an extremely general principle which potentially allows any sort of association to be learned. Modern work on associative learning is more sophisticated (see, e.g. L. B. Smith 2000), but the full ramifications of that sophistication are not as yet handled by the UBTL. In particular, L. B. Smith (2000: 171) notes the importance of blocking, which is the fact that "if one has already learned a cue that reliably predicts an outcome, it is harder to learn a *new* cue that predicts the same outcome". According to Smith, blocking means that "associative learning based on the regularities in one's previous experience *alters* what regularities will be learned in the future".

On the simplest level, this raises questions about how associative mechanisms can produce mutually intelligble grammars given that children do not all experience the same stimuli in the same sequence. Perhaps their experiences are "similar enough" — but if so, that requires substantial work to document how similar children's overall experiences are (not just their linguistic input!).

However, of greater concern for the purposes of this dissertation is the question of how association works in the web of multiply contingent predictors that make up the linguistic system. The associations between individual words and their meanings are among the simplest in the scheme of linguistic associations. Can cross-modal association really allow people to learn, for instance, the various semantic and pragmatic factors conditioning use of the passive voice, the transitive construction, etc.? If so, how do people avoid "blocking" later conditioning factors with the first one? For instance, Hopper and Thompson (1980) identify ten different features cross-linguistically associated with the notion of transitivity, and argue that each of these features contributes to how or whether an event will be expressed transitively, with the relative strengths of the contributions varying from one language to another. How is that none of these associations blocks the others? And if blocking does not apply in language as it does in other sorts of associative learning, then do speakers' mental representations of their languages become cluttered with associations that are reliable but not useful for communication? If so, why have there been so few demonstrations of direct associative links between linguistic and nonlinguistic regularities? If not, why not? Or, if cross-modal association is not at work here, then what is the process by which people learn the meanings and pragmatic connotations of syntactic constructions or other grammatical structures?

Some researchers within linguistics have investigated the related phenomena of "statistical preemption" and "negative entrenchment" (Boyd and Goldberg 2011; Goldberg 2011; Stefanowitsch 2008), in which "speakers learn not to use a formulation if an alternative formulation with the same function is consistently witnessed" (Boyd and Goldberg 2011: 55). These studies show that language users are sensitive to contexts where a given structure would be expected to occur (based on available evidence from their experience so far), but in fact does not, and that they treat this kind of nonoccurrence as evidence that the structure is dispreferred or ungrammatical in that context. However, these studies have focused only on language-internal negative entrenchment — that is, where there is particular pattern of cooccurrences among linguistic forms. It is an open question how or whether negative entrenchment extends to extralinguistic associations of the sort that would be mediated by cross-modal association. Moreover, changes in language usage can apparently occur in adulthood even when we might expect the new usage to be blocked via negative entrenchment[2], so even within the language domain there remains doubt about

[2]For instance, Bybee (2010: 118) points out the recent spread of *you guys* as a second-person plural pronoun, even among adults, who presumably have spent much of their lives hearing and using *you* or *y'all* in positions where *you guys* could have occurred but did not.

what associations might or might not block others.

L. B. Smith (2000: 170) also argues that "we learn what we attend to," and Ellis (2001: 42) suggests that "attentional focus in WM [working memory] can result in the formation of cross-modal associations." This complicates the picture for a thoroughgoing cross-modal-association-based theory of language, since it requires an explanation of what it means for language users to "attend to" semantic categories which are commonly encoded by the world's languages but which are abstract — such as transitivity, modality, or aspect — as well as a demonstration that language users actually do attend to those dimensions of experience.

An additional question is which aspects of experience are available to be cross-modally associated with linguistic structure. In numerous writings, Bybee and other UBLT adherents have suggested that the nonlinguistic contexts of utterances are of relevance to pattern extraction. Bybee and Hopper (2001: 9) say that "tokens of experience with language are organized into exemplars on the basis of high similarity of phonetic shape and function or meaning, and such exemplars are tagged for their contextual associations, both linguistic and extra-linguistic." Likewise Bybee (2010: 74) says: "Since constructions relate meaning to form, all syntactic relations in a construction grammar in contrast have semantic import and are grounded in the linguistic and extralinguistic contexts in which they have been used." Despite this, there has been little discussion in linguistics of how exactly the language is matched up with the nonlinguistic context. As Ibbotson (2013) points out, we still need "a mechanistic account of the dimensions over which children and adults make (and do not make) analogies." In the absence of such an account, and of any explicit constraints, the implication seems to be that any nonlinguistic contextual element that reliably cooccurs with a linguistic form is available to be bound with that form into a linguistic form-meaning pairing.

Thus, there is substantial evidence that cross-modal association does occur, but questions remain about how much it can occur and whether it occurs more or less than we would expect. The abstract notion of cross-modal association as described by Bybee (2010, passim) and Ellis (2001), and as leveraged in the acquisition and modeling studies cited above, could in principle allow for far more associations than our languages seem to actually make use of. One possible explanation for this is that such associations cannot be formed because "useless" patterns of cooccurrence between language and nonlanguage simply do not exist. That would certainly be an interesting fact, but as mentioned above, it would require linguists to take a detailed look at the nonlinguistic world in its own right, to see what regularities it offers.

Another possibility is that such regularities do exist, but are filtered out by language-specific processes. It may be that, although the Law of Contiguity allows us to learn any relationship, language is more than just the Law of Contiguity, and restricts us to learning only some relationships. This is certainly a plausible view. On the most general level, pragmatic factors (e.g., Grice's cooperative principle) are likely to prejudice us in favor of certain cues over others. Tomasello (Tomasello 2009) and others have argued that the origin of language lies in such factors. However, if this is the case, then part of the task of developing the UBTL is documenting those prejudices in detail. Importantly, it is not sufficient to show that they are preferentially encoded in language. If the form-meaning pairs that make up a language are supposed to arise from cross-modal association, it would be tautological to say that language users preferentially *associate* the sorts of cross-modal regularities that language prefers. Rather, it must be shown that the nonlinguistic component of the cross-modal association is preferentially *perceived* or *attended* in and of itself.[3]

As mentioned above, these possibilities are not inherently inconsistent with the UBTL, because the UBTL does not require domain-generality. The point of this discussion is simply to show that, although we know cross-modal association does happen, we are still some way from knowing how wide its reach is — that is, what kinds of associations can actually be incorporated into the linguistic system — and exactly how or whether these associations

[3]Of course, it would also be possible to propose that the associative faculty itself is language-specific — that is, that we not only attend preferentially to certain cues, but preferentially encode, rehearse, or otherwise retain certain kinds of cue-target pairs. However, to do so requires an even further weakening of the notion that the cross-modal association is domain-general. An associative mechanism that is specially designed to recognize certain types of associations for the language domain is as much a linguistic mechanism as an associative one.

can successfully build up the complete web of structure and meaning that constitutes a grammar.

Chunking

According to Bybee (Bybee 2010: 7; see also Bybee 2002; Ellis 1996; Haiman 1994; Miller 1956), "chunking is the process by which sequences of units that are used together cohere to form more complex units." Newell (1990: 7) calls chunking "a ubiquitous feature of human memory."

In the context of linguistics, chunking (although not always called by that name) has been cited as an explanation for the gradual conventionalization of groups of words into syntactic constructions (Boyland 1996; Bybee and Thompson 2000; Krug 1998), the phonetic reduction of common sequences of words (Bybee and Scheibman 1999; Hay and Bresnan 2006; Jurafsky et al. 2001) and the grammaticalization of words into morphemes (Bybee 2007; Bybee et al. 1994; Hopper and Traugott 2003). In each case, the chunk develops "autonomy" — that is, it begins to have properties of its own, above and beyond those of its parts (see Bybee 2010, Chap. 3 for discussion of the causes and effects of autonomy).

Evidence that chunking is a domain-general process can be found in the area of artificial-language studies. In these studies, participants are exposed to sequences of symbols that are semantically meaningless but display regular statistical patterns. The participants are then tested by exposing them to new sequences that either do or do not display the same statistical regularities, and seeing if they can distinguish between the two kinds of sequences. It has been shown that infants[4] (e.g. Aslin et al. 1998; Saffran et al. 1999) as well as adults (e.g. Saffran et al. 1999) are able, at above-chance levels, to distinguish sequences that display the regularities from sequences that do not, with only a few minutes of exposure. Notably, the ability to group regularly cooccurring units is not modality-specific: the statis-

[4] The infant studies use typical methodologies for testing infants' discriminative abilities, such as head-turning paradigms.

tical regularities can be learned whether the symbols in the sequence are auditory or visual (Saffran 2002).

In the framework of Bybee (2010), this chunking ability is identified as a separate ability from cross-modal association. The main difference that seems to be recognized (e.g. Bybee 2010; Ellis 2001) is that chunking involves the creation of larger units from coocurrence regularities within a single sequence of smaller units, whereas cross-modal association involves the recognition of a cooccurrence regularity between two separate sequences (e.g., language and visual input). In other words, chunking apparently involves combining units of "the same kind" into one, whereas cross-modal association creates a new unit from two units of different kinds.

However, in principle it is possible that the two are really the same: that is, perhaps cross-modal association is simply cross-modal chunking. To date, no artificial language studies have tested whether these sorts of sequential regularities can be learned when the artificial language is itself cross-modal — the artificial language has always been auditory or visual, but not both. Nonetheless, if cross-modal association and within-modality chunking are both globally available processes for recognizing and processing stimuli, as suggested by Bybee (2010, Chaps. 1 and 3), then there is no reason that people should not be able to learn regularities that cross domain boundaries. In a similar vein, Ellis (2001: 44) argues that links between various kinds of mental representations underlie many if not all levels of linguistic structure:

> Links with conceptual representations underlie reference and grounded semantics. Links with frequent local collocations underlie syntax and idiomatic meaning. Links with local and more distant lexical neighbours underlie lexical semantics. Links between L2 and simultaneously active Ll representations underlie translation and language transfer effects.

Here "links" are central, whether they are with other linguistic elements or with nonlinguistic elements.

In short, if we can recognize coocurrence patterns within and across domains by means of a general chunking mechanism, we should be able to learn patterns in just the same way regardless of whether they cross domain boundaries or not. This is a crucial consequence of the acceptance of both cross-modal association and chunking as fully general cognitive processes, but its significance for the UBTL does not seem to have been recognized within linguistics. What it means is that, in principle, the current UBTL explanation for how people learn linguistic structure actually amounts to a description of how people might learn *any* structure — that is, any statistically reliable pattern encountered in the world.

1.1.2 Memory

Memory plays an important role in the UBTL. In Bybee's (2010: 14) treatment, the central building blocks are exemplar representations, which "are rich memory representations; they contain, at least potentially, all the information a language user can perceive in a linguistic experience." Bybee (2010: 18) argues that individual tokens of experience must be stored, because "there is no way for frequency to matter unless even the first occurrence of an item is noted in memory." By "frequency mattering" Bybee refers to the other domain-general processes, such as chunking and cross-modal association, which are claimed to operate on the rich memory representations, and to be sensitive to the frequency with which particular units cooccur.

Importantly, in Bybee's characterization, the exemplar includes "the meaning, inferences made from this meaning and from the context, and properties of the social, physical and linguistic context" (Bybee 2010: 14) — that is, nonlinguistic aspects of experience are stored along with linguistic elements (see also Beckner et al. 2009). This information presumably needs to be available in order for cross-modal association to occur, as discussed above. Although the UBLT thus requires a domain-general memory store for processing to operate on, there is little discussion of the nature of this memory store as it relates to nonlinguistic information. There is considerable evidence that phonological/phonetic detail is stored (e.g. Bybee 2001; Hay and Bresnan 2006; Pierrehumbert 2001), and there is also evidence for memory effects on other linguistic levels (such as syntax, e.g. Bybee 2007; Bybee and Thompson 2000), but the nonlinguistic component of the putative exemplars is rarely mentioned.

Work in the UBTL tradition sometimes blurs the distinction between memory and frequency of use. The idea seems to be that memory is the means by which frequency can exert an effect on mental representations. However, although frequency is typically computed for a particular exemplar that is determined to already exist in the system, Bybee's argument would also imply that the rich-memory store must retain all detail even in order to create a new exemplar whose frequency can be subsequently computed.

For instance, Bybee (2010: 19) says that "each of the phonetic forms of a word that are distinguishable are established in memory as exemplars". Likewise in Pierrhumbert's (2001: 141) exemplar model of phonology, "the JND (just noticeable difference) for f0 in any given part of the range is determined by the resolution of the anatomical and neural mechanisms which are involved in encoding f0. Thus, it is reasonable to suppose that speech tokens differing by less than one JND in f0 are stored as if they had identical f0s." Clearly here the only detail that does not make it into the exemplar is that which cannot be perceived in the first place. However, in discussing syntactic exemplars — specifically, the emergence of *in spite of* as a syntactic constituent with its own properties that masks its internal prepositional-phrase structure — Bybee (2010: 138) says that "repetition of the phrase also sets it up as an exemplar in its own right (because it is a chunk)". By Bybee's argument, repetition of the phrase cannot matter unless each repetition was already stored in memory *before* it was set up as an exemplar in its own right; the system would never know when to create a new exemplar unless it kept memories of previous experiences and compared them to determine whether any were similar enough to warrant the creation of a new node in the "network of associations" (Bybee 2010: 22).

Again, there are some possible ways out of this conundrum. One is to say that the storage of all phonetic detail takes place or is relevant only for phonological exemplars, and that the exemplar system for syntax works differently, with its network of associations being determined by some other mechanism. As will be discussed in a later section, however, positing separate processes for different levels of structure in this way is at odds with other tenets of the UBTL. At the least, the discussion and use of similar network diagrams in (Bybee 2010, 22ff.), however, clearly suggest that the same associative network is meant to be active for all levels of linguistic structure. Moreover, the validity of the exemplar model as a domain-general explanation for language structure is seriously weakened if it does not even generalize to different levels of structure within language.

A second option would be to distinguish chunking as a separate cognitive operation which does not rely on memory per se in the same way that exemplar categorization does. However, if this option is taken, it is unclear why the additional memory-reliant exemplar model is needed at all. If chunking can already identify a sequence of words such as *in spite of*, in spite of the phonetic variation across multiple encounters with that phrase, then why can it not also do the work of exemplar phonology by gathering multiple phonetic tokens into a chunk representing a phoneme? In an earlier section I suggested that the UBTL presently does not clearly distinguish chunking from cross-modal association; here we see that it does not clearly distinguish chunking from memory-reliant exemplar categorization either.

This is not to say that the model is therefore incorrect. It is clear that, even if some theoretical work is duplicated by chunking, cross-modal association, and rich memory storage, it is because they are all working in the direction of the overall claim of the UBTL: that language structure is emergent from experience. My point is, again, simply to say that the generality of the proposed mechanisms is a two-edged sword; the mechanisms are powerful, but they can be so powerful that it is not clear why one or two of them alone cannot do all the explanatory work. The solution is not necessarily to cut back the theory in the name of parsimony; rather, as this dissertation will argue, empirical results may suggest that the theory needs more specifics, not more generality.

One reason that the additional utility of the rich-memory-based exemplar system is unclear is that the nature of the memory system itself is not made explicit. It is somewhat puzzling that work within linguistics has relied on the concept of memory while drawing little connection to the vast psychological literature on memory.

On the most general level, it is appropriate to ask, if the UBTL talks about memory, what kind of memory is it talking about? Within psychology, Squire's (1987) categorization of types of memory has been influential. Squire distinguishes short-term memory (which roughly corresponds to what is now more commonly called "working memory") from long-term memory. The latter is divided into procedural memory — "memory that is contained within learned skills or modifiable cognitive operations" — and declarative memory — "memory that is directly accessible to conscious recollection" (Squire 1987: 152) — with declarative memory itself divided into episodic memory — "memory for past events in an individual's life" — and semantic memory — "knowledge of the world ... [which] does not refer to particular events in a person's past" (Squire 1987: 169). More recently "declarative memory" has sometimes been referred to as "explicit memory", and Squire's procedural memory has been grouped into a complementary category of "implicit memory" which also includes perceptual priming, classical conditioning, and other nonconscious mechanisms. This framework has evolved over time (see, e.g. Schacter and Tulving 1994; Squire 2004; Squire and Zola 1996) and alternatives have been proposed (see, e.g. Henke 2010), but the basic distinctions remain valuable.

Where does the sort of memory posited by the UBLT fall in such a scheme? Although Ellis (2001) has discussed the relevance of working memory for the actual online execution of cross-modal association and chunking, it is clear that most discussions of "rich-memory representations" as used in exemplar models are referring to some sort of long-term storage. On the one hand, phrases such as "specific experiences with language" (Bybee 2010: 53) suggest an analogy with episodic memory. Port (2010: 44) even explicitly suggests "a rich memory for speech material, resembling episodic memory for everyday events and activities". On the other hand, it is equally clear that this rich-memory store is not explicit: people obviously cannot call to consciousness every situation in which they heard a particular word, and in many cases they might be unable to recall any specific instances.

Ullman (2001) has proposed a Declarative/Procedural model in which declarative memory underlies lexical knowledge while procedural memory underlies grammatical knowledge; however, separating words from rules in this way is a clear break with the UBTL, which (as has been mentioned but will be discussed more fully later) rejects the notion of such boundaries in favor of gradience between lexical and grammatical facets of language.[5] Kidd and Kirjavainen (Kidd and Kirjavainen 2011; see also Lum and Kidd 2012) contrast this model with a "single-route approach" which is essentially in the vein of the UBTL. However, they argue that

> since the single-route model argues for a prominent role for vocabulary acquisition in the development of morphological knowledge, and since vocabulary knowledge is essentially declarative knowledge, a single-route model that incorporates long-term memory processes would identify a prominent role for the declarative system.

Although this apparently posits declarative memory as the type hypothesized by the UBTL, this cannot really be the sort of "rich-memory store" that forms the basis of exemplar models. What is declarative about vocabulary knowledge is precisely the abstracted "output" of the exemplar system, and what is nondeclarative is precisely the detailed token-by-token records which are claimed not to be thrown away. In other words, although the exemplar model *produces* declarative knowledge (i.e., we can say what a word means), its internal workings cannot be based on declarative knowledge (i.e., we cannot say what specific instances of language use resulted in us knowing what that word means).[6]

Reaching beyond language-related work on memory, some characterizations of the functions of implicit memory show a conceptual link with the pattern-recognition operations posited by the UBTL. Squire and Zola (1996: 13515) discuss neuropsychological evidence

[5]For instance, Croft (2001) states that "there is a *continuum* between the lexcion and syntactic constructions" (emphasis in the original. For useful perspective on this issue, see for instance 2005, 2007, and 2007.)

[6]Nonetheless, Kidd and collaborators (Kidd and Kirjavainen 2011; Lum and Kidd 2012) found evidence that declarative memory was correlated with morphological acquisition in English and Finnish, suggesting that this memory system does play some role.

that implicit memory (which they call nondeclarative memory) is essential for "probabilistic classification learning" in which

> subjects attempt to learn a set of associations. The associations are not obvious, and they are difficult to memorize because of the probabilistic structure of the task. As a result, information from a single trial is not as reliable or useful as information accrued across many trials.

This obviously has much in common with the sort of pattern extraction that is claimed to underlie language in the UBTL. Squire and Zola also note that "artificial grammar learning is nondeclarative", as are "category learning and prototype extraction"; again, these are the types of processes that UBTL proponents say are at the root of linguistic knowledge.

Here I will briefly draw attention to one particular phenomenon within language which some have argued is a form of implicit learning, namely structural priming. A fuller discussion of structual priming will be given later. For the moment, suffice it to say that it is a phenomenon in which use of some particular language structure — for instance, the English passive voice or double-object construction — facilitates future processing or production of that structure. A number of researchers (e.g. Bock and Griffin 2000; Chang et al. 2000; Savage et al. 2006) have suggested that this phenomenon is the result of implicit learning. Structural priming is the specific phenomenon investigated in this dissertation, as will be described in later sections.

On the whole, it seems most plausible that, if the UBTL is on the right track, the memory system involved in linguistic congition is some sort of implicit memory. However, UBTL practicitioners (e.g. Bybee 2010; Port 2010) have tended to emphasize the "richness" of the memory store and its storage of "experiences"; this terminology misleadingly suggests a more episodic store.

Some of this confusion may be merely due to a terminological issue which sometimes arises in distinguishing the different types of memory. Even in the psychological literature there is sometimes a blurring of the distinction between the directness or indirectness of the *task* and the explicitness or implicitness of the memory system that is accessed. S. M. Smith et al. (1990: 230) thus give us the apparently oxymoronic statement that "implicit memory tests reveal episodic memory" — but they then go on to note that the term "indirect measure" unambiguously refers to the task and not to the memory system.

With regard to memory as it figures in the UBTL, it might be useful to follow the terminological approach used by, for instance, Squire and Zola (1996), in which terms for the overall learning effect are mixed with terms for the memory systems. Thus, although their article is about *memory* systems, Squire and Zola (as quoted above) nonetheless say that artificial grammar *learning* is nondeclarative. Most UBTL work on the question to date seemingly implies that the storage of experiences is one thing, and the extraction of the regularities is a separate processing step; but in Squire and Zola's framing, the registering of the regularities *is* the implicit memory — there is no need to deal with the specific memory traces of each learning trial, and thus no need to become entangled in the details about exactly which features of an individual episode are stored and which are not.

It is tautological that every form of memory is "episodic" in the sense that some particular events in the past resulted in the creation of the memory trace. However, this does not necessarily mean that every aspect of those events is part of the memory trace. Likewise, it must of course be true that each linguistic token encountered has an effect on the mental representation, but the mechanism of this may be considered internal to the memory system and need not be specified in a discussion of how that memory system subserves a larger cognitive task (e.g., language). To claim that the language system relies on the "richness" of these individual tokens needlessly exposes the UBTL to questions about how or if the individual tokens can be accessed, when in fact all that is needed is the overall result that the regularities are learned.

Setting aside such issues of presentation, the more significant fact is that UBTL work within linguistics has played somewhat fast and loose with the term "memory". The claim is that various details of experience are stored. However, without grounding such a claim by reference to existing knowledge of memory systems, it is unclear exactly what is being claimed (i.e., whether the knowledge is implicit, explicit, or both) and how or whether the storage of these details is distinct from the process of extracting regularities from them.

As before, this lack of specificity in the theory does not mean it is wrong, but it means its boundaries are currently not well defined. It is, for instance, not clear exactly what sorts of information, and what sorts of regularities in that information, really can be incorporated into the linguistic system. Saying that exemplar representations "contain, at least potentially, all the information a language user can perceive in a linguistic experience" (Bybee 2010: 14) is an extremely broad claim, because people can obviously perceive all sorts of things. The word "potentially" does not make the claim any less broad, but simply hedges it and opens the door for future research to provide a demonstration that a given kind of information is or is not available. In particular, allowing the possibility that linguistic exemplars can contain unconstratined nonlinguistic context opens the floodgates to a multitude of hypotheses about what elements of the nonlinguistic context are or can be relevant, and, under the ground rules of the UBTL, each such hypothesis — that is, each kind of nonlinguistic context — must be evaluated empirically.

1.1.3 Rejection of discrete levels of structure

A third core position of the UBTL is its rejection of the idea that different levels of linguistic structure are governed by separate processes. Rather, in most recent UBTL work, all levels of linguistic structure are taken to be represented by "constructions"; the term derives from construction grammar (see e.g. Fillmore et al. 1988; Goldberg 1995), but the concept has been generalized beyond the syntactic constructions that originally motivated that line of research. Bybee (2010: 56) argues that "chunking and the gradual increase in autonomy has effects at all levels of grammar", and Bybee and Beckner (2009) say that "all of the units of language—segments, phonemes, morphemes, words, phrases, constituents—can be arrived at by the simple categorization process described above [the exemplar model]". Likewise Beckner et al. (2009: 5):

> The basic units of grammar are constructions, which are direct form-meaning pairings that range from the very specific (words or idioms) to the more general (passive construction, ditransitive construction), and from very small units (words with affixes, *walked*) to clause-level or even discourse-level units

Croft (2001: 17) makes the point even more explicit:

> The constructional tail has come to wag the syntactic dog: everything from words to the most general syntactic and semantic rules can be represented as constructions. The final step is to recognize that the internal structure of words are also constructions. [...] The only difference between morphological constructions and syntactic ones is that the former are entirely made up of bound morphemes while the latter are largely made up of free morphemes.
>
> In other words, construction grammar has generalized the notion of a construction to apply to any grammatical structure, including both its form and its meaning. The logical consequence of accommodating idioms in syntactic theory has been to provide a uniform representation of all types of grammatical structures from words to syntactic and semantic rules.

And Bybee and McClelland (2005: 397-398) go so far as to state their position in the most uncompromising italics (emphasis in the original):

> *there is no analysis into units at any level or set of levels that will ever successfully and completely capture the realities of synchronic structure or provide a framework in which to capture language change.*

Needless to say, these are bold statements. They are bold not only because they challenge decades of linguistic tradition, but because they do away with a possible defense mechanism for the theory they advocate (the UBTL). Namely, once it is accepted that there are no a priori distinctions between levels of linguistic structure, it is no longer possible to counter any adverse empirical finding by claiming that the rules are different for that particular sort of linguistic structure. Chunking, cross-modal association, memory (in whatever form) and all the rest of the domain-general apparatus are, according to the UBTL, fully available to operate on anything. In addition, as I have argued at intervals in the preceding sections, since these operations are domain-general and operate over all aspects of experience, they are available to operate on the nonlinguistic correlates of linguistic structure as well as on the linguistic structure itself.

This does not of course mean that any such adverse finding is a death blow to the UBTL, or even a serious wound. To say that there are no *a priori* distinctions between levels of structure is not to say that there are never emergent levels with gradient boundaries but with distinct properties. The point is that any appeal to such levels must always be grounded in specific empirical data that justify the identification of that level as a level. For instance, if it is found that a certain syntactic process does not operate as predicted, there can be no counterclaim that such processes only operate at the level of the lexicon, and not the level of syntax, because in the UBTL there is no such distinction as such; there can still be a counterclaim that in the language(s) investigated, it has been empirically shown that some processes are different than others, and that processes like the one under discussion are different in that they have been empirically shown to operate only on units of a certain size, and that the units under investigation were too big (or too small). Obviously, the available counterclaim is more labor-intensive than the one of which the UBTL has disarmed itself.

This disarmament is both an advantage and a disadvantage. It is an advantage in that it strengthens the theory from a logical perspective: an argument with fewer hypotheses yields more informative conclusions. It is a disadvantage in that it exposes the theory to a much broader range of potential attacks than would otherwise be feasible. Perhaps even this disadvantage can be seen as an advantage, if it leads to a more vigorous and rapid fleshing out of the UBTL.

For present purposes, I point out the level-agnosticism of the UBTL simply to make it clear that the arguments made in the preceding sections are, in the absence of other evidence, prima facie valid for any level of linguistic structure. If cross-modal association can create form-meaning links between words and situations in the world, then it can also create such links between syntax and situations in the world. If chunking can group regularly cooccurring items together, then it can group regularly cooccuring things together at any level of linguistic structure. And, perhaps most importantly, if rich-memory representations contain nonlinguistic contextual information, then we must assume that any level of linguistic structure is potentially contingent on the nonlinguistic context.

1.1.4 Change throughout the lifespan

An additional point made by Bybee (2010: 9) is that "change is postulated to occur as language is used rather than in the acquisition process". Although this position is not made explicit in much UBTL work, Bybee (2010: 114) notes that it is essentially inherent in the usage-based position: "if usage is the basis of grammar and change in the grammar, then there is no a priori reason why change cannot occur over an adult's lifetime." Moreover, the claim is borne out somewhat by some of the demonstrations of domain-general processes discussed in previous sections. Studies have shown that adults can learn artificial-languages (e.g. Saffran et al. 1999) and anomalous constructions within an existing language (Boyd et al. 2009), suggesting that there is scope for continued learning through the lifespan. As mentioned earlier, structural priming has also been thought of as a form of implicit learning, and has been readily demonstrated with adult subjects. Ellis (2001, 2013) has also extended usage-based associative-learning approaches to second-language acquisition.

Of course, it is obvious that adults do not learn as much or as rapidly as children do. A natural explanation for this within the UBTL is that, if mental representations arise from experience, children's mental representations would naturally be much more malleable because they have less experience. Any individual experience that might be registered would carry greater weight for a child, whereas for an adult each experience could more easily blend in with the large store of already-accumulated experiences.

Thus, like the other tenets of the UBTL already discussed, the exact details of this explanation remain to be filled in. Clearly, some change in linguistic representations can be effected in adults; equally clearly, adults' representations are less malleable than children's (see Bybee 2010: 118); the theory does not provide an overall explanation for just what degrees of malleability should be found for which kinds of representations at which points in the lifespan.

1.1.5 Summary of components of the UBTL

To briefly recap: in broad terms, the UBTL proposes that mental representations of linguistic structure arise from the operation of domain-general processes — such as chunking and cross-modal association — on memory representations that potentially include all aspects (linguistic and nonlinguistic) of each situation in which a linguistic item or structure was used or encountered. The UBTL also proposes that these shaping factors are the same for all levels of linguistic structure, and that the representations they create are subject to modification throughout adult life.

There is evidence for each component of this overall theory. However, the posited processes are extremely general in nature, and the extant work, although demonstrating instances of their operation, leave considerable uncertainty about just how broad they may be and precisely how they are supposed to operate in any given situation. When mechanisms of such generality are posited, it is difficult to predict all the possible ways in which they might conceivably interact. In addition, the rejection of a priori distinctions between levels of structure, and of first-language acquisition as the sole locus of change, enlarges even more the field in which these processes are at play.

Recently, Ibbotson (2013) has discussed many of the same issues in a similar review of the UBTL (which he calls "the usage-based approach"). Ibbotson points out that "being able to accommodate a large range of cross-linguistic findings means that stipulating in any detail, a priori, which processing constraints and input properties are most important, and when, is very difficult" and that "usage-based theories need more psychologically plausible models of what gets treated as a chunk when, and hence 'counted' in any distributional analysis”. For the UBTL, “a key part of responding to this challenge will be to specify in greater detail the mechanisms of generalization”. I echo those concerns here.

In addition, I echo the sentiment of Bybee and McClelland (2005: 382) that we should sometimes “ask not what linguistics can do for psychology, but what psychology can do for linguistics.” If the processes underlying language are truly domain-general, then we should be able to “import” knowledge about how those processes work in other domains to make direct predictions about language. The need for this is particularly evident when it comes to the role of memory in the UBTL. Memory is a fundamental psychological notion that has been studied experimentally for over a hundred years. In light of that, it is quite surprising that the exemplar models put forth by Bybee (2001, 2010), Pierrehumbert (2001) and others make no attempt to link their concept of memory with any of that research. Various processes and properties of memory have been deeply investigated within psychology, and if the UBTL really deals with memory *qua* memory, it should be accountable to what is known about memory.

Finally, I wish to emphasize again the nebulous role of nonlinguistic information within the UBTL. The memories are supposed to contain the nonlinguistic context of each linguistic token; the pattern-recognition processes are supposed to be domain-general; and yet there is so far no detailed account of how the exemplar model integrates nonlinguistic experience with linguistic experience. In short, if the UBTL is on the right track, then why does our system of linguistic patterns seem so minimally mixed with the various other patterns that we learn? Why is it that our patterns of usage of syntactic constructions are so closely intertwined with our patterns of usage of particular verbs, and yet so little intertwined with our patterns of eating, sleeping, paying the electric bill, blinking, breathing, or any of the other regularly patterned aspects of our lives? Or is it possible that such patterns *do* exist, but that we are not aware of them, just as we were not, until recently, aware of the subtle language-internal usage patterns which motivated the UBTL?

It is important to recognize that the inclusion of nonlinguistic information in stored exemplars is not an incidental claim that could be withdrawn while leaving the rest of the

UBTL intact. Without inferencing from the nonlinguistic context, the UBTL provides no way for people to learn even single word meanings, let alone apply the complex pragmatic inferencing that is said to underlie language change (e.g. Bybee 2010; Bybee et al. 1994; Hopper and Traugott 2003). It is thus essential that *some* nonlinguistic information be stored; the UBTL simply leaves it open to empirical investigation to determine how much and of what kinds. As will be seen, the central work of this dissertation was to test various kinds of nonlinguistic information to see whether they are so stored.

The point is that if the cognitive basis of the UBTL is truly domain-general, we should not just be able to show that we learn linguistic patterns just like we learn other patterns; we should be able to show that we learn *patterns*, period, whether they consist of language, nonlanguage, or both. The UBTL does not explicitly predict that we should learn such patterns, but in its current state, with its limited specification of how its components interoperate, it implicitly does so predict. A finding that such patterns are learned would be a powerful demonstration of the true domain-generality of the UBTL; a finding that such patterns are not learned would be a useful empirical result about the limits of the sorts of information that can be combined by the pattern-recognition system.

1.2 Structural priming

Having outlined the general perspective of the UBTL, I wish to examine in more detail a particular phenomenon, namely structural priming. Structural priming is a form of the more general phenomenon of priming, in which exposure to a stimulus (the prime) facilitates subsequent processing of a related stimulus (the target).[7] Priming has been demonstrated

[7] A potential terminological stumbling block must be dealt with here. In the psychological literature, there is a distinction between perceptual priming and semantic (or conceptual) priming (see e.g. Tulving and Schacter 1990). All of the kinds of priming discussed in this dissertation fall under the label of semantic priming. Within psychology, the word "semantic" essentially refers to knowledge, and hence "semantic priming" refers to conceptual priming — that is, all priming that is influenced by concepts above the immediate perceptual level (see e.g. Ochsner et al. 1994). However, within linguistics, the term "semantic" has a narrower sense, referring specifically to linguistic meaning, in contrast to various other levels of linguistic structure; thus "semantic priming" refers specifically to a kind of priming in which prime and target bear some relation in meaning, rather than a relation in linguistic form. Therefore structural priming, in which the primed element is a syntactic construction, is semantic priming in the psychological sense but not in the linguistic sense,

for linguistic stimuli that are related semantically (e.g. Meyer and Schvaneveldt 1971), morphologically (e.g. Marslen-Wilson et al. 1994), and phonologically (e.g. Carreiras et al. 2005).

In structural priming, the relation between the prime and the target is one of linguistic structure. Structural priming has typically focused on syntactic relations, although the term "structural priming" allows for other sorts of structural relations (see Pickering and Ferreira 2008). Bock (1986)'s early demonstration of syntactic priming has been influential (although a few earlier studies had addressed the phenomenon, e.g. Estival 1985; Schenkein 1980; Weiner and Labov 1983). In Bock's study, which was done on English, participants heard sentences read aloud and periodically had to describe pictures. The pictures were designed to be describable in either active or passive voice. Before each picture-description task, participants heard a sentence in either active or passive voice. It was found that participants were more likely than chance to describe a picture using the same voice they had just heard; that is, hearing an active-voice sentence primed subjects to use active voice themselves, and likewise for passive voice. The same effect was also observed for dative constructions, in which either the double-object construction (e.g., *Bob gave his friend a book*) or the prepositional dative (e.g., *Bob gave a book to his friend*) could be primed.

There have been numerous subsequent demonstrations of this effect (e.g. Bock 1989; Bock and Loebell 1990; Bock et al. 1992). The priming effect has been shown to be active in dialogue as well as picture description (Branigan et al. 2000; Gries 2005), and has been replicated in Dutch (Hartsuiker and Kolk 1998) and German (Scheepers 2003). It has also been shown for other syntactic constructions, including the order of subject and locative (Hartsuiker et al. 1999), the order of verb and auxiliary (Hartsuiker and Westenberg 2000), and the form of complex noun phrases (Cleland and Pickering 2003).

Pickering and Ferreira (2008), in addition to providing a useful review of structural priming research, discuss theoretical implications of and competing explanations for the

because the crucial shared feature of prime and target is syntax rather than meaning. Because perceptual priming is not germane here, I simply use the linguistic terminology; thus "priming" alone refers to what in psychology would be "semantic priming", and "semantic priming" has the narrow reading as in linguistics.

phenomenon. Of particular note is the possibility that structural priming is a form of implicit learning. Pickering and Ferreira give a clear description of this position (although they do not endorse it):

> The core idea is that for people to produce and comprehend language, they must learn how their different linguistic and language-related representations relate or map onto one other. To produce or understand the word cat, a language user must learn that the meaning of domestic feline maps onto the word cat (which is a noun), which in turn maps onto the sound sequence /kæt/ and so forth. Everyday language experience can drive the learning of this knowledge, if retrieval of the word cat along with its meaning causes the connection or mapping between these representations to become strengthened. Though more abstract, implicit learning of syntax is similar: To produce or understand a passive structure, a language user must learn that certain meaning relationships (typically, of a patient having something done to it by an agent) map onto certain functional elements (subject and oblique objects), which map onto critical features of constituent sequences (how noun phrases and verb phrases are to be configured), and so forth. These mappings are acquired through experience, such that if a person produces or comprehends an utterance in which a particular meaning relationship is expressed through a particular processing sequence corresponding to constructing a passive or active structure, the mappings among the responsible representations should be strengthened, thereby allowing the same meaning relationships to be expressed through the same processing sequences and so the same syntactic features again. By this account, structural priming is a reflection of this extended process: Hearing or producing the prime strengthens the processing sequence that yielded the prime structure, and target processing reveals this strengthened knowledge.

Pickering and Ferreira note several studies advocating this interpretation (e.g. Bock and

Griffin 2000; Chang et al. 2000; Savage et al. 2006).

As suggested by the account above, structural priming is a natural fit for the UBTL. The implicit-learning account of structural priming essentially claims that it is a result of just the sort of pattern extraction posited by the UBTL.

Structural priming has the additional appeal that it can produce effects on actual linguistic production. Experimental results that show performance on a non-production task (such as the artificial-language and weird-word-order studies on adults cited earlier) have an unclear status within the UBTL. Although the UBTL posits that a great deal of information is stored and patterns extracted, it does not make clear predictions about how these representations should affect comprehension versus production. If the theory is fundamentally grounded in usage, which includes both comprehension and production, what are we to make of results which show that people can be "taught" generalizations which they can nonconsciously perceive (at least probabilistically, i.e., by picking the correct picture in a forced-choice task) but will not actually use in speech? The problem is compounded when the comprehension task is remote from realistic activities of language use (e.g., a forced-choice task). Methodologies that demonstrate effects on language production are the most powerful evidence, since a change in production patterns unequivocally shows that mental representation of language has been affected on some level.

As will be described in due course, the empirical investigation of the dissertation was in the form of a structural priming experiment. However, in order to test the role of nonlinguistic information in pattern recognition, the structural priming paradigm was combined with paradigms from psychological work on context-dependent memory.

1.3 Context-dependent memory

Having mentioned the importance of memory within the UBTL, I wish to introduce one memory phenomenon which has been well-studied in psychology and which bears conceptual similarities to the "network model" of the UBTL. This phenomenon is context-dependent memory. Context-dependent memory is the phenomenon that "reinstatement of the incidental environmental context in which learning took place has often been found to improve memory relative to testing memory in a different context" (S. M. Smith et al. 1990; see also S. M. Smith 1994).

Typically, context-dependent memory effects are found by having subjects learn some materials in one context, and then testing their memory for the materials in either the same context or a different context; memory tested in the same context is enhanced relative to testing in a different context. The "context" can apparently be virtually any aspect of the environment; as S. M. Smith (1994: 169) reports, results have been found with contextual differences such as "under water (with scuba gear) vs. on dry land (Godden and Baddeley 1975), with classical vs. jazz music playing in the background (S. M. Smith 1985), in a sensory deprivation flotation tank vs. a lounge (S. M. Smith and Sinha 1987) and, most commonly, in one laboratory room vs. another (e.g. S. M. Smith 1979; S. M. Smith et al. 1978, 1990)". The effect is extensively documented, with work going back as far as Dulsky (1935). S. M. Smith and Vela (2001: 215) conducted a meta-analysis of 75 studies from 41 published articles spanning more than 60 years and concluded: "In spite of some failed attempts to find environmental context-dependent memory effects, it is clear from our meta-analysis that across all reported studies, the effects are reliably found."

Most context-dependent memory research has used verbal stimuli such as word lists as the material to be remembered. However, some studies (Dalton 1993; Krafka and Penrod 1985; Malpass and Devine 1981; S. M. Smith and Vela 1992) have found similar effects for remembering faces, so the effect does not appear to be specific to verbal material.

Context-dependent memory has a clear similarity to some of the mechanisms proposed by the UBTL. In the pattern-rcognition system of the UBTL, it is easy to suppose that contextual elements could form part of the recognized patterns. Moreover, the apparent effect of reinstantiating the context shows that incidental information about the context was "stored" in the way that incidental information about linguistic events (e.g., phonetic detail that does not affect meaning) is supposed to be stored in exemplars according to the UBTL. In particular, the assertion that the UBTL makes use of "memory storage of the details of experience with language, including [...] contexts of use" (Bybee 2010: 7) almost directly suggests that linguistic knowledge should display context-dependent memory effects.

An important issue, again, is whether the context-dependent memories are implicit or explicit — or, alternatively, whether they are revealed by a direct memory task or by influencing performance on an indirect task. The vast majority of context-dependent memory findings have used a direct memory test, such as recognition or cued recall. However, some studies have found effects on implicit tasks. This is important because structural priming is an implicit task: the test is not, for instance, whether the participants can remember having seen passive voice, but whether their having seen passive voice implicitly makes them more likely to use it themselves.

S. M. Smith et al. (1990) had participants memorize a list of word pairs, which included low-frequency homophones of high-frequency words (e.g., subjects saw the word *beet*, which is infrequent compared to the homophonic *beat*). Smith et al. then gave the participants a direct memory task, which was in fact only a decoy task to make the participants believe that their memory would not be further tested. Subsequently, participants left the room and either returned to the same room or went to a different room (this was the context manipulation), at which point they were asked to write down a series of words which they heard spoken on a tape recording. The words spoken included some of the homophones heard before (e.g, subjects heard [bit], which they could have transcribed as either *beat* or *beet*). It was found that participants were more likely to use the low-frequency spelling if they performed this task in the same room where they had initially studied the word pairs. Smith et al. finally asked participants to explicitly indicate words they recognized as having been on the original list. The successfuly primed low-frequency words were not significantly more likely than their unprimed high-frequency homophones to be recognized; that is, although participants used the low-frequency spellings in the transcription task, they apparently did not consciously remember having seen the low-frequency spellings as opposed to their high-frequency counterparts. Smith et al. interpreted this as evidence that the implicit priming effect of exposure to the low-frequency words was context-dependent, in that it was enhanced if subjects stayed in the same room.

Other studies have found conflicting results for context-dependent implicit memory. Jacoby and Witherspoon (1982) reported no effect of context on priming. Amanda Parker et al. (1999) argued that the effect might depend on whether the indirect memory test relied on conceptual or perceptual features of the learned material; they found context-dependence for a conceptual task (category completion) but not a perceptual task (word-fragment completion). However, Andrew Parker et al. (2007) and Mulligan (2011) have challenged those findings.

Two intriguing studies have investigated the effect of an olfactory context on implicit memory. Amanda Parker et al. (2001) had participants solve two Tower of Hanoi puzzles and trace through a maze while exposed to either lavender or lemon scent. Subjects returned the following day to complete the same tasks and were exposed to either the same scent or a different scent; there was a context-dependent effect for the maze task (i.e., it was easier if the same scent was present) but not the Tower of Hanoi. Ball et al. (2010) found olfactory-context-dependent effects on a word-fragment completion task (an implicit measure), but discovered that the effect is sensitive to the particular scents used, specifically to how distinctive and how pleasant they are.

In addition, two studies have demonstrated effects on linguistic production which closely mirror context-dependent memory effects, although only one of these explicitly makes the link to context-dependent memory research. In that study, Horton (2007) investigated whether another person can serve as a "context" which induces an effect on linguistic production. Participants first engaged in a category-cued word-fragment-completion task with two different partners (who were research assistants); the partner would provide a category cue (e.g., "musical instrument"), and the participant would then have to complete the word fragment (e.g., complete B__JO to BANJO). The participant performed this task in two blocks, one with each of the two partners. Subsequently, participants performed a picture naming task, again divided into two blocks, one for each partner. The names of the depicted objects were words that had been fragment-completed in the earlier task, but some words that had been completed with Partner A were named with Partner B and vice versa. The dependent measure was the naming latency, i.e., how long participants took to name the depicted object. Horton found that participants were quicker to name pictures if the same partner was sitting next to them who had been with them when they fragment-completed that word (e.g., a participant who completed BANJO with the "help" of Partner A the would more rapidly name a picture of a banjo if Partner A was also present for the naming task than if Partner B was present). Although participants had better-than-chance explicit recall of which partner had been present when the fragment was completed, there was no significant correlation across subjects between performance on this recall task and their naming latency, suggesting that the reaction-time priming effect was not driven by explicit retrieval of the associated partner's identity. (That is, participants who were better at consciously remembering which partner was there when they saw each word were not necessarily faster to name the depicted objects.)

Horton (2007) suggests that "conversational partners act as contextual cues for the automatic retrieval of associated information just as different rooms or different physical contexts can facilitate memory depending on the type of overlap with the context of encoding", and argues that memory may thus play a role in the activation of information as pragmatically relevant "common ground" (Clark 1996). This is of obvious relevance for the UBTL, since the identity of an interlocutor, and the common ground shared with that interlocutor, are natural candidate for the sort of information that could be stored in an exemplar model as part of the record of the language event.

The final study I wish to discuss in detail is Gurevich et al. (2010). In a remarkable series of experiments, Gurevich et al. provide a counterpoint to earlier findings (e.g. Bransford and Franks 1971; Johnson-Laird 1970; Johnson-Laird and R. Stevenson 1970; Sachs 1967) that suggested that listeners almost immediately discard the surface form of language that they hear, retaining only the meaning or "gist". In these experiments, Gurevich et al. had participants read stories, and then compared their recognition and recall of sentences in the stories as compared to erroneous "recognition" and "recall" of semantically equivalent sentences they had not seen; results were above chance, showing that participants were able to distinguish between sentences they had seen and ones they had not seen, even though the "gist" was the same. The finding is all the more notable in that participants were not told before reading the stories that there would be a memory test later.

Most notable for present purposes, however, was Gurevich et al.'s Experiment 5. Each participant was paired with an experimental confederate (ostensibly another participant), who watched two animated cartoons not visible to the participant and then verbally described them. In fact, each description was a prewritten script; two such scripts were created for each cartoon, and were semantically equivalent but differed in phrasing. (The confederate had practiced speaking the scripts as if naturally describing the cartoon, and in debriefing no participants reported suspicion that the description was staged.) Participants then took a "quiz" in which they had to assess freeze frames from similar cartoons and pick the one that matched the description. In fact, this quiz was merely a cover task; based on supposedly good performance, the experimenter invited the subject back for another, ostensibly unrelated experiment several days later. When the participants returned, they performed the same task, but now in the role of the "describer", watching the cartoons they had previously heard described, and giving their own descriptions. The clauses of the participants' descriptions were compared with those of the description they had heard, as well as those of the semantically equivalent description they had not heard. Despite the fact the participants were not in any way cued to remember the specific linguistic forms they had heard, and despite the fact that they produced their own narration a mean of six days after hearing the confederate's, they were, at higher-than-chance levels, more likely to re-use clauses they had heard in the earlier narration *word-for-word verbatim*, as compared to semantically equivalent clauses they had not heard.

Gurevich et al. (2010) focus on the exemplar-model implications of their findings (namely that the exact linguistic forms heard apparently had an effect several days later, and thus were apparently stored), and shy away from an implicit-memory account, noting that "while participants were not asked to consciously recall verbatim utterances, the task did require that they consciously access the content of the earlier speech." Nonetheless, they note that "retention can be both incidental (no deliberation), and implicitly mediated (no explicit recall)". From a context-dependent memory perspective, Gurevich et al.'s results can be seen as a manipulation of the "semantic context" — that is, the "environment" was the situation depicted in the video. This environment was instantiated during the priming phase, when they heard the narration, and reinstantiated during the test phase by seeing the video; when the context was reinstantiated, incidental features of it (i.e., the specific phrasing of the narration) were reactivated, and implicitly influenced performance on the narration task.

Such studies show that there is scope for context-dependent effects on indirect memory tasks, including language production tasks. Moreover, as I have argued in the preceding sections, the conceptual underpinnings of the UBTL predict that some sort of context-dependent effects should be found. If mental representations of language include the nonlinguistic context, then varying the nonlinguistic context should alter those representations in detectable ways.

1.4 Summary

What I have argued above can be summarized as follows. The UBTL as it presently stands has argued for an explanation of language grounded in very general cognitive mechanisms. In support of this proposal, UBTL adherents have offered ample evidence that many aspects of language comprehension, production, and acquisition can be accounted for by these general mechanisms, without the need for additional mechanisms specific to language or to a particular level of linguistic structure. However, what has not been so well recognized (although see Ibbotson 2013) is that a theory of this generality must also take account of its potential predictions even when they reach *beyond* the realm of what has traditionally been studied within linguistics. It is not sufficient to look at known linguistic phenomena and attempt to explain them by means of the theory; we must also look at the theory itself and see what it predicts or implies, even when these predictions and implications are not something that the theory had intended to explain.

In other words, although the UBTL is a theory developed for language, its conceptual machinery is so general that it potentially explains a great deal more, and it must be evaluated based on how reasonably it explains everything it actually applies to, not only what it wants to explain. Testing the theory only on a subset of its conceptual domain is like gauging one's marksmanship by shooting at a fly with a shotgun; one must check not only whether the fly was hit, but also whether anything else was hit.

The UBTL's postulation that language use is sensitive to all manner of specific cooccurrence patterns is as yet unaccompanied by a full explanation of which cooccurrences win out over which others under which circumstances. As Ibbotson (2013) points out:

> Usage-based approaches might respond by saying that the meaning is the sum total of how the form is used in a communicative context, however for those seeking more detail, usage-based theorists need to provide a more mechanistic account that integrates semantic and formal generalizations. One way to tackle this is to ask, where does the meaning of linguistic form *x* come from in the child's environment?

In addition, as noted earlier, the rejection of distinctions between levels of structure, together with the rejection of acquisition as the sole locus of development of mental representations, means that this question actually becomes "Where does every property of linguistic form *x* come from in a given person's environment?"

The UBTL's reliance on the "environment" means that it offloads into reality a good deal of what more traditional theories keep within assumptions. That is, where other theories might assume that certain sorts of patterns arise due to innate knowledge or learning constraints, the UBTL argues that, as Croft (2003: 102) puts it, "the structure of the language reflects in some way the structure of experience, that is to say, the structure of the world". As mentioned before, this move strengthens the theory from a logical perspective, insofar as it can explain just as much with fewer assumptions. However, it also puts the onus on the theory to demonstrate that the world really does have such structure, and, moreover, that regularities of experience really can and really do translate into regularities of mental representations (typically assessed via behavioral measures, but also potentially via neuroimaging or other techniques).

1.5 The present study

The present study takes a small step towards addressing the issues raised above. Essentially, I ask: Are the broad implications of current UBTL proposals really valid? Can there be incidental learning of arbitrary patterns of cooccurrence between language and the nonlinguistic context? Specifically, can the structural priming effect be modulated by a context-dependent effect, as the notion of rich memory storage proposed by UBTL researchers would seem to imply?

Put another way, the hypotheses of the present study is essentially an omnibus package of strong hypotheses from each of the components of the UBTL discussed earlier: can *adults* (because there is learning across the lifespan) learn a *cross-modal association* between a *syntactic* structure (because there is no difference between that and learning any other language structure) and a feature of the *nonlinguistic* context (because that context is stored in the exemplar representation) in such a way that we can observe an effect on linguistic production? There is evidence for each individual piece of this hypothesis, as described in the preceding pages, but there is no direct evidence that all of these claims will actually hold at the same time.

To test this, a series of experiments were conducted in which participants were exposed to cooccurrence patterns of language and nonlanguage that were completely reliable but also incidental (i.e., it was not necessary to pay attention to the pattern to complete the task). As outlined above, the UBTL hypothesis in this situation is that the nonlinguistic context should be stored along with the linguistic exemplars, and thus that the pattern of cooccurrence should be able to influence participants' subsequent use of language.

Chapter 2

Methods

The five experiments that were conducted were broadly similar in design. To simplify presentation, this chapter gives an overall description of the methodology. Subsequent chapters report on each individual experiment, and each such chapter will describe how the basic methodology was modified for that particular experiment.

On the most general level, the procedure involved two phases: first, a "prime phase", in which participants were exposed to a combination of language and nonlinguistic context; and second, a "test phase" in which participants were exposed to the nonlinguistic contexts from the prime phase while performing a production task. The nature of nonlinguistic context differed from one experiment to another: it was either the background color of the screen, a sound which played along with each stimulus, background music that played throughout the task, or the room in which the subject performed the task.

2.1 Materials

Stimuli for the experiments consisted of 48 sentences and 24 pictures. The sentences consisted of 24 pairs, each with one active sentence and one passive sentence. The two sentences in each pair were semantically equivalent, differing only in voice. The pictures were black-and-white line drawings used in Bock (1986).

2.2 Participants

All participants were undergraduates recruited from linguistics courses. Most were enrolled in an introductory class and had little or no prior exposure to linguistics. Some were recruited from upper-division classes that did not deal directly with psycholinguistics or syntax (e.g., historical linguistics). Students received extra credit in their course as compensation for their participation. As described below, participants answered a post-test question to assess whether they were aware of the experimental manipulation.

2.3 Procedure

Subjects were greeted either by the experimenter or a research assistant, but interaction between them was limited as all instructions were delivered via the computer interface. The experimenter/research assistant's only role was to provide the participant with the consent form and obtain student ID information so that course credit could be given. Although the experimenter and research assistant were not blind to the hypotheses of the study, subjects were assigned to conditions in a randomized manner by the computer only after the experiment had actually begun. Thus, it was not possible for the person setting up the experiment to know what condition the subject would be assigned to.

After reading and signing a consent form, participants were seated at a computer, in front of which, on the desk, was a microphone on a small stand. An initial instruction screen told participants that they would see a series of sentences displayed, and that they should read each one aloud. The microphone was calibrated so that participants did not need to learn toward it or speak loudly.

Participants then began the prime phase, during which they saw a series of 24 sentences, half active and half passive. Each sentence was chosen from a different active/passive pair; thus, each subject saw one sentence from each pair, and no individual subject saw active and passive versions of any individual sentence. Each sentence appeared in black text inside a white box on a screen by itself. After reading each sentence, participants clicked a "done" button at the bottom of the screen to move to the next sentence. Participants' readings of the sentences were recorded but not analyzed; the purpose of the task was simply to ensure that participants actually read the sentences instead of blindly clicking through them, and the recordings were listened to just to make sure they had done so.

Along with each sentence there was a nonlinguistic "context," which varied from one experiment to another. For each experiment, there were two contexts, and for each participant, one of these two contexts was present for all active sentences, while the other was present for all passive sentences. Thus each participant saw 12 active sentences, each accompanied by a nonlinguistic context, and 12 passive sentences, each accompanied by another nonlinguistic context. I use the terms "active context" and "passive context" to refer to the context shown with active sentences and the context shown with passive sentences, respectively.

After reading all 24 sentences, participants saw a second instruction screen, which told them that they would now see a series of pictures, and that they should speak aloud a one-sentence description of each picture. They then began the test phase, in which they saw 24 stimulus pictures in a row. The pictures were black and white line drawings and displayed in a white box of similar size and position to that used for the text. (The box sizes differed slightly from one picture to another because the pictures had different dimensions.) As with the sentences, participants spoke their description and then clicked a "done" button to proceed to the next picture.

Along with each picture, one of the two contexts was present. For half of the pictures, the active context was present, and for the other half, the passive context was present. Thus each participant saw 24 pictures, 12 of which were accompanied by a context previously encountered when reading active sentences, and the other 12 of which were accompanied by a context previously encountered when reading passive sentences.

The presentation of contexts was varied between subjects in certain ways to guard against order-of-presentation effects. For each phase of the experiment, some subjects saw "blocked" stimuli and others saw "random" stimuli. Blocked stimuli meant that subjects saw all stimuli with one context first, then all stimuli with the other context; random stimuli meant that subjects saw stimuli from the two contexted randomly intermixed. For subjects seeing blocked stimuli, the choice of which stimulus block came first (active or passive) was counterbalanced across subjects. The choice of blocked/random presentation was counterbalanced independently for the two phases — that is, an individual subject might see blocked stimuli in the prime phase but random stimuli in the test phase.

Thus, a "blocked prime" subject saw either 12 active sentences in a row followed by 12 passive sentences in a row, or 12 passive sentences in a row followed by 12 active sentences in a row. A "random prime" subject also saw 12 active sentences and 12 passive sentences, but saw them randomly intermixed. A "blocked test" subject saw either 12 pictures in a row accompanied by the active context followed by 12 pictures in a row accompanied by the passive context, or 12 pictures in a row accompanied by the passive context followed by 12 pictures in a row accompanied by the active context. A "random test" subject also saw 12 pictures accompanied by the active context and 12 pictures accompanied by the passive context, but saw them randomly intermixed.

Regardless of whether stimuli were blocked or random, the order of the individual sentences and pictures was random for each subject, as was the choice of which particular sentences were active and which were passive (i.e., whether the active or passive sentence was chosen from each sentence pair), as well as the choice of which individual pictures were paired with which context. For instance, although all "blocked prime" subjects saw 12 active sentences in a row and 12 passive sentences in a row, they did not all even see the same sentences, let alone in the same order. Some subjects, for instance, saw "The rain was soaking the patio furniture" in the active block, while others instead saw "The patio furniture was getting soaked by the rain" in the passive block. Also, among these "blocked prime" subjects, some saw an active block first and some saw a passive block first. Likewise, in the test phase, all subjects saw 12 pictures accompanied by the active context, but not all subjects saw the same pictures with the active context; rather, a random set of 12 pictures was chosen for each individual subject, and those 12 were shown with the active context, while the other 12 were shown with the passive context.

In short, everything was randomized as much as possible, so that, across subjects, no individual stimulus had any consistent sequential position, nor any reliable pattern of occurrence with either of the two voices or with any other stimulus.

After describing all 24 pictures, subjects were taken to a short questionnaire screen. There, they were asked to provide their age and sex. They also were asked whether they had noticed any pattern in the sentences, pictures, and/or contexts, or in the relationship between the sentences/pictures and the contexts. After completing this questionnaire subjects saw a “Thank you” screen and were done with the experiment.

Chapter 3

Experiment 1a: Color

3.1 Participants

Participants were 42 undergraduates recruited from introductory linguistics classes. They received extra credit in their course as compensation for participating.

3.2 Methods

This experiment followed exactly the pattern outlined above. The nonlinguistic context that was manipulated was the color of the screen: either red or blue. Thus, for a given subject, if the "active context" was blue, then all active sentences were shown on a blue background, while all passive sentences were on a red background; if the active context was red, then it was the other way around.

3.3 Results

Subjects' readings of the prime sentences were listened to only to verify that subjects did in fact perform the reading task (which all subjects did); their readings do not figure in any analyses. Data from one subject, however, were excluded from analysis because many of the subject's responses on the picture-description task were ungrammatical (e.g., "a man getting stinged by a bee"), raising doubt about whether the subject was really a native speaker.

Subjects' responses to the post-test question were inspected to see if they reported any awareness of the link between the colors and the sentences. No subjects did. Interestingly, many subjects claimed to have noticed correlations which did not in fact exist, typically involving a link between the color and the unpleasantness of the depicted or described situation. A typical subject claimed: "The worst scenarios were displayed on a red background and more positve ones on blue backgrounds." The willingness of subjects to reach for dubious semantic relationships while remaining oblivious to the 100%-reliable correlation between color and syntax underscores how unaware the typical language user is of syntactic structure.

The audio recordings of subjects' picture descriptions were coded for which voice they were in. Response were coded as "active", "passive", "agentless passive" or "other". Responses were coded as active if they contained an active transitive clause whose arguments referred to the central participants in the depicted event; they were coded as passive if they contained a full passive (with agent in a *by* phrase) whose arguments referred to the central participants. They were coded as "agentless passive" if they contained a one-argument passive referring to one of the central participants, with the other participant either omitted or expressed in some other fashion (e.g., in an oblique phrase). They were coded as "other" if they met none of the above criteria.

A few particular cases should be noted. First, responses could be coded as passive whether they used *be* or *get* as the auxiliary. Second, agentless passives where the "agent" was expressed, but not in a *by* phrase, were coded as "agentless passive" rather than passive; this included for instance cases where the agent was expressed as an instrument (e.g., *a boy getting hit in the head* ***with*** *a baseball*). Third, the criteria for coding as active/passive were applied by reference to the central participants in the depicted event, not to any minor details of the picture that the subject happened to include in the description. Thus, responses were still coded as active or passive even if the described event was not exactly the "intended" depicted event, as long as it had the same participants. For instance, in describing the picture of a veterinarian being bitten by a cat, some participants used a description like *a vet holding a cat*; this was still counted as transitive because it described the veterinarian and the cat and was potentially passivizable (e.g., *a cat being held by a vet* is possible). Likewise, for the picture showing a soldier being run over by a tank, the coding was done based on how the soldier and tank were expressed, ignoring any references to the soldier's gun; thus a description such as *a soldier getting run over by a tank and dropping his gun* was coded as "passive" because the clause involving the soldier and tank is passive, even though the conjoined clause describing the soldier and his gun happens to be active.

Each of the 41 participants provided 24 responses, for a total of 984 responses. Of these 790 (80.3%) were coded as active or passive and 194 (19.7%) were coded as "agentless passive" or "other". Except where noted, the analyses reported below were conducted only on responses coded as active or passive.

The results were submitted to a logistic regression using the R package lme4 (Bates et al. 2013; R Core Team 2013, version 1.0-5) with ObservedVoice (active or passive) as the dependent variable on each trial. The predictors included in the regression were:

1. ActiveContext — Which color was primed for active voice in the priming phase

2. Context — Which color was shown on the current trial

3. BlockedPrimes — Whether the primes were blocked or random

4. BlockedTest — Whether the test trials were blocked or random

5. StimIndex — The index of the trial within the sequence of trials, i.e., 1 for the first picture, 2 for the second, etc.

6. ActiveProp — The proportion of trials so far for which the subject's response was in the active voice

7. SubjID — A random effect for the identity of the subject

8. PicName — A random effect for the particular picture described on this trial

All interactions of up to three predictors were also included.[1]

The ActiveProp proportion was calculated relative to all active/passive responses by the subject so far, ignoring "other" responses, and was smoothed to avoid giving undue weight to early trials. Thus ActiveProp was calculated as:

$$\frac{1 + \#\text{ of active responses so far}}{2 + \text{total }\#\text{ of responses so far}}$$

The 1 and 2 are the smoothing factors. The smoothing was done because otherwise there would be no ActiveProp for the first trial (since there are no "responses so far" to divide by), and every second trial would have ActiveProp either 1 or 0 (because on every second trial either "all" responses so far — that is, the one response so far — were active or all were passive. The smoothing "hedges" the proportions so that, after N trials, the minimum possible ActiveProp is $\frac{1}{N+1}$ (instead of 0) and the maximum is $\frac{N}{N+1}$ (instead of 1). This reflects the fact that a high or low proportion of active voice is more meaningful after a large number of trials than after a short number.

The ActiveProp variable was included in order to capture a self-priming or persistence effect within each subject. Given earlier research on syntactic priming (e.g. Gries and Wulff 2009; Jaeger and Snider 2008), it was reasonable to expect that subjects who began using a given voice would prime themselves to use the same voice on the next trial, thus potentially leading to a cascade effect in which each subsequent use of a particular voice makes them even more likely to use it yet again. Therefore, ActiveProp was included to measure the extent of such syntactic perserveration within each subject.[2]

[1] The random effects included here were adjustments to intercepts. Models with random-slope predictors for the targeted effect — that is, effects for the Context-ActiveContext interaction, grouped by SubjID and by PicName — were also run for this and all other experiments. These models did not differ significantly from those with random intercepts only, so the final models reported here are those with random intercepts only.

[2] It is important to distinguish between the self-priming represented by ActiveProp and the subject-specific voice preference captured by the random effect of subject. The random effect takes account of an individual subject's *overall* preference for one voice or the other across all trials; it does not take account of how that preference is distributed over the sequence of trials. The ActiveProp does take this temporal distribution into account, because it measures, for each trial, the proportion of active voice used on *preceding* trials only. As

A stepwise model selection process was used, eliminating predictors one at a time until all remaining predictors were at least marginally significant (at the $p < 0.1$ level) or were qualified by a significant higher-order interaction. This rather liberal p-value threshold was chosen simply to allow the model to retain predictors whose influence on the model might be of interpretive interest despite not reaching the traditional $p < 0.05$ significance level, without implying that these predictors were in fact *interpreted* as signficant. In other words, marginally significant predictors were left in the model to avoid discarding any lurking hints of a theoretically relevant effect, but they do not figure in the discussion below except where they suggest that such a hint may actually be lurking, and in such cases they are clearly distinguished as "marginally significant" and not interpreted as being of equal importance with effects which did reach the $p < 0.05$ level.

The resulting model had a marginal R^2 of 0.10 (computed using the method of Nakagawa and Schielzeth (2013), with the R package MuMIn), a classification accuracy of 73.8% (compared with a baseline of 58.5% based only on the overall total proportions of active and passive responses[3]), and a point-biserial response correlation of 0.55[4]. The final model is summarized below[5]:

an example, consider two hypothetical subjects: one who uses active voice on the first 12 trials, then passive voice on the last 12 trials; and another who alternates between active and passive voice on every trial. Both will be assigned the same offset by the random effect, because their overall proportions of each voice are equal (both use active exactly half the time and passive the other half). But their behavior from the point of view of ActiveProp is very different: the first subject has a tendency to keep using whatever voice he has used the most up until that time; the second subject has a tendency *not* to use the voice he has used the most up until that time (i.e., he has a strong tendency to oscillate between voices). Thus the ActiveProp and the random effect of subject capture different information.

[3]That is, 58.5% of the responses were active voice, meaning that a naive classification that always predicts active voice would be right 58.5% of the time.

[4]The point-biserial response correlation is the Pearson correlation coefficient of the predicted probability of passive voice on each trial and a binary response variable set to 0 if active voice was observed on that trial and 1 if passive voice was observed. While the classification accuracy only measures whether the model was "in the right direction" — i.e., whether the observed result had a predicted probability of 50% or higher — the point-biserial correlation also takes account of the model's "confidence" in its predicted outcomes. Given that passive voice was observed on a trial, a model that predicts a 49% probability of passive and a model that predicts a 1% probability of passive both make the wrong prediction; but the second model is "more wrong" because it is very confident about its incorrect prediction, i.e., it gives a very low probability to the event that actually occurred.

[5]All model tables use Type II Wald Chi-squared tests.

	Chisq	Df	Pr(>Chisq)	
ActiveContext	0.9081	1	0.340628	
Context	0.7031	1	0.401754	
BlockedPrimes	3.1170	1	0.077477	.
BlockedTest	1.5512	1	0.212964	
ActiveProp	5.7711	1	0.016292	*
StimIndex	1.0912	1	0.296199	
ActiveContext:Context	0.4971	1	0.480764	
ActiveContext:BlockedPrimes	1.0888	1	0.296742	
ActiveContext:ActiveProp	0.4189	1	0.517496	
Context:BlockedPrimes	0.1412	1	0.707082	
Context:BlockedTest	6.0364	1	0.014013	*
Context:ActiveProp	1.4144	1	0.234324	
Context:StimIndex	0.1015	1	0.750092	
BlockedPrimes:BlockedTest	2.7740	1	0.095807	.
BlockedTest:StimIndex	0.0002	1	0.988462	
ActiveContext:Context:BlockedPrimes	9.5816	1	0.001965	**
ActiveContext:Context:ActiveProp	3.8810	1	0.048834	*
Context:BlockedTest:StimIndex	4.2502	1	0.039245	*

The predicted effect was an interaction between ActiveContext and Context, such that active voice was predicted to be more common for pictures displayed with the "active color" (i.e., the color present with active sentences in the prime phase), and likewise for passive voice. This interaction was not found to be significant. In fact, it was only partially in the right direction: the nature of the nonsignficant interaction was that passive voice was always preferred when the passive color was red, but it was especially preferred when red was also the color seen with the picture. Thus the hypothesis was not supported.

However, there was a significant ($p < 0.01$) three-way interaction among ActiveContext,

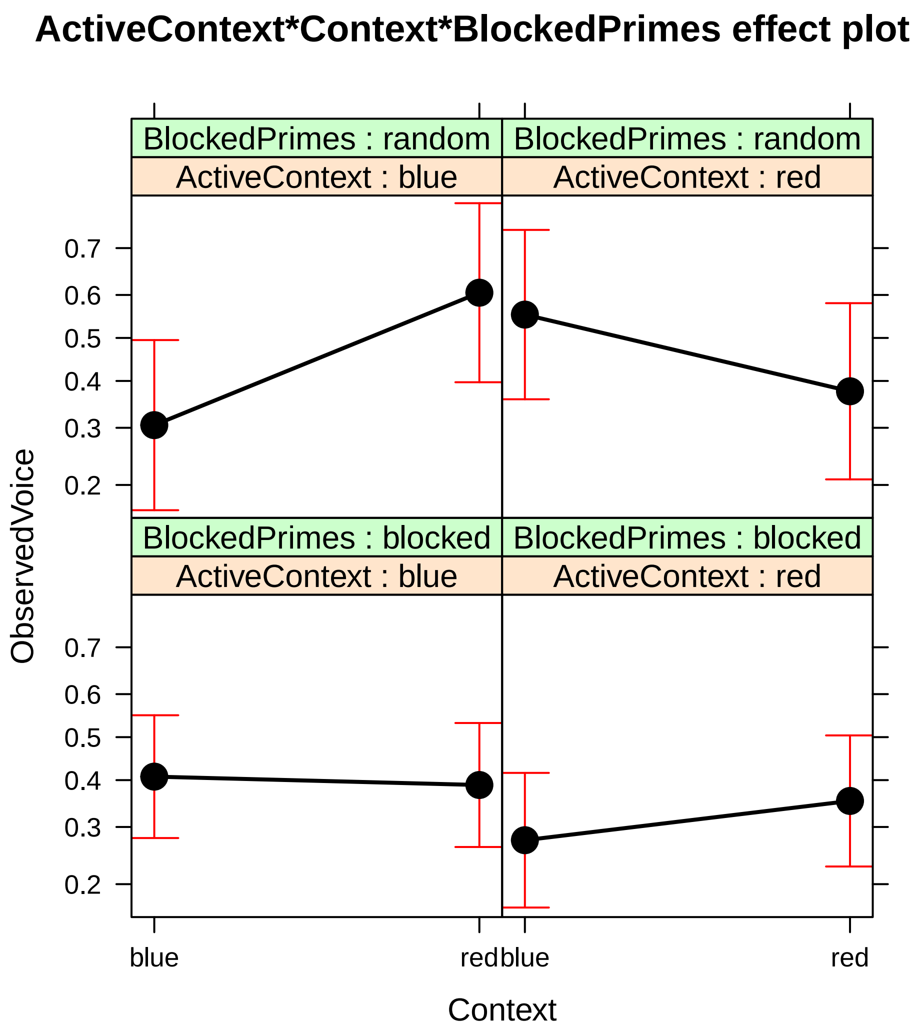

Figure 3.1: Effect plot for interaction of ActiveContext, Context, and BlockedPrimes. The Y axis represents probability of passive voice. The top two panels represent the cases where subjects saw random primes; here the predicted effect is observed, with participants using comparatively more active voice when the active color was shown, and likewise for passive voice. The bottom panels represent cases where subjects saw blocked primes; in these cases there was little difference in voice usage across experimental conditions.

Context and BlockedPrimes. An effect plot for this effect is shown in Figure 3.1.[6] The nature of the effect was that the predicted effect was observed only for participants who saw randomly mixed primes (rather than blocked).

There was also a main effect of ActiveProp; this was a self-priming effect, with participants being more likely to use active voice to describe a picture the more often they had

[6]In this and all other effect plots, the Y-axis values measure the probability of passive voice – the higher a point, the more likely passive voice is. The effect plots were created using the "effects" package for R (Fox 2003).

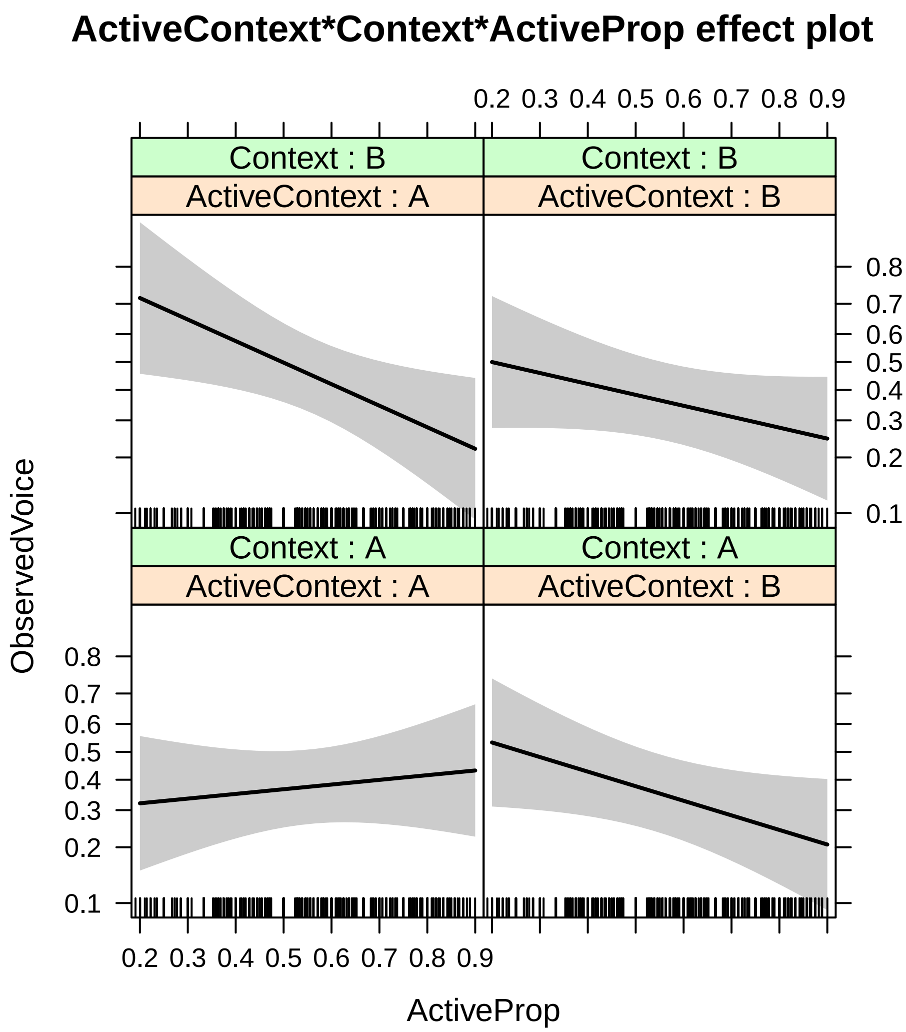

Figure 3.2

used it to describe earlier pictures.

This effect was qualified by a significant ($p < 0.05$) interaction between ActiveContext, Context, and ActiveProp, indicating that the strength of the self-priming tendency varied depending on the relationship between the trial context and the active context. Figure 3.2 summarizes the effect. The overall nature of the effect was that the self-priming tendency was strongest when the passive color was shown, and weaker or even reversed when the active color was shown. This effect thus does not have any clear interpretation with regard to the influence of the contextual prime on subjects' production. One could speculate that the interaction is due to the asymmetric markedness of active and passive voice. That is, it might be that the learned association between the color and the voice did not succeed in priming participants to use each voice with its color, but did succeed in interfering with their production processes to the extent that it disrupted their natural tendency to persist with the use of a single voice. However, there is no obvious reason why such interference would suppress self-priming only with the marked (i.e., passive) voice; if anything, the natural hypothesis would be that the less-frequent passive would be more easily disrupted, whereas in fact the reverse was found (self-priming was *stronger* when the passive color was shown).

Finally, there was a significant interaction of Context and BlockedTest, which was qualified by a signficant three-way interaction between Context, BlockedTest, and StimIndex. The nature of these effects was that for some combinations of color and blocked/random presentation, participants became more likely to use active voice as they progressed through the trials, while for other combinations, they became less likely to use active voice. Since the interaction does not involve ActiveContext, it is unrelated to the priming manipulation. It also has no clear interpretation with regard to the hypotheses. It is unclear why, for instance, participants would be expected to use more and more active voice, but only on screens with a red background — let alone why they would do this only when test stimuli were blocked but do the reverse when test stimuli were randomly interleaved.

All these interactions, however, must be qualified with the statement that a random-effects-only model actually does a slightly better job of accounting for the results. A model was fit on the same data, using only the random effects for SubjID and PicName. This random model had a classification accuracy of 74.4%, slightly higher than the 73.8% of the original model. In addition, the point-biserial correlation between the predicted probabilities and observed outcomes was 0.551, slightly higher than the 0.545 for the original model.

To assess whether the inclusion of agentless passives would affect the results, the same analysis was rerun, but now with all "agentless passive" responses included as passive responses. (For instance, both *a car getting towed* and *a car getting towed by a tow truck* were both considered passive.) There were 61 such responses.

The results were not meaningfully affected. The significant effects described above remained significant, with the exception of the Context-BlockedTest-StimIndex interaction, which, as mentioned, did not carry hypothesis-relevant information anyway. The only new significant effect was an interaction of BlockedPrimes and BlockedTest, indicating that passive voice was more common when primes were blocked but test trials were random, or vice versa; this effect does not involve either Context or ActiveContext and thus also does not bear on the hypothesis. The relationship between the model including agentless passives and its corresponding random-effects-only model was similar to that for the model without agentless passives (CA of 70.7% for the full model vs. 72.2% for the random model; point-biserial correlation .50 for the full model vs. .52 for the random model), and the R^2 value of 0.10 was the same as for the model without agentless passives.

3.4 Discussion

Overall, the results from Experiment 1a do not lend support to the hypothesis suggested by the UBTL, namely that the priming manipulation can create an association between the background color and the voice. There were some three-way interactions that involved the link between the primed voice and the voice used in picture descriptions, but these interactions did not relate in any simple way to the 2x2 crossing of these two variables. That is, participants did apparently behave differently depending on which voice was primed with which color, but the differences in their behavior do not bear any obvious relationship to that priming. The inclusion of agentless passives did not materially impact these conclusions.

In addition, a model using random effects only accounts for the observed data slightly better than a model that takes account of the various predictors that were supposed to explain the variability. This indicates that, although significant effects were observed, the overall pattern of results can just as easily be explained by a combination of per-picture biases (i.e., some pictures are more likely to be described with one voice than the other) and per-subject biases (i.e., some subjects tend to use active voice more than others). This conclusion is also consistent with the low marginal R^2 of the full model (0.10), which indicates that the fixed effects of the model explain little variance beyond that accounted for by the random effects. In other words, there is no real reason to suppose that the experimental manipulation had any effect beyond the intrinsic preferences of individual speakers and the intrinsic linguistic affordances of the stimulus pictures.

The one result that did support the hypothesis was that the predicted effect *was* present for a subgroup of subjects, namely those who saw randomly mixed primes rather than blocked primes. It is therefore possible that the block structure of the priming influences its effectiveness. When the primes were blocked, only one color change occurred during the prime phase (between the two blocks), but if the primes were mixed, the color would change back and forth between red and blue many times. It could be that this made the color changes more salient and thus enhanced the priming effect. In order to test this hypothesis, Experiment 1b was run, duplicating the procedure of Experiment 1a but restricting only to random-prime presentation, as will be described in the next chapter.

Chapter 4

Experiment 1b: Random-prime color

4.1 Participants

Participants were 18 undergraduates, recruited as described for Experiment 1a. None of the participants had participated in Experiment 1a.

4.2 Methods

As discussed in the results section, the outcome of Experiment 1a suggested a followup version. Experiment 1b was the same as Experiment 1a, except that all participants saw random primes (there were no blocked primes). Both Random Test and Blocked Test presentations still existed, however.

4.3 Results

Data from three subjects had to be excluded. One subject noticed the association between the color and the voice. Another subject reported noticing a relationship between the color and the "sentence structure", which essentially meant this subject had noticed the experimental manipulation. A third subject gave bizarre free-association responses to all pictures (e.g., the picture of lightning striking a church was described with *Jesus is back*), rather than describing them as such.

This left 15 subjects. Of the 360 responses from these subjects, 283 (78.6%) were coded as active or passive and the other 77 (21.4%) were coded as agentless passive or “other”. These proportions are similar to those observed in Experiment 1a.

The active/passive responses were submitted to a logistic regression as described in Experiment 1a. The BlockedPrimes predictor was not used, since in this experiment all subjects saw random primes. The resulting model had a marginal R^2 of 0.12 and a classification accuracy of 80.6%, compared with a baseline of 59.0%. The model is summarized below.

	Chisq	Df	Pr(>Chisq)	
ActiveContext	1.8141	1	0.17802	
Context	1.3888	1	0.23860	
ActiveProp	6.3539	1	0.01171	*
StimIndex	0.2463	1	0.61970	
ActiveContext:StimIndex	3.0431	1	0.08108	.
Context:StimIndex	3.9853	1	0.04590	*

Again, the predicted interaction of ActiveContext and Context was not found. This time the effect was essentially totally absent, with virtually no difference in preference for active vs. passive depending on which color had been associated with which voice. Thus the hypothesis was not supported.

The only meaningful significant effect was a main effect of ActiveProp, summarized in Figure 4.1. This was a self-priming effect, such that the more active voice a participant had used at a given point in the test phase, the more likely he or she was to again use active voice.

There was also a significant interaction of Context and StimIndex, as well as a marginally significant interaction of ActiveContext and StimIndex. Given that these effects do not involve the experimental manipulation (i.e., the relationship between Context and ActiveContext), they have no clear interpretation with regard to the hypotheses. Conceptually,

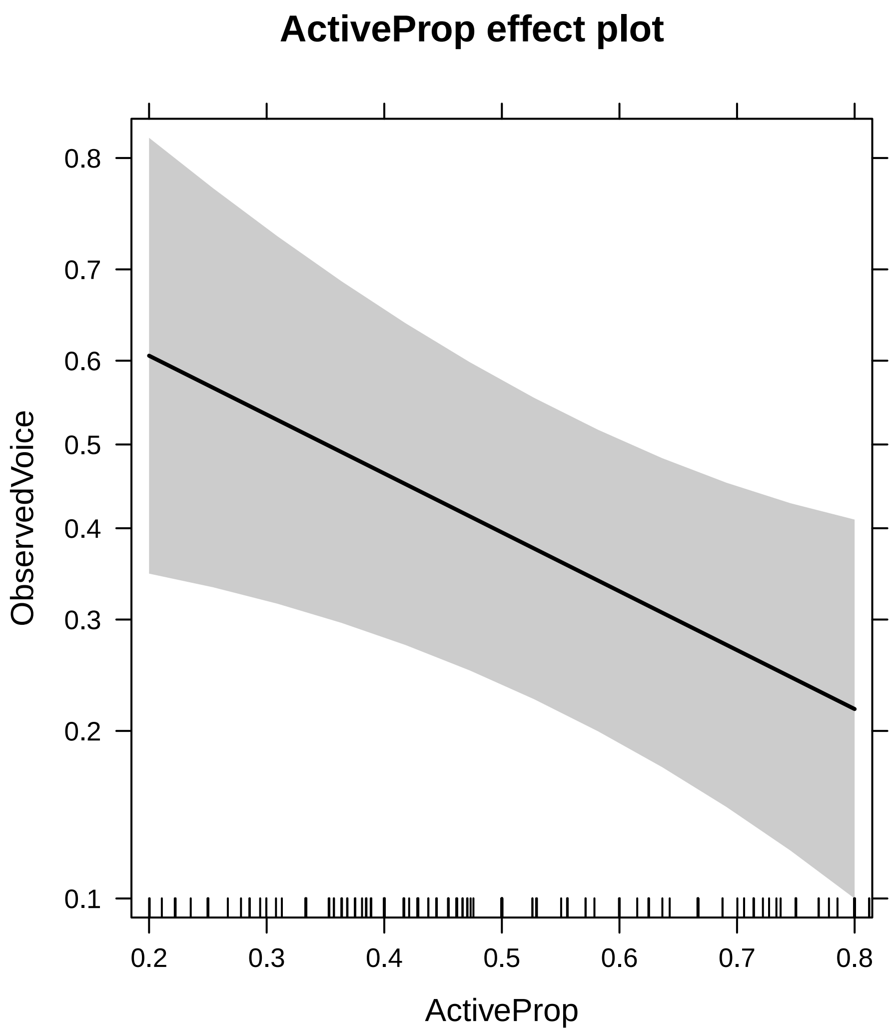

Figure 4.1: Effect plot for the effect of ActiveProp. The greater the proportion of active voice used so far during the test phase, the lower the likelihood of using passive voice.

effects involving Context or ActiveContext but not both indicate that the particular color used had an influence on the choice of voice — that is, the color red had an intrinsic impact on the use of active vs. passive voice, regardless of which voice it was associated with, and that the color blue had a different intrinsic impact. Such effects are difficult to explain since there is no obvious reason why a color in and of itself would induce such effects on language use.

Again, however, a random-effects-only model predicted the results slightly better (CA 83.7% vs. 80.6% for the full model; point-biserial correlation 0.684 vs. 0.677 for the full model).

Again, a version of the model was fit including agentless passives as passives. (There were 24 agentless-passive responses.) The result was effectively the same as the original model. A new significant effect of ActiveContext and StimIndex emerged, such that participants became progressively more likely to use active voice if the active color was red, and less likely to use active voice if the active color was blue; however, this effect did not involve the Context variable, and so did not involve participants responding differently to different colors in the test phase, and hence the effect makes no statement with regard to the hypotheses. Although the model with agentless passives had a higher marginal R^2 (0.22) than the model without agentless passives, its classification accuracy (77.5%) and point-biserial response correlation (0.63) were slightly lower. The agentless-passives model and its corresponding random-effects-only model had essentially equivalent performance (random-effects-only CA was the same at 77.5%; point-biserial slightly lower at 0.61).

4.4 Discussion

Although Experiment 1a suggested that the predicted effect might hold conditionally, when subjects were exposed to primes randomly mixed between the two voice/color pairs, this hypothesis was not borne out by Experiment 1b. In fact, no meaningful effects were found except for a simple self-priming effect. Also, the results can again be explained more parsimoniously by simply allowing for individual differences in voice preference and picture-specific differences in linguistic affordances (i.e., how naturally each depicted scene can be described with active vs. passive voice). Thus, Experiment 1b does not support the hypothesis that speakers can learn incidental relationships between language and nonlanguage.

Chapter 5

Experiment 2: Sound

5.1 Participants

Participants were 41 undergraduates recruited as described previously. None of the subjects had participated in either of the other two experiments.

5.2 Methods

This experiment followed the same paradigm as Experiment 1a, except that the nonlinguistic context that was manipulated was a sound clip that played when each stimulus was shown. The two sounds were a "sad trombone" riff and the sound of tinkling chimes. The two clips were approximately equal in length (about 1 second each). Participants wore headphones when doing this experiment. When each stimulus (either sentence or picture) was displayed, one of the two sounds played over the headphones.

5.3 Results

Data from one participant were excluded because the participant indicated in the post-test question that she "noticed I said things differently active/passive". This was regarded as sufficient awareness of the purpose of the experiment to warrant excluding the data.

Of the 960 responses from the remaining 40 subjects, 748 (77.9%) were coded as active/passive and the other 212 (22.1%) as agentless passive or other.

The active/passive responses were submitted to a logistic regression as described above. The resulting model had a marginal R^2 of .20. a classification accuracy of 75.7% (compared with a baseline of 52.1%), and a point-biserial response correlation of 0.55. The model is summarized below.

	Chisq	Df	Pr(>Chisq)	
ActiveContext	0.0653	1	0.79825	
Context	1.4549	1	0.22774	
BlockedPrimes	0.0429	1	0.83597	
BlockedTest	0.3465	1	0.55611	
ActiveProp	53.9345	1	2.073e-13	***
StimIndex	0.0477	1	0.82720	
ActiveContext:Context	5.8703	1	0.01540	*
ActiveContext:BlockedTest	3.1653	1	0.07522	.
ActiveContext:StimIndex	5.5444	1	0.01854	*
Context:BlockedTest	0.3798	1	0.53773	
Context:StimIndex	0.1575	1	0.69150	
BlockedPrimes:BlockedTest	0.7398	1	0.38971	
BlockedPrimes:ActiveProp	0.4083	1	0.52284	
BlockedPrimes:StimIndex	0.1877	1	0.66485	
BlockedTest:ActiveProp	2.6105	1	0.10616	
BlockedTest:StimIndex	0.4430	1	0.50568	
ActiveProp:StimIndex	4.5194	1	0.03351	*
ActiveContext:Context:BlockedTest	3.5953	1	0.05794	.
ActiveContext:Context:StimIndex	3.2704	1	0.07054	.
Context:BlockedTest:StimIndex	4.5977	1	0.03201	*
BlockedPrimes:BlockedTest:ActiveProp	3.1105	1	0.07779	.
BlockedPrimes:BlockedTest:StimIndex	3.3840	1	0.06583	.

In this case, the interaction between ActiveContext and Context was significant — but the direction of the effect was the opposite of the prediction (see Figure 5.1). When the chime sound was associated with active voice, there was little difference in participants' usage of voice depending on which sound they heard; but when the "sad trombone" sound

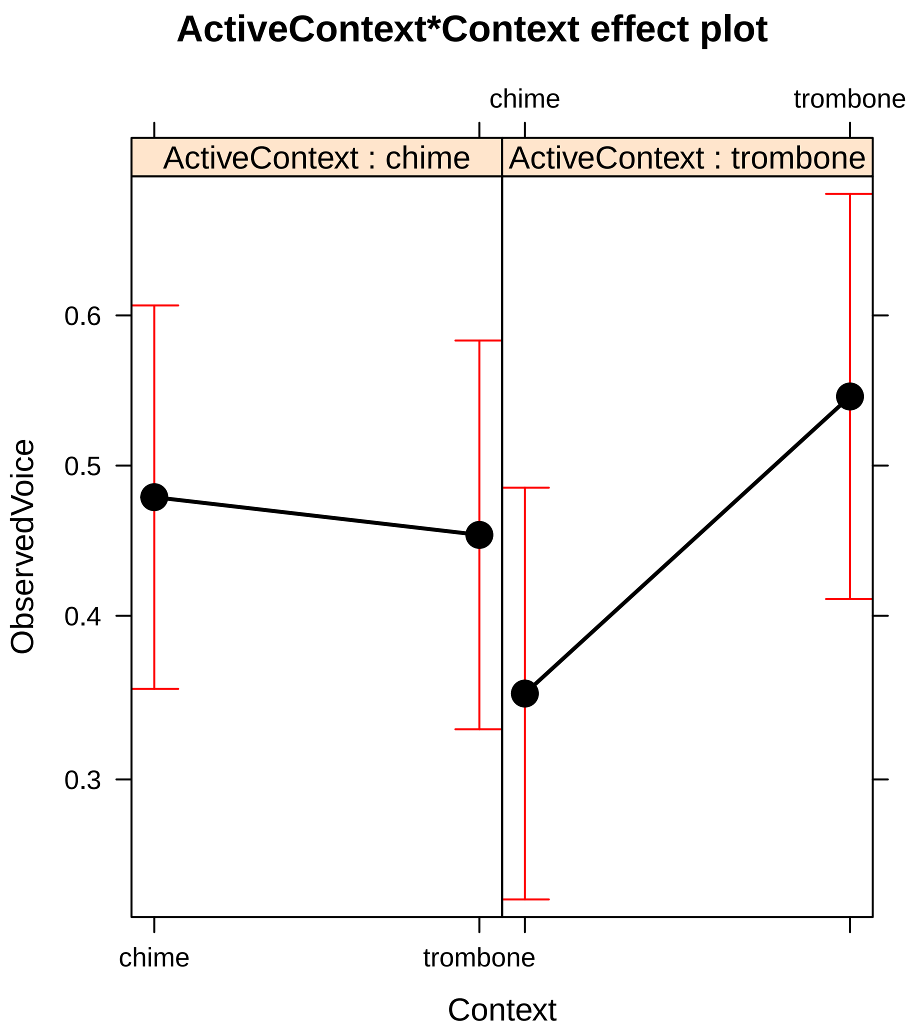

Figure 5.1: Interaction between ActiveContext and Context. The Y axis represents probability of passive voice. The interaction was in the opposite of the predicted direction: when the "sad trombone" sound was associated with active voice (right panel), participants were in fact more likely to use the *passive* voice when hearing that sound.

was associated with active voice (right panel), participants were in fact more likely to use the *passive* voice when hearing that sound. Thus the hypothesis was definitely not supported.

This interaction was qualified by two marginally-significant three-way interactions: ActiveContext-Context-BlockedPrimes and ActiveContext-Context-StimIndex. The first of these is summarized in Figure 5.2; the nature of the effect was that the previously-mentioned interaction of ActiveContext and Context occurred only when participants had blocked tests (not random tests). The second of the three-way interactions, with StimIndex, was such that

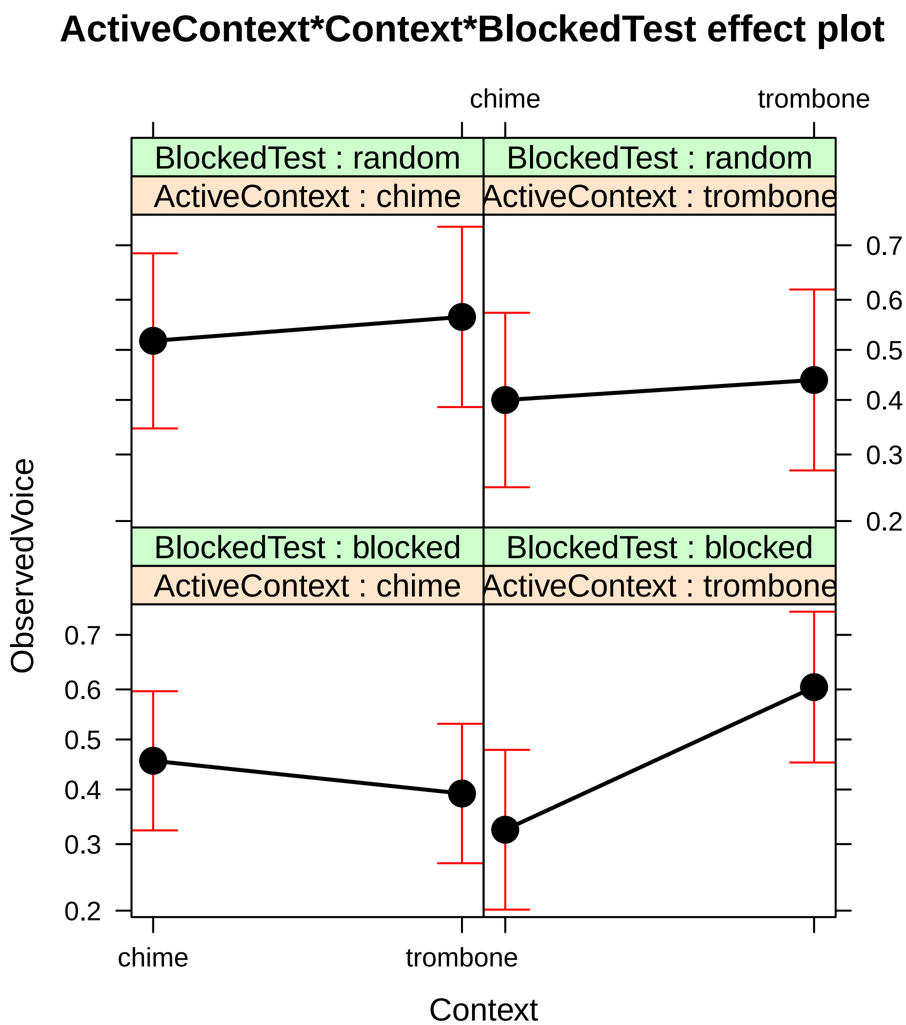

Figure 5.2: Interaction between ActiveContext, Context, and BlockedTest. The reverse-direction interaction of ActiveContext and Context was essentially confined to cases where the participants had blocked tests (bottom panels). When tests were randomly mixed (top panels), there was no major interaction between ActiveContext and Context.

participants' use of active vs. passive voice when the "chime" sound was present tended to change over the course of the experiment, but their use of active vs. passive when the "sad trombone" sound was present did not shift much. The effect is summarized in Figure 5.3.

The previously-seen self-priming main effect of ActiveProp was also significant. Here it was qualified by an interaction between ActiveProp and StimIndex, such that the self-priming effect became stronger as an individual participant progressed through the experiment.

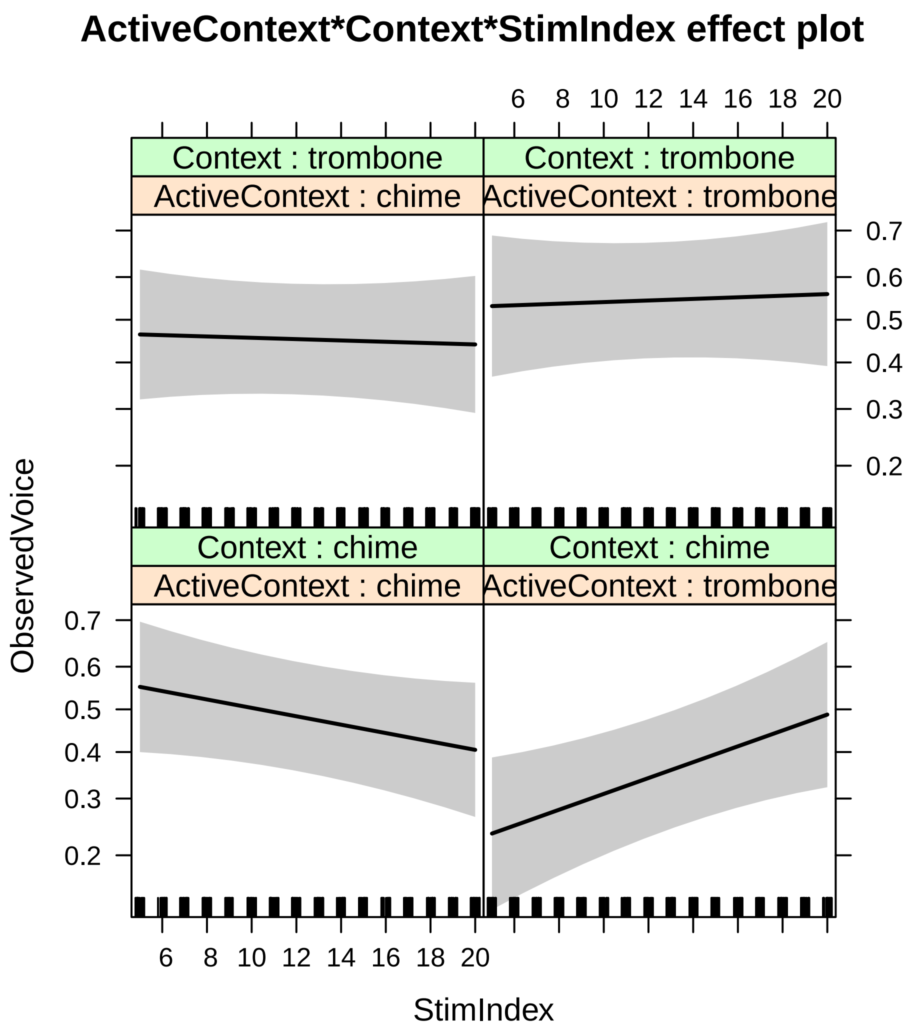

Figure 5.3: Interaction between ActiveContext, Context, and StimIndex. For "chime" trials (i.e., when the chime sound played), participants' use of active vs. passive changed over the course of the experiment, although the direction of the change (more active vs. more passive) varied. For "sad trombone" trials there was little change in participants' use of active vs. passive as the experiment went on.

Finally, there was a significant three-way interaction between Context, BlockedTest, and StimIndex. As with similar interactions discussed before, this effect did not involve the experimental manipulation and thus does not bear on the hypotheses.

Again, a random-effects-only model was fit on the same data, and again this model slightly outperformed the original model (CA 78.5%, point-biserial response correlation 0.62).

The model was again re-fit including agentless passives. The results were substantially unchanged. Some effects that were previously marginally significant now reached the .05 threshold; of these, the only one involving the targeted interaction was the ActiveContext-Context-BlockedTest interaction discussed above. The other effects that reached signifance only with the addition of agentless passives were all extraneous effects not involving the targeted interaction (ActiveContext-BlockedTest-StimIndex, BlockedPrimes-BlockedTest-ActiveProp, and BlockedTest-ActiveProp-StimIndex). The random-effects-only model with agentless passives outperformed its corresponding mixed-effects model by about the same margins as for the models without agentless passives.

5.4 Discussion

In Experiment 2 the results again did not support the hypothesis that people can be primed by an incidental association between language and nonlinguistic context. In this case, there was a significant effect of the sound-voice association, but the effect was opposite to the predicted direction: if participants heard a "sad trombone" sound while reading active-voice sentences, they were in fact more likely to use *passive* voice if the same sound played when they saw a picture they had to describe. Even this effect, however, was again qualified by the finding that inter-subject and inter-picture variability in voice preference describes the results slightly better than do the purported contextual effects.

Chapter 6

Experiment 3: Music

6.1 Participants

Participants were 59 undergraduates recruited as described earlier, none of whom had participated in any of the other experiments.

6.2 Methods

In this experiment, the nonlinguistic context that was manipulated was music. There were two music clips: Mozart's Piano Concerto No. 24 in C Minor, and "People Make the World Go Around," a jazz piece from Milt Jackson's album *Sunflower*. These pieces were chosen because they were previously used by S. M. Smith (1985) to demonstrate that music can produce context-dependent memory effects. Both pieces are entirely instrumental, so there was no possibility of priming from sung lyrics.

For this experiment, there were no random-order stimuli, in either prime or test phase. The reason for this was that the musical context inherently involved a certain contintuity, and random presentation would break this continuity. For random presentation, the music would have to switch back and forth from one piece to another, with a single piece perhaps playing for no longer than the five seconds or so that it took the participant to read the sentence or describe the picture. In such a situation, it would be misleading to say that different "music" stimuli had been presented to the subjects, since isolated five-second clips of music are closer to simple sound clips. Thus, in order to ensure that participants experienced overall "background music" as the context, all participants saw blocked stimuli.

Experiments 1a, 1b and 2 had all found a strong self-priming effect. In addition, in subjectively listening to the subjects' recorded responses, it seemed that some subjects would "get into a rut", giving many answers in a row not only with the same voice, but using similar prosody, with virtually no temporal gap between responses, and with a certain "rhythm" between their verbalizations and button clicks (e.g., always clicking the "Done" button at the instant they finished saying the last word). Although not all subjects exhibited this behavior, it suggested that the self-priming effect observed for voice might be part of a larger-scale automatization in which subjects began to progress mechanically through succeeding trials. Although it was clear that subjects were attending to the task (because their picture descriptions were sensible), it was possible that their attention became somewhat lax, which might reduce the effect of the contextual manipulation.

Accordingly, in an attempt to weaken this perseverative effect, and thus potentially leave more room for the targeted interaction to take effect, this experiment introduced an additional filler task between test trials. After describing each picture, subjects were presented with a screen showing nine checkboxes randomly scattered around the screen. Four of the checkboxes were already checked, and the subjects had to click on the other five to check them in order to proceed to the next picture. (The instructions at the beginning of the experiment were updated to tell subjects what to do on this task.) This served to delay subjects between trials. It also disrupted any developing "rhythm" in their process. Although the pictures and the "Done" button were always in the same screen position from one trial to the next, the checkboxes were randomly repositioned each time, meaning that subjects could not simply leave their mouse cursor hovering over the "Done" button, or mechanically look to the same area of the screen every time, but had to actively engage with the display.

Aside from these changes, Experiment 3 followed the same procedure described above.

6.3 Results

Data from three subjects were excluded. One subject's post-test response indicated that she noticed a change in "tense" associated with the music; while the grammatical change between active and passive is not technically one of tense, it was obvious that this participant had become aware of the link between language and nonlinguistic context. One subject was excluded because she did not do the reading task during the prime phase, raising the possibility that she had not really read the sentences. Finally, one subject was excluded because the majority of her responses were not descriptions of the actual scenes but associative riffs or evaluations (e.g., *this woman shouldn't have been a nurse* for a picture of a nurse about to be hit by a truck).

Of the 1344 responses given by the remaining 56 subjects, 1025 (76.3%) were active/passive and the other 319 (23.7%) were agentless passive or "other".

The active/passive responses were submitted to a logistic regression as described for earlier experiments. The BlockedPrimes and BlockedTest predictors were not used, since as described above all participants saw blocked primes and blocked test trials in this experiment. The resulting model had a marginal R^2 of 0.06, a classification accuracy of 74.8% (compared with a baseline of 57.8%) and a point-biserial response correlation of 0.55. The model is summarized below:

	Chisq	Df	Pr(>Chisq)	
ActiveContext	0.2647	1	0.60690	
Context	1.2312	1	0.26717	
ActiveProp	25.0970	1	5.452e-07	***
ActiveContext:Context	0.3962	1	0.52904	
ActiveContext:ActiveProp	0.2126	1	0.64471	
Context:ActiveProp	0.4088	1	0.52258	
ActiveContext:Context:ActiveProp	3.2540	1	0.07125	.

The only effect significant at the 0.05 level was a main effect of ActiveProp; this was the self-priming effect discussed previously.

There was also a marginally significant ($p = 0.07$) interaction of ActiveContext, Context, and ActiveProp, summarizd in Figure 6.1. The nature of the effect was that the self-priming tendency was stronger when subjects were hearing the "active music" (i.e., the music they had heard while reading active-voice sentences) and weaker when they were hearing "passive music".

Again, however, a random-effects-only model fit on the same data slightly outperformed the full model (CA 75.8%, point-biserial correlation 0.59).

Again, another model was fit including the agentless passive responses (of which there were 85) as passives. This model was essentially identical to the original model, with no additional significant effects. A random-effects-only model including agentless passives likewise outperformed the corresponding mixed-effects model.

6.4 Discussion

The hypothesis of incidental learning of the music-voice association was not supported here. Again the only significant effect found was a self-priming effect.

A marginally significant ActiveContext-Context-ActiveProp interaction was found. Recall that a similar effect was found in Experiment 1a. These effects indicate that the strength

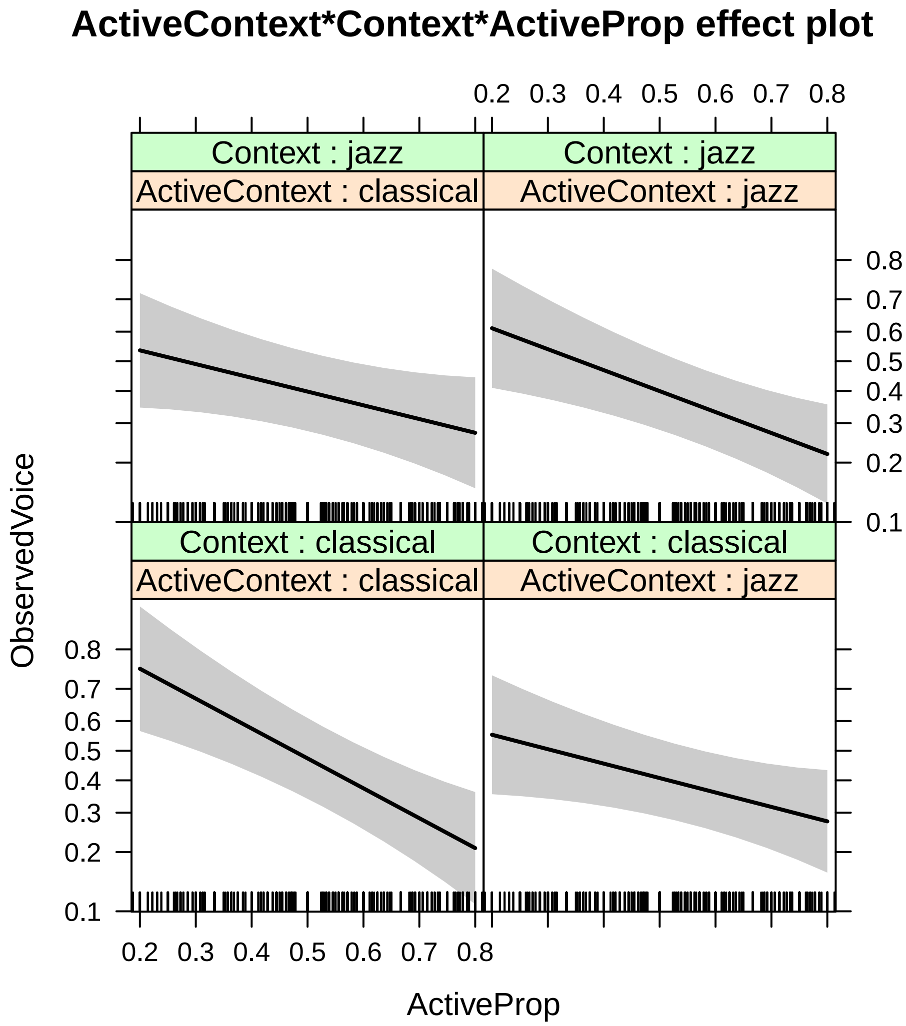

Figure 6.1: Interaction of ActiveContext, Context, and ActiveProp. The slopes of the lines indicate the strength of the self-priming effect across the different conditions. The self-priming effect was strongest when subjects were hearing the "active music" (top right and bottom left panels), and weaker when subjects were hearing the "passive music" (top left and bottom right panels).

of the self-priming effect varied depending on whether subjects were in a context (i.e., seeing a color or hearing music) previously associated with active or passive voice. However, the directions of the effects were different in Experiments 1a and 3. In Experiment 1a, the self-priming effect was stronger when subjects saw the "passive color" and weaker when they saw the "active color"; in Experiment 3, the self-priming effect was stronger when subjects heard the "active music" and weaker when they heard the "passive music".

These contradictory findings reduce the plausibility of an explanation of these effects in terms of the asymmetrical markedness of active and passive voice. If it were true that learning an association between passive voice and nonlanguage was somehow not the same as learning an association between active voice and nonlanguage, due to the greater markedness and lower frequency of passive voice, then it should be the case that the asymmetry is consistent across different nonlinguistic context types. It is difficult to see why an association between active voice and color would enhance the base structural priming tendency, while an association between active voice and music would inhibit it.

Chapter 7

Aggregate analysis of Experiments 1a-3

As noted earlier, the four experiments just described all used the same paradigm, in which participants were exposed to a language-nonlanguage association in a prime phase and then, in a test phase, described pictures accompanied by the nonlinguistic contexts from the prime phase. Because the paradigms were exactly parallel across all experiments, it is possible to consider their aggregate data together. In effect, this is like regarding all of them as different conditions in one large experiment. Notably, this aggregate experiment has a much larger sample size, increasing its statistical power and reducing the likelihood that a genuine effect would be missed.

7.1 Results

Accordingly, a logistic regression was fit as described above, using all of the active/passive responses data analyzed in all of the experiments. This data consisted of 2846 responses from 152 subjects. The same predictors were used as in the models above. There were of course cases where one experiment did not vary some of the predictors (e.g., Blocked-Primes was always "blocked" for Experiment 3), meaning that the predictors were not all as nearly balanced across the entire dataset as they were for the individual experiments, but balanced data is not necessary for a mixed-effects model. The values of the ActiveContext and Context variables were converted to a common format, such that the color blue, the "chime" sound, and the classical music were coded as "Context A" and the color red, the "sad trombone" sound, and the jazz music were coded as "Context B".

The resulting model had a marginal R^2 of 0.05, a classification accuracy of 75.1% (against a baseline of 56.6%), and a point-biserial response correlation of 0.55. The final model is summarized below.

	Chisq	Df	Pr(>Chisq)	
ActiveContext	2.2375	1	0.13470	
Context	0.2976	1	0.58540	
BlockedPrimes	0.8766	1	0.34912	
BlockedTest	0.0820	1	0.77457	
ActiveProp	51.3259	1	7.824e-13	***
StimIndex	0.1803	1	0.67115	
ActiveContext:Context	0.4155	1	0.51917	
ActiveContext:BlockedPrimes	0.0321	1	0.85784	
ActiveContext:BlockedTest	3.2910	1	0.06966	.
Context:BlockedPrimes	0.2588	1	0.61095	
BlockedPrimes:BlockedTest	0.8126	1	0.36735	
BlockedPrimes:StimIndex	0.0813	1	0.77555	
BlockedTest:StimIndex	0.0285	1	0.86587	
ActiveContext:Context:BlockedPrimes	2.8877	1	0.08926	.
BlockedPrimes:BlockedTest:StimIndex	3.1602	1	0.07546	.

The only effect significant at the 0.05 level was the self-priming effect of ActiveProp. There was a marginally significant ($p = 0.08$) three-way interaction between ActiveContext, Context, and BlockedPrimes. As shown in Figure 7.1, this effect did not represent an occurrence of the predicted ActiveContext-Context interaction in some subgroups of participants. Although the likelihood of active vs. passive varied from one condition to another, it did not have any consistent relationship with the nonlinguistic contexts that

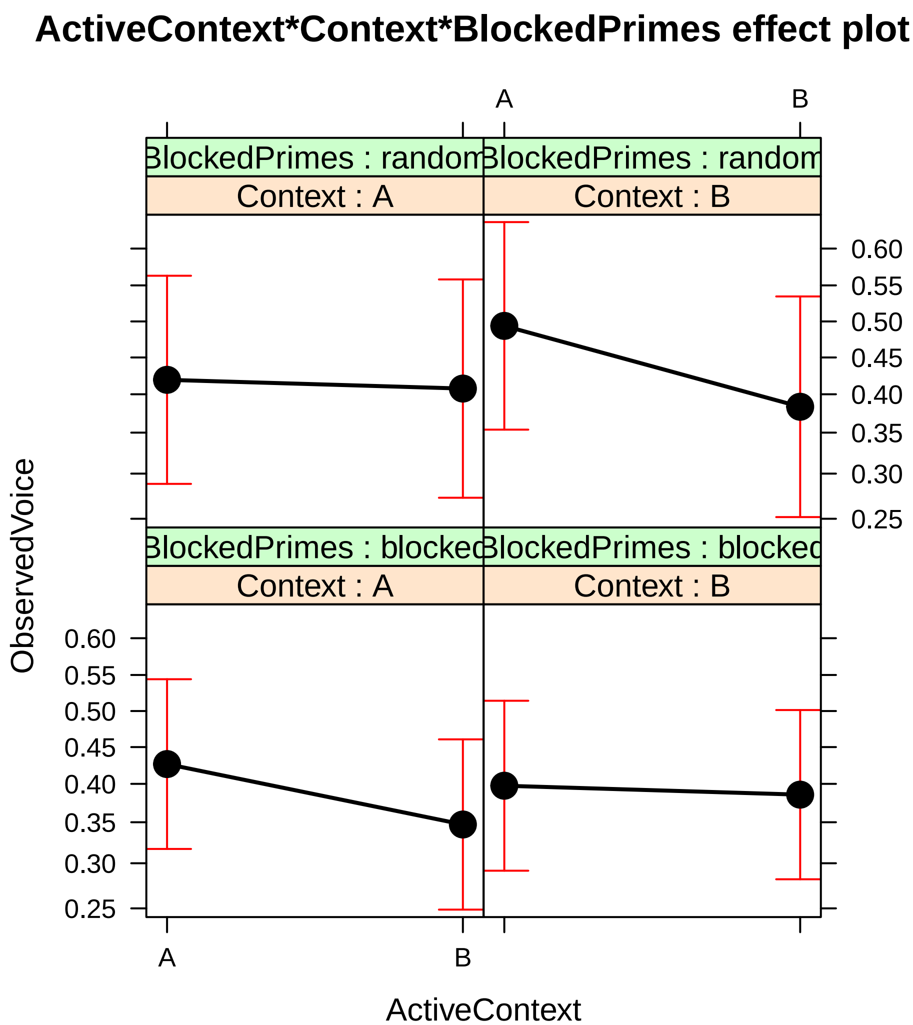

Figure 7.1: Interaction between ActiveContext, Context, and BlockedPrimes. Although there were different patterns of active/passive response in different conditions, the results do not show any clear interaction between ActiveContext and Context.

were associated with the voices.

There were also marginally significant interactions of BlockedPrimes-BlockedTest-StimIndex and ActiveContext-BlockedTest. Because these interactions do not involve the experimental manipulation, they do not make any statement about participants' ability to learn the incidental association between context and voice.

A random-effects-only model was again fit, and again outperformed the model including fixed effects, with a classification accuracy of 76.5% and a point-biserial correlation of 0.58.

An additional mixed-effects model was fit including the agentless passives (of which there were a total of 244) as passives. In this case, the resulting model showed some differences from the model without agentless passives, but the differences were not germane to the hypothesis. Specifically, the agentless-passive model the three marginally-signficant interactions mentioned above became wholly nonsignificant ($p > 0.1$). A new interaction of ActiveContext, ActiveProp, and StimIndex was significant ($p < 0.05$) and an interaction of Context, BlockedPrimes, and ActiveProp was marginally significant ($p = 0.7$). Neither of these interactions involved the targeted ActiveContext-Context relationship, so again they have no bearing on whether subjects' responses were sensitive to the primed associations.

A random-effects-only model including agentless passives outperformed its corresponding mixed-effects model by a similar margin as for the models without agentless passives.

In addition, a similar aggregate model was fit including an Experiment variable, which took on a different value for each experiment. This allowed for the possibility of discovering different patterns of effects across experiments which might not be visible in the separate results of each individual experiment. The results of this model, however, were essentially the same as those of the model without the Experiment variable, with no hypothesis-relevant differences. The only significant predictor involving the Experiment variable was a three-way interaction of Experiment-Context-ActiveContext. Although this interaction involved the targeted predictors, it did not convey meaningful information beyond the results of the individual experiments: the nature of the interaction was simply that the Context-ActiveContext interaction went in different directions across experiments — without being significant in any of the individual experiments except Experiment 3, where its direction was opposite to the prediction. This is what was already reported, and is consistent with the interpretation given for the individual experiments, namely that the contextual manipulations had little or no systematic impact on participant's choice of voice.

7.2 Discussion

Considering the four earlier experiments collectively increased the sample size to 152 participants. This potentially increases the chance of detecting an effect that was too small to reach significance in any of the individual experiments. Instead, however, what was found was that the various interaction effects observed in the earlier experiments vanished when aggregated with additional data, leaving essentially nothing but the self-priming effect.

For instance, although in Experiment 1a it was found that the predicted context-dependence existed only for subjects who saw blocked primes, no such effect was found with the aggregated data. In Experiments 1a and 3, interactions were found between ActiveContext, Context, and ActiveProp. Although the effects were in opposite directions in the two experiments, they raised the possibility that the experimental manipulation might affect production only indirectly, by modulating the otherwise dominant self-priming effect. In the aggregate data, however, this interaction also disappeared. This suggests that, if those additional interactions represent genuine effects, they are very small.

Chapter 8

Experiment 4: Room

8.1 Participants

Subjects were 55 undergraduates, recruited as described for earlier experiments. None had participated in any of the other experiments.

8.2 Methods

In this experiment, the nonlinguistic context that was manipulated was the room itself: participants performed the test phase of the experiment either in the same room where they had performed the prime phase, or in a different room.

The procedure had to be altered somewhat for this experiment. The base paradigm described above required multiple shifts from the Active Context to and from the Passive Context. However, if changing contexts required the participant to physically move to another room, this could require the participant to move back and forth between two rooms up to three times during the experiment. This would likely tip participants off that the change of room had some significance; it would also introduce logistical problems in syncing a single participant's data across multiple sessions spread over two computers in two rooms.

Therefore, in Experiment 4, the paradigm was altered so that the manipulation was done between subjects instead of within subjects. In the prime phase, each subject saw sentences in only one voice (either active or passive).

There were two rooms, the Dark Room and the Light Room. All participants began the experiment in the Dark Room and performed the prime phase in that room. After each participant had completed the prime phase, the experimenter led the participant into the Light Room to wait for a moment. The experimenter then checked the computer in the Dark Room for an indication of whether this participant was assigned to a Same Room or Different Room condition.[1] If the participant was in the Same Room condition, the experimenter set up the test phase on the computer in the Dark Room and brought the participant back into the Dark Room to perform the test phase. If the participant was in the Different Room condition, the experimenter set up the test phase on the computer in the Light Room and the participant remained in the Light Room for the test phase.

Thus, all participants performed the prime phase in the Dark Room, and all participants then waited for a moment in the Light Room, but half of the participants went back into the Dark Room for the test phase, while the other half stayed in the Light Room for the test phase. Half of all participants saw active-voice sentences in the prime phase, while the other half saw passive-voice sentences. Thus, the design created a 2x2 crossing of Prime Voice (active or passive) with Test Room (same or different).

As a result of these change, the nature of the hypothesis was somewhat different for Experiment 4. Because each participant saw sentences in only one voice, some degree of syntactic priming was to be expected for all subjects. The hypothesis was that subjects in the Same Room condition would show a stronger priming effect (i.e., be more likely to use the voice they had been primed with) than subjects in the Different Room condition.

The two rooms were adjacent and similar in size, but were set up to have different atmospheres. In the Dark Room, the blinds were closed, the lights were turned off, and the desk faced away from the window. (The light switch in the room was taped over so

[1]For this experiment, the software was modified to include a "secret button", not visible to the participants, which enabled the experimenter to access this condition information. However, the experimenter remained blind to whether the subject had been primed with active or passive voice.

that participants could not turn the lights on themselves.) The computer screen was lit, and some light filtered through the blinds and through the door (which had a grating of wooden slats), so the room was far from pitch black, but it was noticeably dark. In this room, participants sat on a hard plastic classroom chair. In general, the two room were set up so as to differentiate, as much as possible, the overall sensory experiences of being in each of them.

In the Light Room, on the other hand, the desk faced the window, which had the blinds raised. The window had a view of the sky and a courtyard outside. The overhead light in this room was kept on, so that it was quite bright. In addition, participants sat in a soft, cushioned conference chair.

8.3 Results

Data from 9 subjects were excluded. Six subjects noticed that all the sentences they saw were in the same voice. Two subjects indicated a suspicion that the experiment was attempting to influence the phrasing of their picture description. Finally, no audio was recorded for one subject, apparently due to a problem with the microphone.

There was thus a much higher rate of reported awareness of linguistic form in this experiment, relative to the others (eight in this one experiment versus a total of four in all the other experiments). This is probably due to the between-subjects design, which exposed each participant to only one voice. In the earlier experiments, although some subjects were exposed to blocks of a single voice, all eventually saw equal numbers of sentences in both active and passive voice. Since use of both voices is unremarkable in discourse, this was not likely to attract attention. In this experiment, though, participants were only exposed to one voice, and the mere fact of seeing only one voice might be noticeable. Interestingly, only one subject noticed when all sentences were in active voice, while five noticed when all were in passive voice, perhaps due to the greater markedness of passive voice. This interpretation is bolstered by the fact that, of the eight subjects who reported some awareness of voice (or "sentence structure" or the like), seven had been exposed to passive voice and only one to active voice. Since the passive voice is regarded as the more marked, it makes sense that subjects would be more likely to notice when the sentences were all passive than when they were all active.

Of the 1104 responses from the remaining 46 subjects, 848 (76.8%) were active/passive and the other 256 (23.2%) were agentless passive or "other".

The active/passive responses were submitted to a logistic regression. Since the experimental design for this experiment differed somewhat from the others, some of the predictors were different. There were no Context, ActiveContext, BlockedPrime, or BlockedTest predictors; since each participant saw only one voice and performed the test phase in only one room, there was no distinction between different contexts within subjects, and all stimuli were effectively "blocked".

Instead, the two predictors used were PrimeVoice and SameRoom. PrimeVoice indicated the voice (active or passive) with which a participant had been primed during the priming phase. SameRoom indicated whether the participant did the test phase in the same room as the prime phase (i.e., the Dark Room) or in a different room (i.e., the Light Room). In addition to these two, the predictors ActiveProp and StimIndex, as well as the random effects for SubjID and PicName, were included as described for earlier experiments.

In this case, the prediction was that participants in the same-room condition would be more likely to use the voice they had been primed with than would participants in the different-room condition. That is, an interaction between SameRoom and PrimeVoice was predicted. Note that there was no crucial prediction about the main effect of PrimeVoice. Clearly, one would expect that participants primed with active voice would be more likely to use active voice than participants primed with passive voice; however, the crucial prediction in this case was that such an effect should be enhanced if the participants additionally remained in the same room.

The resulting model had a marginal R^2 of 0.15, a classification accuracy of 73.5% (compared with a baseline of 52.6%), and a point-biserial response correlation of 0.54. The

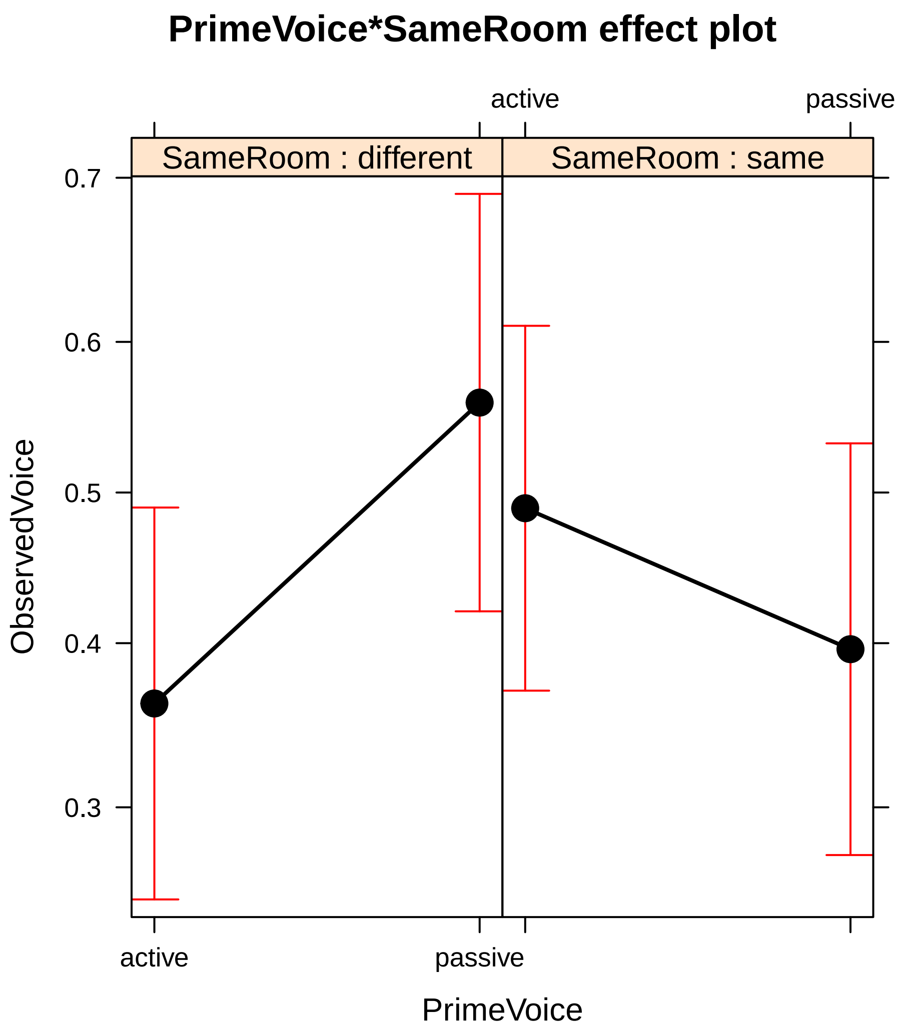

Figure 8.1: Interaction between PrimeVoice and SameRoom. When participants remained in the same room, they were more likely to use the voice they had *not* been primed with. When they went to a different room, they were more likely to use the voice they *had* been primed with.

model is summarized below:

	Chisq	Df	Pr(>Chisq)	
PrimeVoice	0.7511	1	0.3861248	
SameRoom	0.0048	1	0.9450003	
ActiveProp	35.0642	1	3.19e-09	***
StimIndex	2.2147	1	0.1366985	
PrimeVoice:SameRoom	7.5433	1	0.0060234	**
PrimeVoice:StimIndex	11.1204	1	0.0008539	***

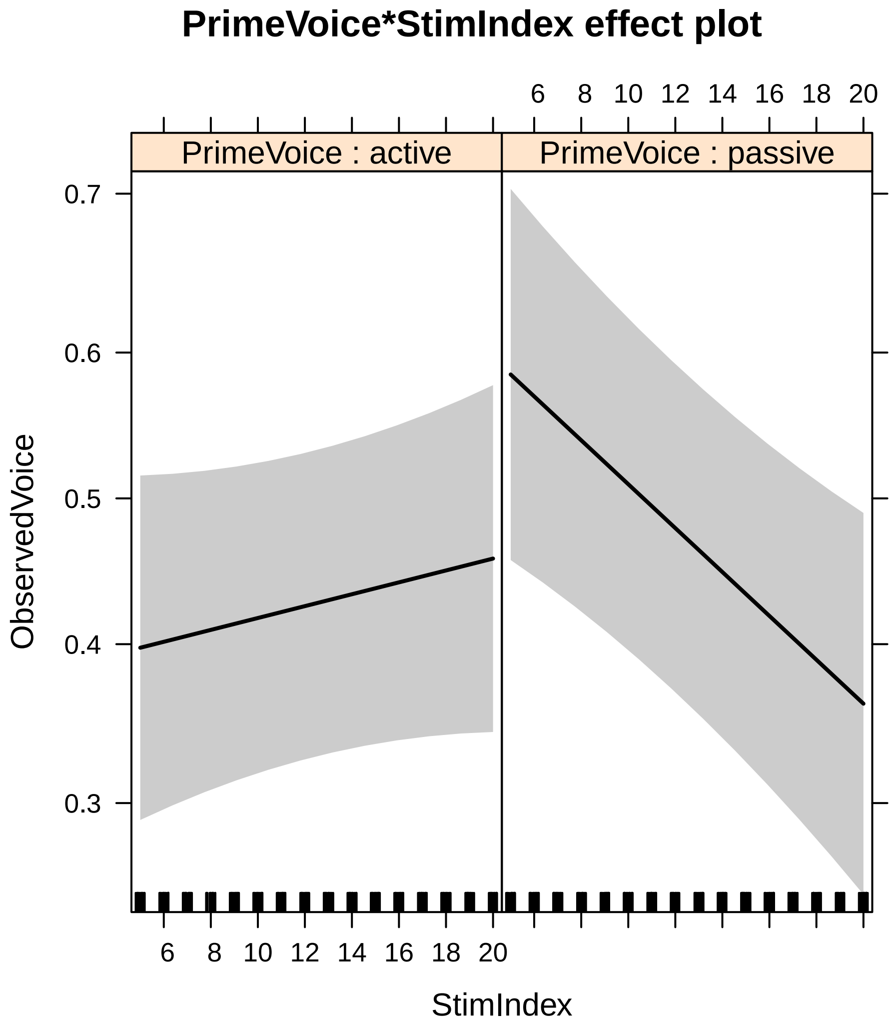

Figure 8.2: Interaction between PrimeVoice and StimIndex. The self-priming effect was strong for participants primed with passive voice, and weakly reversed for participants primed with active voice.

The predicted interaction of PrimeVoice and SameRoom was significant. However, as shown in Figure 8.1 the direction of the effect was a complete reversal of the prediction: when they remained in the same room for prime and test, participants tended to use the voice they had *not* been primed with, whereas when they went to a different room, they tended use the voice they *had* been primed with. This effect is all the more puzzling because participants were exposed to only one voice. This means that, for instance, participants, when sitting in the original room to describe pictures, actually used more passive voice than average only if they had not been exposed to passive voice at all.

There was also a significant interaction between PrimeVoice and StimIndex, shown in Figure 8.2. Recall that the Y axis in these plots represents the likelihood of using passive voice. The ascending line in the left panel, therefore, indicates that participants primed with active voice had a slight tendency to use more and more *passive* voice over the course of the picture-description task; conversely, participants primed with passive voice had a stronger tendency to use more and more *active* voice as the experiment went on. This can be interpreted as a decaying effect of the prime voice: participants primed with active voice used more active voice at the beginning of the picture description task, when the prime was "fresh", but gradually reverted to using more passive voice, and vice versa for those primed with passive voice. The asymmetry of the slopes again may reflect an asymmetry in markedness between active and passive: participants drifted back to the use of the unmarked active voice more quickly than they drifted back to the use of the marked passive.

Finally, there was a strong main effect of ActiveProp, again indicating a self-priming effect.

A random-effects-only model was again fit on the same data, and, as in the other experiments, it slightly outperformed the model with fixed effects (classification accuracy 75.4% vs. 73.5%; point-biserial response correlation .58 vs. .54).

The analysis was redone including agentless passive responses (of which there were 64) as passives. The resulting model was not meaningfully different from the model without agentless passives.

8.4 Discussion

The hypothesis of a context-dependent priming effect was not supported. There was a context-dependent effect, but the effect was opposite to the prediction, such that participants for whom the prime context was reinstated (i.e., who were in the same room for prime and test phases) preferentially used the voice they had *not* been primed with. This effect is difficult to explain in terms of mechanisms based on cooccurrence patterns. It is, for instance, difficult to reconcile this result with the claim that "repeated inferences made in context become part of the meaning of a word" (Bybee 2010: 55). Even if one were to argue that incidental information, such as the ambiance of the surrounding room, is not likely to be the basis of such inference and would not influence later use, it is hard to see why such information should produce the opposite result, whereby participants in effect "inferred" that a particular syntactic structure was *not* to be used just in the very context where they had seen it used.

Chapter 9

Power analysis

One potential criticism of the experimental results reported here is that the sample size was not large enough, and/or the design not sensitive enough, to detect any effect of plausible size. In order to assess this, a Monte Carlo-based power analysis was conducted: a series of simulations were run in which simplified versions of the logistic regression models were fit on data that was computer-generated to fit the overall profile of the actual data. This analysis essentially simulated results of experiments using the same paradigm as Experiments 1a-3.

Specifically, for each simulation, 40 fake "subjects" were generated, each of which gave responses of either "active", "passive" or "other" according to certain probabilistic parameters. (Note that all experiments reported in the preceding chapters except Experiment 1b had a sample size at least this large.) Each "subject" had a baseline propensity to respond with "active" rather than passive voice, with this propensity chosen from a normal distribution whose mean and standard deviation matched those of the distribution of active/passive ratios of the subjects in the actual data (aggregated over Experiments 1a-3). In addition, to represent each picture's baseline "affordance" for description with each voice, a baseline active/passive probability was assigned to each picture stimulus, equal to the overall ratio of active and passive responses given for that picture (again, aggregated over all experiments).

For each "subject", 24 responses were then generated probabilistically as follows:

1. One response was generated per subject, per stimulus picture (as in the actual data).

2. Each subject had a 22% chance of giving an "other" response (neither active nor passive) on each trial. (This was the approximate overall proportion of "other" responses in the data.)

3. If the subject did not give an "other" response, its probability of responding with "active" was the average of the subject-specific baseline and the stimulus-specific baseline, except that. . .

4. In order to mimic the self-priming effect observed with actual subjects, this probability was adjusted in a manner proportionate to the value of ActiveProp for that trial — that is, simulated "subjects" who had given more active responses so far were more likely to give another active response, and those who had given more passive responses were more likely to again give a passive response.[1]

Each response had an associated Context and ActiveContext as in the real experiments[2], but these were not used in generating the initial response. Thus, these generated responses were sampled randomly from the space of possible responses, according to general parameters that approximately matched the distribution of the observed data.

Then, in a separate "effect-generation" step, to simulate an "effect size" of the targeted context-dependence effect, individual responses were altered to conform to the hypothesis. For each subject, some number of individual responses which were contrary to the hypothesis — that is, for which an "active" response was predicted but a "passive" was produced,

[1]Specifically, the adjustment was $\frac{ActiveProp-0.5}{2}$: the subject's probability of responding "active" was adjusted by this proportion in the direction of the currently-more-frequent response. For instance, if 75% of a subject's responses so far had been active, the adjustment would be $\frac{0.75-0.5}{2} = 0.125$; thus the probability was moved 12.5% of the distance from the pre-adjustment probability toward 1. (If 25% of the subject's responses had been active, the probability would have been moved 12.5% toward zero.) So if the pre-adjustment probability (average of subject and stimulus probabilities) was 0.6, the adjusted probability would be 0.65 — that is, 12.5% closer to 1. No theoretical status is claimed for the details of this adjustment mechanism; it was simply a heuristic designed to "nudge" response probabilities in a manner that roughly approximated the self-priming behavior observed in actual subjects.

[2]That is, each "subject" was assigned an "active context" of either A or B, and then gave 12 responses in that context and 12 in the other.

or vice versa — were switched so that they conformed with the hypothesis. The amount of alteration — how many trials per subject had to be "forced" into conformance above and beyond the aforementioned "foundation" of subject-specific baselines, item-specific baselines, and self-priming adjustments — thus represents the size of the targeted context-dependence effect. Three different effect sizes were simulated, as described below.[3]

The above procedure was run 500 times to generate 500 simulated datasets of 40 "subjects" each. Each of these datasets was then submitted to a logistic regression using only the predictors Context, ActiveContext, and ActiveProp, and their two-way interactions, together with random effects for SubjID and PicName. The full model-selection process was not run, and the model was not whittled down to retain only significant predictors; rather, what was inspected was how often the *initial* model — including all three predictors and all their interactions — had a significant Context-ActiveContext interaction in the correct direction (such that context-congruent responses were more likely).[4]

The proportion of simulated models which found such an effect gives an indication of the statistical power of the procedure. Since the hypothesized effect was built into the simulation process (in the effect-generation step), an ideal model ought to be able to detect it. The proportion of models that failed to find the effect represents the probability of failing to detect a genuine effect (a so-called "Type II error", whose probability is denoted by β). Traditionally, a β of 0.2 is considered acceptable — that is, researchers accept that, about 20% of the time, their procedures will fail to detect an effect even when it really exists. A β of less than 0.2 is conservative (i.e., the analysis is very unlikely to miss a genuine effect), and a β exceeding 0.2 is generally considered unacceptable (i.e., it poses too high a risk of

[3]Note that the term "effect size" here does not refer to an effect size *measure* such as a coefficient in the logistic regression. Rather, "effect size" refers to the actual strength of the effect on subjects' responses, in the practical sense of how many responses were altered from the random-baseline response that would be expected by chance. For instance, an effect that managed to induce each participant to respond in a context-congruent manner on two trials is a stronger effect than one that induced a context-congruent response on just one trial per subject.

[4]This abbreviated model fitting process was chosen both in order to make the simulation computationally tractable, and to simplify comparison across the simulated and real datasets. Essentially, this process restricted focus to only those predictors which had a consistent impact on the models fit on the actual experimental data — namely, ActiveProp and the random effects — and then added the targeted predictors (ActiveContext and Context).

missing a genuine effect).

Three such 500-dataset simulations were run, with different effect sizes. The first simulation used an effect size of one response per subject, meaning that in the effect-generation step, one response per subject was switched into conformance — that is, one trial that would otherwise have had active response in a passive context, or vice versa, was switched so that the "right" response was produced. In this case, 404 of the 500 simulated datasets produced a Context-ActiveContext interaction that was in the right direction and significant at $p < 0.05$. This corresponds to a β of 0.192 — almost exactly the same as the traditional 0.2 value. This means that the sample size and procedures used in these experiments had sufficient power to detect an effect at traditionally accepted levels of significance, even if the effect was only strong enough to influence a single response for each subject.

The second simulation used an effect size of two responses per subject: that is, two responses per subject were switched into conformance. In this case, all 500 of the simulated datasets found the effect. This corresponds to a β of 0, which is considerably more conservative than the traditional level. (In fact, it represents a 0% chance of a false-negative error, which is not really achievable in practice.) This means that the sample size and procedures used here had ample power to detect an effect that was strong enough to influence two responses per subject — indeed, if the real effect in any of the experiments had been this strong, it would almost certainly have been detected.

Finally, in order to check that the simulation procedure was not inherently biased towards producting significant results, a third simulation was run without the effect-generation step — that is, only the subject-specific and picture-specific baselines and the self-priming adjustments were used, so the simulated effect size was zero. In this case, only 25 of the 500 simulations produced Context-ActiveContext interactions signficant at $p < 0.05$ (in any direction); thus the proportion of actual spurious effects was 0.05, exactly matching the p-value criterion. This means that the simulation was not inherently biased towards generating significant effects in line with the hypothesis, so we can be confident that the simulation was a reasonable recreation of the chance-baseline situation in the actual experimental task.

Note that, since each subject completed 24 trials, even changing two responses represents an observable effect on less than 10% of trials. With simulated data for this effect strength, the effect was detected 100% of the time. Thus, if the experiments conducted here missed a genuine effect, that effect was most likely rather small.

Obviously, this simulation process only crudely represented the chance-baseline behavior of the subjects, since the only parameters used were subject-specific and stimulus-specific active/passive response probabilities. It should be remembered, however, that in all of the actual experiments, a random-effects only model, making use of essentially the same information, performed better than the model accounting for the fixed effects. In other words, subject-specific and item-specific baselines apparently go a long way toward explaining the entirety of the actual experimental results — and yet the simulations show that a genuine effect, even a rather small one, *could* have been detected despite this high level of "background noise". This strongly suggests that the actual experiments did not fail to find an effect simply because the design or statistical analysis were weak, or because the sample sizes were inadequate, but because there really was no sizable effect to be found.

Chapter 10

General discussion

Taken as a whole, the results reported here do not provide support for the foundational UBTL hypothesis that speakers' usage of linguistic forms across contexts is a direct reflection of their experience with those forms across those contexts — although neither do they provide strong evidence against this hypothesis. In these experiments, although the association between voice and nonlinguistic context in the prime phase was always 100% reliable, there was no consistent effect of context on participants' choice of voice in the test phase.

Of course, there are many possible explanations for a null result. I will here address several such explanation, and argue that there is reason to believe that the null result here is of theoretical interest.

Perhaps the simplest explanation for the null result is that the sample sizes were too small. The power analysis presented earlier offers strong evidence that this is not the case. An additional argument against the insufficient-power criticism is that the sample sizes in these experiments were at least as large as those in previous experiments which did find a linguistic priming effect or a context-dependent memory effect. The five experiments reported here had effective sample sizes (i.e., not counting excluded participants) of 41, 15, 40, 56, and 46; the aggregate analysis of Experiments 1a-3 covered 152 subjects. As regards linguistic priming, Bock (1986) reported syntactic priming effects in three experiments with 48 subjects each; Branigan et al. (2000) reported syntactic priming in a dialogue task with 24 subjects; (Hartsuiker et al. 2004) found cross-linguistic syntactic priming with 24 subjects. Within research on context-dependent explicit memory, (S. M. Smith 1985, Experiment 1) obtained significant results using a musical context with 54 subjects; (S. M. Smith et al. 1978, Experiments 1-3) found results using a room context with sample sizes of 16, 24, and 20; (Godden and Baddeley 1975) found results using an underwater/on-land context manipulation with subject sizes of 16 and 18. Work on implicit context-dependent memory effects has tended to involve larger sample sizes, e.g., 192 subjects in the homophone-spelling task of S. M. Smith et al. (1990, Experiment 1), 90 subjects in the incidental word-memory study by Amanda Parker et al. (2001, Experiment 1), 65 to 93 subjects in the olfactory-context studies by Ball et al. (2010). However, the aggregate data of Experiments 1a-3 fits comfortably within even this range, and implicit effects have also been found with smaller sizes, e.g., 48 subjects in Amanda Parker et al. (2001, Experiment 2). As noted earlier, the studies by Gurevich et al. (2010) and Horton (2007) blend priming and context-dependence, and these studies essentially obtained implicit context-memory effects on language using quite small sample sizes. The verbatim-recall results of Gurevich et al. were found in five experiments, two with 20 subjects and three with 24 subjects; the partner-specific lexical priming effect was found by Horton in two experiments with sample sizes of 16 and 20 subjects.

Clearly, a larger sample size increases the chance of finding an effect. Nonetheless, given the substantial previous findings on the robustness of both syntactic priming and context-dependent memory effects, even with relatively small sample sizes, it should be equally clear that the sample sizes in the experiments reported here were not obviously too small. In particular, the aggregate results of Experiments 1a-3, encompassing 152 subjects, draws on a sample size considerably exceeding that in most syntactic priming research. The fact that aggregating the results resulted in the *disappearance* of effects from the smaller individual experiments suggests that, if anything, it was the positive results from those studies that were more likely to be red herrings.

It could also be argued that the contextual manipulations used in these studies were not strong enough to produce a context-dependence effect. However, this argument is unpersuasive, since earlier studies have used essentially the same context manipulations. Dulsky (1935) found background-color dependence effects; S. M. Smith (1985) obtained a context-dependence effect using the same musical contexts used in Experiment 3; numerous studies cited above have found a context-dependence effect when the manipulated context was the room in which priming/testing took place. In addition, as discussed earlier, the UBTL as currently articulated posits no particular constraints on what elements of the context might be relevant[1]; as matters currently stand, therefore, any contextual element is theoretically "in play" when it comes to effects on language.

Moving to more theoretical rather than methodological explanations, it should be noted that, if we relax any of the UBTL assumptions outlined in the first chapter of this dissertation, the findings are easily explained. If we allow that the pattern-recognition process may be domain-specific rather than domain-general, the lack of a result here could be due to such domain-specificity; since the pattern was not linguistic, it might not be recognized. If we allow that the rich-memory representations of an exemplar storage mechanism do not in fact always retain all information about every aspect of speech situation, then it may be that the nonlinguistic contextual information simply was not stored.[2] If we allow that different cognitive mechanisms may be at work for different levels of linguistic structure, then perhaps the syntactic level is not subject to the same cross-modal mapping processes that may be active in lexical learning. Finally, if we allow that recognition of linguistic patterns is restricted to (or at least highly privileged during) childhood, it may simply be that adult participants are not able to learn associations of this sort.

Another possibility is that the frequency of the voice-context association was insufficient to produce a noticeable difference in the participants' mental representations. A good deal of research (e.g. Bybee 2007; Bybee and Hopper 2001; Bybee et al. 1994) has argued that

[1] See the previously-cited statement by Bybee (2010) that rich memory representations of language "contain, at least potentially, all the information a language user can perceive in a linguistic experience."

[2] The small interaction effects that did involve the contextual manipulations could be explained by supposing that the contextual information was retained in some cases but not others — for instance, only on trials where the subject's attention was for some reason drawn to the context — so that the association between context and voice was more patchy.

frequency is a key factor in the emergence of language structure. Although it is certainly possible that an effect might be observed if the same experiment were re-run with a larger set of stimuli, there are no specific theoretical grounds for this belief. Goldberg et al. (2004) and Boyd and Goldberg (2012) found that adult subjects were able to learn an argument-structure generalization from only 16 short film clips (although this learning was assessed via a comprehension task rather than production). Horton (2007) reports that participants formed an association between words and the identities of specific people based on only 32 critical stimuli (not counting fillers). Participants in the study by Gurevich et al. (2010) apparently retained a significant memory trace of two 65-clause narrations over a duration of roughly a week; although this is more than the 24 sentences seen by participants in the present study, the fact that participants retained this information over such a long duration, even without any specific contextual pattern to associate it with, at least suggests that massive numbers of stimuli are not needed to successfully get tokens of language lodged in participants' brains. In short, while more stimuli, like more subjects, are always preferable, there is no particular reason to think that frequency was the determining factor here, and there are some previous studies suggesting that associations can be successfully created with even fewer stimuli.

One could also make the argument that the associations between syntactic voice and nonlinguistic context were ignored by participants because they were not meaningful. The most straightforward response to this objection is to ask how we know what counts as meaningful. For instance, in the studies by Boyd and Goldberg (2012) and Goldberg et al. (2004), children and adults were exposed to novel verbs in a novel syntactic construction. The "meaning" of the novel construction involved the act of appearance, and the different verbs indicated different manners of appearance; this meaning was communicated via videos of the various appearances, accompanied by sentences using the novel construction and novel verbs. It is not immediately clear why participants would be sensitive to the abstract notion of "appearance" as represented by a range of distinct videos, and yet not be sensitive to extremely concrete regularities such as the recurrence of the exact same sound or color with individual sentences — nor is there currently any UBTL explanation for this.[3]

In addition, UBTL proponents have cited the artificial-language-learning literature (e.g. Aslin et al. 1998; Saffran et al. 1999) as evidence of the power of pattern-recognition and cooccurrence tracking even in the absence of meaning; for instance, Beckner et al. (2009: 5-6) say that "such studies indicate that subjects learn patterns even when the utterance corresponds to no meaning or communicative intentions", implying that meaning is only incidental to the pattern-recognition abilities proposed by the UBTL. In this case, it should not matter whether the nonlinguistic context was "meaningful"; on the straightforward UBTL view, it should be enough that it was a pattern. It should also be noted that the experiment did not require participants to attribute meaning to the construction as such in order for an effect to be found; the only hypothesis was that the cooccurrence pattern would influence their syntactic choice — even if only in a nonconscious, noncommunicative, mechanical way — not that it would influence the meaning they perceived in the scene or attempted to convey in their description.

It is indeed unfortunate that we do not currently have a semantic theory that can clearly tell us which kinds of experiences are meaningful and which are not. However, lacking such a theory, it would be dubious to say that only "meaningful" associations can be learned, since we do not really know which associations are meaningful a priori. Indeed, if, as in Ibbotson's (2013) characterization of the UBTL position, "the meaning is the sum total of how the form is used in a communicative context", then to stipulate meaning as a prerequisite for associative learning is to put the cart before the horse: on this view, associations actually *become* meaningful by virtue of our learning them.

[3]Goldberg (1998, 2006) has suggested that "constructions that correspond to basic simple sentence types encode as their central senses event types that are basic to human experience" (Goldberg 1998: 205). As examples, she offers "for example, someone causing someone to receive something (the double object construction), something causing something to move (the caused-motion construction), or an instigator causing something to change state (the resultative construction)" (Goldberg 1998: 205). However, there is no explanation of how we are to know a basic pattern of experience when we see one. Even with Goldberg's own examples, it is not obvious why, for instance, "causing someone to receive something" is more "basic" than simply receiving something, or why "causing something to move" is more basic than moving, or, on the other hand, why either is more basic than "someone causing something". If we are to use meaningfulness as a criterion for pattern-learning, we need some independent, practical means of assessing how meaningful a situation is.

A related objection might be that, rather than the nonlinguistic context being too meaningless, the linguistic context was too meaningful. Given that active and passive voice already have different pragmatics and preferred contexts of use, it could be argued that these existing associations were too strong to be overridden, even partially, by the novel nonlinguistic association. However, this argument is difficult to reconcile with the extensive structural priming literature; if even a single hearing of passive voice can apparently influence a speaker's subsequent choice of active or passive (Bock 1986), it is hard to explain why such a choice would be impervious to a patterned array of hearings of passive voice. Moreover, the experiment only sought a quantitative shift in the proportion of the two constructions, which is exactly the short of incremental shift which, in theory, might occur of its own accord and instigate a larger-scale grammatical evolution (Bybee 2010, Chap. 7). If "there is no reason to suppose that quantitative changes in construction use cannot occur in adults" (Bybee 2010: 118), one would imagine that, if ever there were a situation where such quantitative change *could* occur, it would be in an associatively loaded situation such as that provided by the experiments reported here.

A more general version of these objections might simply be that, as suggested by L. B. Smith (2000: 170) and Ellis (2001: 42), attention is critical for the learning of associations, and therefore it is possible that subjects did not learn the context-voice association because they did not (for whatever reason) attend to the context. There are a number of rebuttals to this argument. First, as discussed in the introduction, although Ellis and other UBTL theorists have mentioned the potential role of attention in a general way, there is no straightforward hypothesis about when and how attention should be relevant. It is unclear, for instance, what it would mean for subjects in the "weird word order" studies (e.g. Akhtar 1999) to be "attending" to the specific manners of appearance which constitute the meaning of a novel lexical item, or how we would know whether they were doing so (rather than just attending to the overall video stimulus), or indeed how we would know whether they were doing so. More broadly, the notion that language users must attend to dimensions of experience in order to extract regularity from them is conceptually at odds with most formulations of exemplar-based theory, which posit that features need only be *perceived* (not attended), and that recognition of regularities is a separate cognitive process operating on "raw" data in the rich-memory store (see, e.g. Pierrehumbert 2001: 141).

A second piece of evidence against an insufficient-attention explanation comes from participants' responses to the post-test question which probed their awareness of the experimental manipulation. As mentioned earlier in passing, a sizable number of subjects responded to this question by inventing associations between the nonlinguistic context and the *meaning* of the sentences, ignoring their syntactic form, or the scenes in the pictures. A typical example comes from a subject in Experiment 3 (the music experiment), who claimed to have noticed a pattern: "The pictures tended to have some sort of disaster happening, or injuries were involved. As the injuries got potentially more fatal the music crescendoed." This subject obviously was attending to the music well enough to believe it bore some thematic relationship to the pictures (even though, since the music played in its own fixed tempo while the subject controlled her own pace through the pictures, it was not possible for there to be a prearranged relationship between the two). Similar responses were observed for the other contexts. Although subjects' responses were not coded in detail to assess how many made specific reference to the context, the repeated occurrence of references like this makes it implausible to suppose that many subjects totally ignored the nonlinguistic context.

Perhaps the strongest counter to an insufficient-attention explanation is that lack of explicit attention to the context is very nearly de rigeur in context-dependent memory studies. Indeed, such studies have often been at pains to ensure that subjects were *not* paying special attention to the context, in order to eliminate the possibility that the effects were driven by subjects' deliberate, conscious retrieval of the context. S. M. Smith et al. (1990: 239) say:

> At the study session, subjects were not instructed or encouraged to attend to the environment, nor was the context obviously related to the study material. On the homophone spelling test, subjects were not instructed or encouraged to

> attend to either the context or the studied events. Therefore, the findings of both experiments indicate that incidental background contextual information is both stored in memory and used to probe memory without the obvious intent of the subject.[4]

Since one goal of this dissertation was to test the relationship between context-dependent memory and the UBTL's exemplar/associative framework, it would not have been appropriate to draw subjects' attention to the context. This does not mean that subjects need to be totally oblivious to the context — as noted above, many subjects clearly were aware of the contexts, just as subjects in the studies of S. M. Smith et al. (1990) were no doubt aware of the visual appearance of the room around them. The point is simply that context-dependent memory does not require "special" attention or deliberate intent to attend, and thus, if such attention is necessary for the operation of the UBTL's rich memory system, then context-dependent memory as currently understood is apparently not operative in that rich memory system.

Finally, another possible explanation for the lack of significant results in the studies reported here is that the dependent measure — use of active versus passive voice — is too coarse for a subtle effect such as context-dependent memory to manifest itself. In other words, one might argue that syntactic priming can occur and context-dependent memory can occur, but that context-dependent memory is too weak a force to influence syntactic choice. This approach does not require abandoning any of the UBTL's core positions per se, but it requires facing up to their vagueness. Essentially, this objection amounts to saying "the UBTL exemplar model doesn't work like context-dependent memory" without saying what the UBTL exemplar model *does* work like. It is certainly reasonable to conclude on the basis of these results that the UBTL is probably not driven by context-dependent memory, but this is in fact new and useful information, given that there are no other comparably specific proposals for how the UBTL's driving mechansisms actually operate.

[4] It is also worth noting that S. M. Smith et al. (1990) used implicit *tests* of memory — that is, subjects not only were not told to attend to the context, but were not aware at test that their memory was what was being tested (they believed that an earlier cover task was the memory task and was already over).

In other words, the UBTL as it currently exists does not provide sufficient detail to craft a realistic account of what sorts of experiences influence what sorts of linguistic behavior. The current position within the UBTL is essentially that any contextual element could influence any linguistic choice. It is clear from context-dependent memory studies that contextual elements can influence various sorts of behaviors; it is also clear from structural priming studies that syntactic choice can be influenced. If the two do not match up, then the story is apparently not as simple as the current UBTL statements would have it. In short, the UBTL allows that any contextual feature can *potentially* influence language use; but because it does not say how to decide whether a given feature will *actually* influence language use, it does not give us much traction on the task of actually grounding language structure and patterns of use in patterns of experience. In such a situation, the natural approach is to take known mechansisms with similar properties and try them out to see if they produce the results the UBTL predicts.

Likewise, there are so far few or no accounts of the specifics of the memory mechanism which supposedly underlies linguistic knowledge. Context-dependent memory is one phenomenon whose properties make it a plausible candidate for this role, but the results presented here do not support the proposal of context-dependent memory as the operative force here. Therefore, again, more detail is needed in specifying the nature of the memory system posited by the UBTL, and of the cognitive operations which rely on it, in order to generate more precise predictions about how they should influence behavior.

It is also worth noting another implication of the results reported here: if contextual changes can affect linguistic production, the effects are probably very small. In particular, they are probably at least as small as the numerous other observed significant effects that do not involve the hypothesis. As has been seen, there were many effects involving interactions of "conditions" that were not really hypothesis-relevant manipulations, but simply methodological hedges or counterbalances to avoid undue influence of details of the experimental task. For instance, the UBTL hypotheses that motivated the study do not provide any particular reason to suppose that the sound of a sad trombone would produce different priming effects than that of a chime, and yet in several cases effects of this type were observed: effects were asymmetrical depending on which of the two nonlinguistic contexts was used. The fact that such effects occurred along with the hypothesized priming effect, and then disappeared along with it in the aggregate analysis of experiments 1a-3, suggests that even instrinsic influences of individual contextual elements may be as powerful as reliable regularities in their cooccurrence with language. In other words, even if the association of a sad trombone sound with passive voice has an effect on syntactic production, it may be that that effect is as weak as the intrinsic effect (whatever its source) of a sad trombone sound on syntactic production, regardless of what voice that sound was associated with. Although such "intrinsic" influences of particular stimuli are themselves presumably the result of associations formed throughout life, this illustrates the difficulty of providing a clear specification of what sorts of patterns can be incorporated into linguistic representations: there is no obvious way for the system to separate the regularities that represent signal from those that represent noise.

The one consistent finding across all experiments is the self-priming effect: the more participants have used a particular voice, the more likely they are to continue using it. This is essentially a replication of the basic structual priming effect: exposure to a structure such as voice (in this case, by using it oneself) increases the likelihood of use of the same structure again. It is important to note, however, that this effect is *non-associative* with regard to the parameters of the experiment. That is, although there was a completely reliable correlate of voice during the priming phase (namely the nonlinguistic contextual variable), it does not appear to be the case that use of a particular voice in that context strengthened the association between the voice and that correlate, at least not enough to influence subsequent production.

Thus, if we take as given that, as Pickering and Ferreira (2008) suggest, "hearing or producing the prime strengthens the processing sequence that yielded the prime structure", the present results do not support the hypothesis that the processing sequence yielding passive voice takes account of the nonlinguistic context during that processing. Alternatively, if instead we take as given that processing of the passive voice is part and parcel of processing the entire nonlinguistic context, then the present results do not support the hypothesis that perceiving the prime strengthens that processing sequence. In other words, the present results do not support the dual assertion that language processing takes account of the nonlinguistic context, and that things which are processed together thereby form an associative link. These are both core tenets of much usage-based work. Therefore, either one or both of these assumptions should be revised or abandoned.

At the same time, it is important to note that the current findings do not in themselves falsify any of these tenets. Taken jointly and at face value, the various core claims of the UBTL would predict results that were not found here, and thus, as noted above, the current findings do not support these claims taken jointly and at face value. However, as also discussed above, relaxing any one of several core claims would provide an explanation for the null results; therefore, even if some of the claims are false, the present results do not allow us to know which ones they are.

Moreover, it is possible that all the claims could be preserved in their general form if they were tempered by additional constraints on the kinds of context that are stored, the kinds of patterns that can be extracted, the amount of data required for such patterns to be extracted, etc. One simple and plausible constraint would be that the pattern-recognition system does not generally track cooccurrences between language and non-focal elements of nonlinguistic experience. Since the focus of the experimental task was obviously on the displayed sentence/picture and not the color/music/sound/room, these contextual elements may have been ignored by the pattern-recognition system, even if they were attended by subjects at least part of the time. Another plausible constraint would be that adults do not track cooccurrences along dimensions which have not been informative in their experience so far — that is, that we learn what aspects of experience to ignore, and that we have learned to ignore color/music/sound/room, and disregard them as possible sources of information about language-relevant patterns. Such constraints are very likely necessary in order to rein in the scope of the patterns recognizable under the UBTL. Crucially, however, detailed work is needed to decide what constraints are operative, and until that work is done, the UBTL leaves us in the dark about what is going on when regularities of experience do not seem to have any effect on language use.

As outlined in the introduction, various researchers have provided considerable evidence *for* the various tenets of the UBTL; therefore, the conclusion here should not be that the theory, or even any of its individual claims, are hopeless, but rather that a great deal more work is necessary to flesh those claims out and explain how they interact with one another.

Chapter 11

Conclusion

Although the results reported here are not supportive of the broad claims of the UBTL, I do not wish to argue that the UBTL is therefore on the wrong track. Rather, what I suggest is that the UBTL has well-enough established its overall gist, and that it is time to begin nailing down the specifics. As noted earlier, we have ample evidence for many of the individual tenets of the UBTL — that domain-general processes are active in language use, that memory plays a role, etc. The next step is to move from arguing that such factors play a role to crafting explicit, predictive accounts of exactly how they work — that is, under what conditions do what kinds of language structures, mental representations of linguistic structures, or patterns of linguistic behavior emerge?

Like many functionally-oriented theories of language, the UBTL claims that what speakers know about their language is influenced by a broad array of factors, many of them not purely or inherently linguistic in nature. More specifically, it is claimed that cooccurrence patterns among various dimensions of experience will affect mental representations of language. However, it is important to distinguish between three levels of strength at which such a claim may be framed, each of which makes correspondingly strong predictions. The weakest version of the claim is that language can be sensitive to such cooccurrence patterns; this claim simply predicts that we will find at least some cases where patterns of experience affect language structure, but they may be quite rare or peripheral. A stronger claim is that language, by and large, *is* sensitive to such patterns; this predicts that in general most language structures will show the influence of patterns of experience in some way, but leaves open the possibility that language structure may also be fundamentally driven by other factors, with the pattern-based component as only a minor part of how any given language structure works. The strongest claim is that patterns of experience are the core mechanism by which language structures arise; this predicts not only that all language structures will show the influence of patterns of experience, but that essentially everything about these structures can be explained by reference to patterns of experience.

In short: is it the case that language *can be influenced* by patterns of experience, that language *is influenced* by patterns of experience, or that language *is created* by patterns of experience?[1] I believe we have substantial evidence for the first claim, a smaller but growing body of evidence for the second claim, and relatively little direct evidence for the third. Research in usage-based linguistics has accumulated a vast amount of evidence that linguistic congition is not a hermetically sealed, autonomous module unaffected by life experience; this research has moreover demonstrated that the interface between language and patterns of experience is fairly pervasive. We have, in other words, evidence that "the set of cognitive and neuromotor mechanisms or activities that are put into use in online communication and in the mental storage of language" (Bybee 2010: 33) is sensitive to patterns of usage. However, this is a long way from establishing Bybee's subsequent claim that "the particular way these processing mechanisms work determines fairly directly the facts about the nature of language" (Bybee 2010: 33) — in large part because our understanding of "the particular way these processing mechanisms work" is still quite vague.

Insofar as the UBTL's basic claims make predictions about what kinds of behavior should be observed in many situations, those predictions are extremely broad, and, as the results presented here suggest, are not necessarily borne out. The UBTL predicts generally that "the structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes" (Beckner et al. 2009: 2), but has so far placed few

[1] Perhaps an even stronger version would claim that language is *constituted* by patterns of experience.

constraints on what sorts of patterns may thus give rise to language. This leaves many unanswered questions about exactly how the emergence happens. If there is a regular pattern of occurrence linking a linguistic structure and a contextual feature, will this cause people to use the structure more, or differently? Will it produce a qualitiative change (e.g., use of a different construction) or a quantitative one (e.g., use of the same constructions in different proportions)? Will it produce an effect only on comprehension but not on production? How regular must the association be for these effects to be observed? How much exposure must people have to the association for any of these effects to be observed? Is the effect modulated by the specifics of the stimuli (i.e., the particular construction and/or particular aspect of the context)? If so, how?

I suggest that our understanding of how language works would be greatly advanced by a concerted effort to address these questions in a systematic way. By "a systematic way" I mean the straightforward scientific enterprise of formulating specific proposals and testing whether they work — that is, whether they are demonstrably powerful and versatile enough to drive the construction of all language structure out of patterns of experience.

I emphasize that the broad gist of the UBTL may be correct even though it does not yet address these questions — indeed, the evidence from various studies cited throughout this dissertation strongly points in that direction. However, it is essential not to regard questions such as the above as mere details which are needed only to iron out the wrinkles. These questions are central to the causal import of the UBTL. If "the particular way these processing mechanisms work" is supposed to determine language structure, an explicit characterization of that particular way is vital to the theory.[2]

In particular, it is important that the theory evolve beyond retrospective explanation of existing grammar to the identification of causal factors which operate in a specifiable and predictable way. A good deal of existing UBTL research reasons from effects to causes — for instance, from existing constructions to past patterns of usage which are believed

[2] Ibbotson (2013) makes a similar point: "Usage-based theories need more psychologically plausible models of what gets treated as a chunk when, and hence 'counted' in any distributional analysis [...] A key part of responding to this challenge will be to specify in greater detail the mechanisms of generalization, specifically a mechanistic account of the dimensions over which children and adults make (and do not make) analogies."

to have resulted in grammaticalization (e.g. Bybee 2007; Bybee and Hopper 2001; Bybee et al. 1994). What we need now are predictions from causes to effects. It is not enough to show that what already happened can be interpreted in terms of certain proposed past causes; a solid theory needs to actually predict, on the basis of presently observed factors that are believed to be causes, what effects will be observed in the future.

Given that virtually every aspect of language structure is likely to have multiple causes, it is of course not necessary that every purported cause lead infallibly to its predicted effect in every case. Nor it is necessary that the cause-effect relationship be tested exclusively via experimental methods; it is quite possible to look at causes and effects by finding cases where the cause occurred in naturalistic data, and seeing if the effect followed. (In other words, we do not have to predict what will *happen* in the future, we only have to predict what we will *find* in the future, before we actually go looking for it — even if where we eventually go looking for it is also in data from the past.) The point is simply that we cannot assess the power of the theory only by looking for the effects and seeing if the supposed causes preceded them; we must take the hypothesized causes as a starting point, see how reliably the effects follow, and keep score of which proposed causes actually produce the predicted effects.

Currently, we have only a vague notion of what kinds of situations actually *cause* what kinds of changes in language structure (or mental representations thereof). Claims such as "the kind of constituency normally studied by syntacticians also has its source in language use and frequency of co-occurrence. [...] Constructions [...] are conventionalized through frequent use" (Bybee and Hopper 2001: 14) and "an exemplar is built up from a set of tokens that are considered by the organism to be the same on some dimension" (Bybee 2010: 19) imply that similarity and cooccurrence are the driving factors. But does this mean that whenever some experiences are similar and/or cooccur in some pattern, language structure should be affected? If so, then we should look for every possible case of similar and/or regularly cooccurring experiences and see if language structure really is affected. If not, what additional information is necessary to predict what will actually happen? I concur with Bybee 2010: 62 that "the pressing need is for fully elaborated substantive theories that predict which similarities will be important." The present study, for instance, suggests that similarities in incidental environmental context are not (or at most only very weakly) important.

We must also avoid the related pitfall of identifying the hypothesized causes only via their effects. As an example, consider the claim that "repetition of the same inferences builds their strength in the exemplar representation of meaning and context. Eventually the inferences become part of the meaning" (Bybee 2010: 159). If inference is the cause and semantic change the effect, we must be careful to avoid analyzing only for those inferences which have already been absorbed by semantic change. To show that patterns of inference drive semantic change, we need to be able to identify inferences *independently* of their conventionalization into semantic meaning. If we only investigate "successful" inferences (i.e., those that did become conventionalized into semantics), we cannot know what distinguishes them from the unsuccessful inferences (i.e., those did not make it into the meaning).

Again, these are not minor details. A theory that predicts many effects that do not happen is just as faulty as one that fails to predict effects that do happen, and in its current form, the UBTL seems to predict many effects for which, at the least, we have little evidence (such as those which the experiments reported here sought and, for the most part, did not find). As Ibbotson (2013) notes, fixing this overprediction "would also go a long way to engaging those from different theoretical approaches to language acquisition who are skeptical of the usage-based approach for this reason — 'what pattern of results *couldn't* a usage-based theory explain.'"

In the same vein, I stress that this issue of is not specific to any particular formulation of the UBTL or its claims, but is fundamental to the very idea of a usage-based approach to language. As mentioned in the introduction, although I have referred to "the" UBTL, there is no single, monolithic, explicitly articulated theory as such, and hence I am not merely claiming that the findings here support or refute specific claims by particular researchers.

Rather, the core *concept* of language structure arising from usage is what is at issue here. No matter how you slice it, the claim that language structure arises from patterns of usage ultimately requires affirmative demonstrations of a cause-effect relationship between patterns of usage and language structure, and a thoroughgoing explanation of the causal mechanisms by which the patterns of usage effect change in language structure.

It is important to note that the crafting of such a causal account is a fundamentally different task from the gradual accumulation of findings about *influences* of patterns of experience on language structure. For instance, no matter how many results we obtain showing that people retain detailed memory traces of linguistic experiences, and no matter how many results we obtain showing that existing constructions have plausible origins in past patterns of usage, we will never, by those findings, convincingly show that patterns of usage drive language structure. To do that, the findings in the areas of memory and historical change (along with inference, chunking, cross-modal association, etc.) must be integerated into a mechanistic account of how the system actually works as a whole. Findings demonstrating *influence* of particular experiences on language structure are immensely useful in that they draw us toward hypotheses about how the system works, but we still need to actually test the resulting hypotheses directly to see whether the disparate individual findings can in fact be melded into a coherent account.

In its current state, the UBTL predicts that any pattern in experience should be picked up on by the language system. If this claim is true, we should be able to demonstrate this by finding patterns of experience — either manufactured in an experiment or identified in natural settings in a manner independent of their effects — and predicting their effects on mental representations of language. If the claim is false, then there must be additional constraints on how the system recognizes patterns, and those constraints need to be identified and built into the theory. In short, when we claim that language arises from usage, we have committed ourselves to the claim that the patterns that are represented in language must be there because certain patterns of experience "out-competed" other patterns in some way, and thus were noticed and learned by the cognitive system. We need to know what the rules of that competition are: how does our cognitive apparatus select among the many patterns that confront us?

Importantly, this enterprise must involve looking for patterns beyond language. As I have emphasized throughout this dissertation, the domain-generality position of the UBTL implies that whenever we make a claim about how patterns of experience affect linguistic representations, we are also making a claim about how they affect mental representations in general. The UBTL would therefore do well to draw on relevant work in fields outside linguistics, not only at the general level of importing concepts such as memory and chunking, but at the more specific level of assessing in detail how the operation of these processes within language is or is not parallel to their operation in other domains.

I have discussed memory in this context already: if memory is key to the UBTL, we need to know whether properties of memory in other domains apply with regard to language, and if not, we need to characterize the differences. The experiments reported here represent an attempt to connect linguistic exemplar memory with at least one conceptually similar phenomenon, namely context-dependent memory. Similar cross-disciplinary comparisons are possible in other areas. As discussed in the introduction, modern conceptions of classical conditioning (Jara et al. 2006; Perales et al. 2004; Rescorla 1988) suggest parallels to the pattern learning posited by the UBTL. There is also more general research on inductive learning (Holland et al. 1989). Reaching even further afield, processes or models similar to those of the UBTL (e.g., analogy, pattern extraction, connectionism, emergence) have been proposed to underlie consciousness (in philosophy, e.g. P. M. Churchland and P. S. Churchland 1998; Dennett 1993; and in some cases formalized into mathematical models, e.g., Laakso and Cottrell 2000), the self (in neuroscience, see Damasio 2012), and indeed all of human thought — a recent general-audience volume (Hofstadter and Sander 2013) is subtitled "Analogy as the fuel and fire of thinking". Assessment of the domain-generality of the UBTL's proposed mechanisms would benefit from contextualization with respect to their counterparts in these other domains.

The basic point is that a process is not domain-general just because we can label it with a word (such as "chunking", "memory", or "analogy") that also labels processes in other domains; a process is domain-general if we have evidence that it actually operates in the same way across domains. It is just not possible to know whether the cognitive mechanisms underlying language are domain-general without explicitly comparing them to nonlinguistic cognitive processes.

In essence, the arguments above amount to saying that the UBTL's reach currently exceeds its grasp. Although researchers have made immense progress in demonstrating how language is sensitive to patterns of experience, including nonlinguistic experience, as yet we have little evidence that patterns of experience are the fundamental cause of language structure. Within the UBTL, there have been broad claims about the ability of chunking and cross-modal association to provide a framework for language; although there is considerable evidence that these processes are active and relevant, there is as yet no concrete, causal account of how they work to produce actual grammars directly from patterns of experience. There have also been broad claims that many incidental details of experience are stored in a rich memory system; although there is considerable evidence that memory for language may be richer than has been assumed, there is no concrete evidence to support the claim that this memory system actually contains every aspect of every experience with language. In addition, there is little explanation of where this memory system fits in relation to other well-studied memory systems. Finally, the UBTL emphasizes the domain-generality of its proposed cognitive mechansisms, such as analogy, chunking, and cross-modal association; although there is evidence that such mechanisms can in principle be domain-general, there is little direct evidence that what they do in the language domain is actually the same as what they do in other domains. This is in large part because UBTL work to date has focused primarily on language-internal patterns, whereas, if the cognitive underpinnings of language are as powerful as the UBTL would have it, and are essentially domain-general, they ought to also cognitively underpin a great deal of highly structured knowledge in other domains. In other words, the UBTL actually predicts a lot about nonlanguage, but its implications have only been considered with regard to language.

In many of these areas, there are clear paths forward, some of which I have noted in passing throughout this dissertation. For instance, the UBTL would benefit greatly from a clarification of the role of meaning in the pattern-recognition process: is meaning a causal factor in pattern-recognition (i.e., are we more attuned to meaningful patterns than to meaningless ones, and if so, how do we know which is which?), or is meaning only a result of pattern-recognition (i.e., patterns acquire meaning by being recognized)? Studies in language acquisition (e.g., the "weird word order" studies of Akhtar 1999 and others) have already begun to demonstrate causal relations between patterns of experience and language structure; we need to work on extending this causal predictiveness to synchronic adult language, as well as to diachronic studies. Studies such as those by Horton (2007) and Gurevich et al. (2010) offer tantalizing demonstrations of the cross-modal association mechanism that is claimed to underlie language, but the experiments of the present study suggest that there are limits on this mechanism; we need to narrow the gap between these results, systematically investigating what kinds of associations are and are not learned.

In the tale of Ali Baba and the forty thieves, a bandit chalks a symbol on Ali Baba's house so that he and his comrades can return later to take their revenge on the unsuspecting woodcutter. Ali Baba's slave, however, notices the mark, and she cleverly thwarts the thieves — not by erasing the mark, but by drawing the same mark on all the nearby houses. This tactic succeeds because, when the thieves return, it is not enough for them to return to the correct house and see the mark; they cannot know which house is the right one unless they also do *not* see the mark on any other house. When the UBTL's reach exceeds its grasp as described above, we risk being fooled in the same way. Evidence to date has made it quite clear that language structure and use are sensitive to the factors posited by the UBTL — that is, we can see the mark on the door. However, this does not mean that the UBTL's explanation of how these factors operate is correct, because those posited factors ought to have much broader effects than the purely linguistic ones that have been studied — in other words, it seems quite likely that the UBTL, in its generality, has marked every door in the neighborhood. In order to hone in on a coherent, causal theory of language — Ali Baba's true house, as it were — we need to begin erasing the extra marks by narrowing the theory down.

References

Akhtar, Nameera (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of child language* 26.02, pp. 339–356.

Akhtar, Nameera and Michael Tomasello (1997). Young children's productivity with word order and verb morphology. *Developmental psychology* 33.6, p. 952.

Alishahi, Afra, Afsaneh Fazly, and Suzanne Stevenson (2008). "Fast mapping in word learning: What probabilities tell us". In: *Proceedings of the Twelfth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, pp. 57–64.

Alishahi, Afra and Suzanne Stevenson (2008). A computational model of early argument structure acquisition. *Cognitive Science* 32.5, pp. 789–834.

Aslin, Richard N, Jenny R Saffran, and Elissa L Newport (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science* 9.4, pp. 321–324.

Ball, Linden J, Jaswinder Shoker, and Jeremy NV Miles (2010). Odour-based context reinstatement effects with indirect measures of memory: The curious case of rosemary. *British Journal of Psychology* 101.4, pp. 655–678.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5. URL: `http://CRAN.R-project.org/package=lme4`.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann (2009).

Language is a complex adaptive system: Position paper. *Language Learning* 59.s1, pp. 1–26.

Bergen, Benjamin and Kathryn Wheeler (2010). Grammatical aspect and mental simulation. *Brain and Language* 112.3, pp. 150–158.

Bock, Kathryn (1986). Syntactic persistence in language production. *Cognitive psychology* 18.3, pp. 355–387.

— (1989). Closed-class immanence in sentence production. *Cognition* 31.2, pp. 163–186.

Bock, Kathryn and Zenzi M Griffin (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129.2, p. 177.

Bock, Kathryn and Helga Loebell (1990). Framing sentences. *Cognition* 35.1, pp. 1–39.

Bock, Kathryn, Helga Loebell, and Randal Morey (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological review* 99.1, p. 150.

Boyd, Jeremy K and Adele E Goldberg (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87.1, pp. 55–83.

— (2012). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of child language* 39.3, p. 457.

Boyd, Jeremy K, Erin A Gottschalk, and Adele E Goldberg (2009). Linking rule acquisition in novel phrasal constructions. *Language Learning* 59.s1, pp. 64–89.

Boyland, Joyce Tang (1996). "Morphosyntactic change in progress: A psycholinguistic approach". PhD thesis. University of California, Berkeley.

Branigan, Holly P, Martin J Pickering, and Alexandra A Cleland (2000). Syntactic coordination in dialogue. *Cognition* 75.2, B13–B25.

Bransford, John D and Jeffery J Franks (1971). The abstraction of linguistic ideas. *Cognitive psychology* 2.4, pp. 331–350.

Bybee, Joan (2001). *Phonology and language use*. Vol. 94. Cambridge University Press.

Bybee, Joan (2002). Sequentiality as the basis of constituent structure. In: *The evolution of language from pre-language*. Ed. by Talmy Givón and Bertram F. Malle. Amsterdam: John Benjamins, pp. 109–132.

— (2007). *Frequency of use and the organization of language*. Oxford University Press.

— (2010). *Language, usage and cognition*. Vol. 98. Cambridge University Press.

Bybee, Joan and Clay Beckner (2009). Usage-based theory. In: *The Oxford handbook of linguistic analysis*. Ed. by Bernd Heine and Heiko Narrog, pp. 827–856.

Bybee, Joan and Paul J Hopper (2001). *Frequency and the emergence of linguistic structure*. Vol. 45. John Benjamins Publishing.

Bybee, Joan and James L McClelland (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22.2-4, pp. 381–410.

Bybee, Joan, Revere Perkins, and William Pagliuca (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Bybee, Joan and Joanne Scheibman (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37.4, pp. 575–596.

Bybee, Joan and Sandra A Thompson (2000). "Three frequency effects in syntax". In: *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*. Vol. 23, pp. 65–85.

Carreiras, Manuel, Ludovic Ferrand, Jonathan Grainger, and Manuel Perea (2005). Sequential effects of phonological priming in visual word recognition. *Psychological Science* 16.8, pp. 585–589.

Casenhiser, Devin and Adele E Goldberg (2005). Fast mapping between a phrasal form and meaning. *Developmental Science* 8.6, pp. 500–508.

Chang, Franklin, Gary S Dell, Kathryn Bock, and Zenzi M Griffin (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research* 29.2, pp. 217–230.

Churchland, Paul M and Patricia Smith Churchland (1998). *On the contrary: Critical essays, 1987-1997*. MIT Press.

Clark, Herbert H (1996). *Using language*. Vol. 4. Cambridge University Press Cambridge.

Cleland, Alexandra A and Martin J Pickering (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49.2, pp. 214–230.

Croft, William (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

— (2003). *Typology and universals*. Cambridge University Press.

Dalton, Pamela (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition* 21.2, pp. 223–234.

Damasio, Antonio (2012). *Self comes to mind: Constructing the conscious brain*. Random House LLC.

Dennett, Daniel C (1993). *Consciousness explained*. Penguin UK.

Dulsky, Stanley Gollis (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology* 18.6, p. 725.

Ellis, Nick C (1996). Sequencing in SLA. *Studies in second language acquisition* 18.01, pp. 91–126.

— (2001). Memory for language. *Cognition and second language instruction*, pp. 33–68.

— (2013). Frequency-based grammar and the acquisition of tense-aspect in L2 learning. In: *Research Design and Methodology in Studies on Second Language Tense and Aspect*. Ed. by Rafael Salaberry and Llorenç Comajoan. Mouton de Gruyter, pp. 89–118.

Estival, Dominique (1985). Syntactic priming of the passive in English. *Text* 5.1-2, pp. 7–22.

Fazly, Afsaneh, Afra Alishahi, and Suzanne Stevenson (2008). "A probabilistic incremental model of word learning in the presence of referential uncertainty". In: *Proceedings of the 30th Annual Conference of the Cognitive Science Society*.

Fillmore, Charles J, Paul Kay, and Mary Catherine O'connor (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, pp. 501–538.

Fox, John (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software* 8.15, pp. 1–27.

Givón, Talmy (1979). *On understanding grammar*. Academic Press.

Glenberg, Arthur M and Michael P Kaschak (2002). Grounding language in action. *Psychonomic bulletin & review* 9.3, pp. 558–565.

Godden, Duncan R and Alan D Baddeley (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology* 66.3, pp. 325–331.

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

— (1998). Patterns of experience in patterns of language. In: *The new psychology of language*. Ed. by Michael Tomasello. Lawrence Erlbaum Associates Mahwah, NJ, pp. 203–19.

— (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

— (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22.1, pp. 131–153.

Goldberg, Adele E, Devin M Casenhiser, and Nitya Sethuraman (2004). Learning argument structure generalizations. *Cognitive Linguistics* 15.3, pp. 289–316.

Gries, Stefan Th (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34.4, pp. 365–399.

Gries, Stefan Th and Stefanie Wulff (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7.1, pp. 163–186.

Gurevich, Olga, Matthew A Johnson, and Adele E Goldberg (2010). Incidental verbatim memory for language. *Language and Cognition* 2.1, pp. 45–78.

Haiman, John (1985). *Natural syntax*. Cambridge University Press.

— (1994). Ritualization and the development of language. In: *Perspectives on grammaticalization*. Ed. by William Pagliuca. John Benjamins, pp. 3–28.

Hartsuiker, Robert J and Herman HJ Kolk (1998). Syntactic persistence in Dutch. *Language and Speech* 41.2, pp. 143–184.

Hartsuiker, Robert J, Herman HJ Kolk, and Philippine Huiskamp (1999). Priming word order in sentence production. *The Quarterly Journal of Experimental Psychology: Section A* 52.1, pp. 129–147.

Hartsuiker, Robert J, Martin J Pickering, and Eline Veltkamp (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science* 15.6, pp. 409–414.

Hartsuiker, Robert J and Casper Westenberg (2000). Word order priming in written and spoken sentence production. *Cognition* 75.2, B27–B39.

Hay, Jennifer and Joan Bresnan (2006). Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23.3, pp. 321–349.

Henke, Katharina (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience* 11.7, pp. 523–532.

Hofstadter, Douglas and Emmanuel Sander (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.

Holland, John, Keith Holyoak, R Nisbet, and Paul Thagard (1989). *Induction: Processes of learning, inference and discovery*. Cambridge, MA: MIT Press.

Hopper, Paul J (1987). Emergent grammar. *Berkeley Linguistics Society* 13, pp. 139–157.

Hopper, Paul J and Sandra A Thompson (1980). Transitivity in grammar and discourse. *Language*, pp. 251–299.

Hopper, Paul J and Elizabeth Closs Traugott (2003). *Grammaticalization*. Cambridge University Press.

Horton, William S (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes* 22.7, pp. 1114–1139.

Ibbotson, Paul (2013). The scope of usage-based theory. *Frontiers in psychology* 4.

Jackendoff, Ray (2007). Linguistics in cognitive science: The state of the art. *The linguistic review* 24.4, pp. 347–401.

Jacoby, Larry L and Dawn Witherspoon (1982). Remembering without awareness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 36.2, p. 300.

Jaeger, T. Florian and Neal Snider (2008). "Implicit learning and syntactic persistence: Surprisal and cumulativity". In: *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*. Washington, D.C., pp. 1061–1066.

James, William (1890). *Principles of Psychology*. New York: Henry Holt and Company.

Jara, Elvia, Javier Vila, and Antonio Maldonado (2006). Second-order conditioning of human causal learning. *Learning and motivation* 37.3, pp. 230–246.

Johnson-Laird, Philip N (1970). The perception and memory of sentences. In: *New Horizons in Linguistics*. Ed. by John Lyons. Harmondsworth, Middlesex: Penguin, pp. 261–270.

Johnson-Laird, Philip N and Rosemary Stevenson (1970). Memory for syntax. *Nature*, p. 412.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D Raymond (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language* 45, pp. 229–254.

Kidd, Evan and Minna Kirjavainen (2011). Investigating the contribution of procedural and declarative memory to the acquisition of past tense morphology: Evidence from Finnish. *Language and Cognitive Processes* 26.4-6, pp. 794–829.

Krafka, Carol and Steven Penrod (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology* 49.1, p. 58.

Krug, Manfred (1998). String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26.4, pp. 286–320.

Laakso, Aarre and Garrison Cottrell (2000). Content and cluster analysis: assessing representational similarity in neural systems. *Philosophical psychology* 13.1, pp. 47–76.

Langacker, Ronald W (1987). *Foundations of Cognitive Grammar: theoretical prerequisites. Volume 1*. Vol. 1. Stanford university press.

Lieven, Elena VM, Julian M Pine, and Gillian Baldwin (1997). Lexically-based learning and early grammatical development. *Journal of child language* 24.01, pp. 187–219.

Lindblom, Björn, Peter MacNeilage, and Michael Studdert-Kennedy (1984). Self-organizing processes and the explanation of language universals. In: *Explanations for language universals*. Ed. by B. Butterworth, B. Comrie, and Ö. Dahl. Berlin/New York: Walter de Gruyter, pp. 181–203.

Lum, Jarrad AG and Evan Kidd (2012). An Examination of the Associations Among Multiple Memory Systems, Past Tense, and Vocabulary in Typically Developing 5-Year-Old Children. *Journal of Speech, Language and Hearing Research* 55.4, p. 989.

Malpass, Roy S and Patricia G Devine (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology* 66.3, p. 343.

Marslen-Wilson, William, Lorraine K Tyler, Rachelle Waksler, and Lianne Older (1994). Morphology and meaning in the English mental lexicon. *Psychological review* 101.1, p. 3.

Mathe, Stefan, Afsaneh Fazly, Sven Dickinson, and Suzanne Stevenson (2008). "Learning the abstract motion semantics of verbs from captioned videos". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08*. IEEE, pp. 1–8.

McClelland, James L and Joan Bybee (2007). Gradience of Gradience: A reply to Jackendoff. *The Linguistic Review* 24.4, pp. 437–455.

Meyer, David E and Roger W Schvaneveldt (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology* 90.2, p. 227.

Miller, George A (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 63.2, p. 81.

Mulligan, Neil W (2011). Conceptual implicit memory and environmental context. *Consciousness and cognition* 20.3, pp. 737–744.

Nakagawa, Shinichi and Holger Schielzeth (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4.2, pp. 133–142.

Newell, Allen (1990). *Unified theories of cognition*. MIT Press.

Ochsner, Kevin N, C-Y Peter Chiu, and Daniel L Schacter (1994). Varieties of priming. *Current Opinion in Neurobiology* 4.2, pp. 189–194.

Parker, Amanda, Angus Gellatly, and Mitchell Waterman (1999). The effect of environmental context manipulation on memory: Dissociation between perceptual and conceptual implicit tests. *European Journal of Cognitive Psychology* 11.4, pp. 555–570.

Parker, Amanda, Henny Ngu, and Helen J Cassaday (2001). Odour and Proustian memory: reduction of context-dependent forgetting and multiple forms of memory. *Applied cognitive psychology* 15.2, pp. 159–171.

Parker, Andrew, Neil Dagnall, and Anne-Marie Coyle (2007). Environmental context effects in conceptual explicit and implicit memory. *Memory* 15.4, pp. 423–434.

Pavlov, Ivan Petrovich (1927). *Conditioned reflexes*. Oxford University Press.

Perales, José C, Andrés Catena, and Antonio Maldonado (2004). Inferring non-observed correlations from causal scenarios: The role of causal knowledge. *Learning and Motivation* 35.2, pp. 115–135.

Pickering, Martin J and Victor S Ferreira (2008). Structural priming: a critical review. *Psychological bulletin* 134.3, p. 427.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition and contrast. In: *Frequency and the Emergence of Linguistic Structure*. Ed. by Joan L. Bybee and Paul Hopper. John Benjamins, pp. 137–157.

Pine, Julian M and Elena VM Lieven (1993). Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language* 20, pp. 551–551.

— (1997). Slot and frame patterns and the development of the determiner category. *Applied psycholinguistics* 18, pp. 123–138.

Port, Robert F (2010). Rich memory and distributed phonology. *Language Sciences* 32.1, pp. 43–55.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

Redington, Martin, Nick Crater, and Steven Finch (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22.4, pp. 425–469.

Rescorla, Robert A (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist* 43.3, p. 151.

Sachs, Jacqueline Strunk (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2.9, pp. 437–442.

Saffran, Jenny R (2002). Constraints on statistical language learning. *Journal of Memory and Language* 47.1, pp. 172–196.

Saffran, Jenny R, Elizabeth K Johnson, Richard N Aslin, and Elissa L Newport (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70.1, pp. 27–52.

Savage, Ceri, Elena Lieven, Anna Theakston, and Michael Tomasello (2006). Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development* 2.1, pp. 27–49.

Schacter, Daniel L and Endel Tulving (1994). What are the memory systems of 1994? In: *Memory systems 1994*. Ed. by Daniel L Schacter and Endel Tulving. MIT Press, pp. 1–38.

Scheepers, Christoph (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition* 89.3, pp. 179–205.

Schenkein, James (1980). A taxonomy for repeating action sequences in natural conversation. In: *Language production*. Ed. by Brian Butterworth. Vol. 1, pp. 21–47.

Smith, Linda B (2000). Avoiding associations when it's behaviorism you really hate. In: *Becoming a word learner: A debate on lexical acquisition*. Ed. by Roberta Michnick Golinkoff, Kathy Hirsh-Pasek, Lois Bloom, Linda B Smith, Amanda L Woodward, Nameera Akhtar, Michael Tomasello, and George Hollich. Oxford University Press, pp. 169–174.

Smith, Steven M (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory* 5.5, p. 460.

— (1985). Background music and context-dependent memory. *The American journal of psychology*, pp. 591–603.

— (1994). Theoretical principles of context-dependent memory. In: *Theoretical aspects of memory*. Routledge, pp. 168–195.

Smith, Steven M, Arthur Glenberg, and Robert A Bjork (1978). Environmental context and human memory. *Memory & Cognition* 6.4, pp. 342–353.

Smith, Steven M, Fred R Heath, and Edward Vela (1990). Environmental context-dependent homophone spelling. *The American journal of psychology*, pp. 229–242.

Smith, Steven M and A. K. Sinha (1987). *Effects of brief immersion in a flotation tank on memory and cognition*. Tech. rep. CSCS-004. College Station, TX: Texas A&M University, Committee for the Study of Cognitive Science.

Smith, Steven M and Edward Vela (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology* 6.2, pp. 125–139.

— (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review* 8.2, pp. 203–220.

Squire, Larry R (1987). *Memory and brain.* Oxford University Press.

— (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory* 82.3, pp. 171–177.

Squire, Larry R and Stuart M Zola (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences* 93.24, pp. 13515–13522.

Stefanowitsch, Anatol (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19.3, pp. 513–531.

Stefanowitsch, Anatol and Stefan Th Gries (2003). Collostructions: Investigating the interaction of words and constructions. *International journal of corpus linguistics* 8.2, pp. 209–243.

Theakston, Anna L, Elena VM Lieven, Julian M Pine, and Caroline F Rowland (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language* 31.1, pp. 61–100.

Thompson, Sandra A (1988). A discourse approach to the cross-linguistic category "adjective". In: *Explaining language universals*. Ed. by John A. Hawkins. Basil Blackwell, pp. 167–85.

— (1998). A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. In: *Case, typology and grammar: In honor of Barry J. Blake*. Ed. by Anna Siewierska and Jae Jung Song, pp. 309–341.

Tomasello, Michael (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

— ed. (1998). *The new psychology of language: Cognitive and functional approaches to language structure*. Psychology Press.

— (2000). First steps toward a usage-based theory of language acquisition. *Cognitive linguistics* 11.1/2, pp. 61–82.

— (2009). *The cultural origins of human cognition*. Harvard University Press.

Tulving, Endel and Daniel L Schacter (1990). Priming and human memory systems. *Science* 247.4940, pp. 301–306.

Ullman, Michael T (2001). The declarative/procedural model of lexicon and grammar. *Journal of psycholinguistic research* 30.1, pp. 37–69.

Weiner, E Judith and William Labov (1983). Constraints on the agentless passive. *Journal of linguistics* 19.1, pp. 29–58.

Yu, Chen (2008). A statistical associative account of vocabulary growth in early word learning. *Language learning and Development* 4.1, pp. 32–62.

Yu, Chen, Dana H Ballard, and Richard N Aslin (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science* 29.6, pp. 961–1005.

Yu, Chen and Linda B Smith (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18.5, pp. 414–420.

Zwaan, Rolf A, Robert A Stanfield, and Richard H Yaxley (2002). Language comprehenders mentally represent the shapes of objects. *Psychological science* 13.2, pp. 168–171.