UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Generative and Item-Specific Knowledge of Language

Permalink
https://escholarship.org/uc/item/8qm7n5wz

Author
Morgan, Emily

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Generative and Item-Specific Knowledge of Language

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

in

Linguistics and Cognitive Science

by

Emily Ida Popper Morgan

Committee in charge:

Professor Roger Levy, Chair
Professor Sarah Creel
Professor Grant Goodall
Professor Andrew Kehler
Professor Marta Kutas

2016

Copyright
Emily Ida Popper Morgan, 2016
All rights reserved.

The Dissertation of Emily Ida Popper Morgan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, thank you to Roger Levy, for your advice and support throughout my grad school experience. It goes without saying that this dissertation has benefitted hugely from your insight, but furthermore, you've guided me in becoming a fully fledged researcher. It's a privilege to work with someone who has both such expertise and such a desire to make sure that I succeed.

Thank you to the rest of my committee—Sarah Creel, Grant Goodall, Andy Kehler, and Marta Kutas—for being willing to ask me tough questions. Your thoughtful feedback and diverse perspectives have made this dissertation better.

This dissertation would not have been possible without the help of many undergraduate research assistants: Mike Brooks, Hannah Campbell, Miranda Clemmons, and especially Wednesday Bushong and Melody Yen, who coded the majority of the corpus. Thank you all.

To all the CPL lab members past and present, including Klinton Bicknell, Rebecca Colavin, Gabe Doyle, Victoria Fossum, Mark Myslín, Bożena Pająk, Gary Patterson, Till Poppels, Titus von der Malsburg, Eva Wittenberg, and Meilin Zhan: You guys are the ones who motivate me to come to work every day. Thank you for helping me navigate grad school, for reading last minute drafts of papers, for hours spent commiserating in coffee shops, and for laughing at bad linguistics jokes and inadvertently hilarious R plots.

Savi and Ryan, thank you for loaning me your parking spot. Sometimes the little things make all the difference.

Thank you to my parents for your unwavering love and support.

And finally, thank you to my Unlab—Cat, Mark, and Cedric—for always being there for me. We **all** made it!

Chapter 2, in full, contains material being prepared for publication in Morgan, E., & Levy, R. (2015). Abstract knowledge versus direct experience in processing of binomial expressions. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Morgan, E., & Levy, R. (2015). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), 37th Annual Meeting of the Cognitive Science Society (pp. 1649-1654). Austin, TX: Cognitive Science Society. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, contains material being prepared for publication in Morgan, E., & Levy, R. (2016). Frequency-dependent regularization in iterated learning. The 11th International Conference on the Evolution of Language. The dissertation author was the primary investigator and author of this paper.

The research presented here was funded by research grants NSF 0953870 and NICHD R01HD065829 and fellowships from the Alfred P. Sloan Foundation and the Center for Advanced Study in the Behavioral Sciences to Roger Levy, and NIH T32-DC000041 to the UCSD Center for Research in Language.

VITA

2007	B.A. in Mathematics *cum laude*, Harvard University
2009	M.Sc. in Brain and Cognitive Sciences *cum laude*, University of Amsterdam
2015	Ph.D. in Linguistics and Cognitive Science, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Generative and Item-Specific Knowledge of Language

by

Emily Ida Popper Morgan

Doctor of Philosophy in Linguistics and Cognitive Science

University of California, San Diego, 2016

Professor Roger Levy, Chair

The ability to generate novel utterances compositionally using generative knowledge is a hallmark property of human language. At the same time, languages contain non-compositional or idiosyncratic items, such as irregular verbs, idioms, etc. This dissertation asks how and why language achieves a balance between these two systems—generative and item-specific—from both the synchronic and diachronic perspectives.

Specifically, I focus on the case of *binomial expressions* of the form "X and Y", whose word order preferences (e.g. *bread and butter*/#*butter and bread*) are

potentially determined by both generative and item-specific knowledge. I show that ordering preferences for these expressions indeed arise in part from violable generative constraints on the phonological, semantic, and lexical properties of the constituent words, but that expressions also have their own idiosyncratic preferences. I argue that both the way these preferences manifest diachronically and the way they are processed synchronically is constrained by the fact that speakers have *finite experience* with any given expression: in other words, the ability to learn and transmit idiosyncratic preferences for an expression is constrained by how frequently it is used. The finiteness of the input leads to a rational solution in which processing of these expression relies gradiently upon both generative and item-specific knowledge as a function of expression frequency, with lower frequency items primarily recruiting generative knowledge and higher frequency items relying more upon item-specific knowledge. This gradient processing in turn combines with the bottleneck effect of cultural transmission to perpetuate across generations a frequency-dependent balance of compositionality and idiosyncrasy in the language, in which higher frequency expressions are gradiently more idiosyncratic. I provide evidence for this gradient, frequency-dependent trade-off of generativity and item-specificity in both language processing and language structure using behavioral experiments, corpus data, and computational modeling.

Chapter 1

Introduction

1.1 Storage and computation

To what extent are complex linguistic representations *computed* on the fly using generative linguistic knowledge and to what extent are they *stored* (i.e. memorized as chunks and reused holistically)? For example, when we encounter an inflected verb such as *walked*, do we process it by decomposing it into its component morphemes *walk* and -*ed*, or do we recognize it holistically in its inflected form? Likewise, is a common expression such as *bread and butter* always generated in real time from its component words, or can it be recognized holistically as a known expression?

From the perspective of describing language structure, we can ask a comparable set of questions: to what extent are the structures we find in natural language data *compositional* (able to be described as composed from smaller units via generative principles) and to what extent are they *idiosyncratic* (unpredictable from generative paradigms)? For example, should we think of *bread and butter* as merely the compositional conjunction of two common nouns, or does its high frequency and sometimes metaphorical use grant it special status as a known collocation? Similarly, does the extreme phonetic reduction of a high frequency phrase like *I don't know* indicate that it requires special distinction in our theories of language structure, or is it a purely compositional utterance from the perspective of structural description, which just happens to have some special attributes that manifest in online processing?

This dissertation takes up the above questions, asking to what extent language processing relies on *generative* versus *item-specific* knowledge, to what extent language structure is *compositional* versus *idiosyncratic*, and how these two levels of description (language processing and language structure) are mutually constraining.

1.1.1 Some big questions

Big questions in language processing

To what extent are complex linguistic representations computed on the fly using generative knowledge, and to what extent do they rely on pre-computed and stored item-specific representations? Some amount of both computation and storage clearly exists in language processing: the ability to generate novel utterances compositionally is unquestionably a hallmark of human linguistic competence, while some amount of storage is clearly required for, at minimum, monomorphemic lexical items, which are not predictable from any component parts. But beyond these basic facts, theories of language processing vary wildly in how much of language processing is attributed to generative versus item-specific knowledge, from almost-total reliance on generative knowledge (e.g. Hauser et al., 2002) to almost-total reliance on storage and reuse (e.g. Bybee, 2001, 2009), with many in between (e.g. Pollard and Sag, 1994; Pinker, 2000; Goldberg, 2003; O'Donnell et al., 2011, among many others).

In addition to asking, in general, how much these two types of knowledge are recruited in language processing, we can ask what factors influence how much one or the other is recruited in a given instance. In particular, does the frequency of an item influence whether it is processed using generative or item-specific knowledge? We might expect that higher frequency items are more likely to recruit item-specific knowledge than lower frequency items (e.g. compare *walked* to *kayaked,* or *I don't know* to *I do not possess that knowledge*). If frequency is a factor, does reliance upon generative versus item-specific knowledge change categorically at some threshold, or does it vary gradiently? At very high frequencies, might item-specific knowledge come to block the use of generative knowledge altogether, even for items that are in principle entirely compositional, or does generative knowledge always continue to exert a role?

From a functional perspective, we can ask *why* these two types of knowledge might be recruited to different extents. Of course, novel items must be generated compositionally, while idiosyncratic items such as idioms must be processed using item-specific knowledge. But what of non-novel, compositional items (i.e. items which have been previously experienced, which follow the compositional rules)? Does a limit on human memory prevent us from storing everything we have experienced, such that generative knowledge is required even for previously experienced items? Or does a limit on computational speed prevent us from generating everything on the fly, such that more reliance upon stored item-specific knowledge is required? Does the ability to recruit two different types of knowledge provide an ability to hedge one's bets in cases where one isn't certain about how a particular linguistic item ought to be processed? And how must language processing adapt to the structure of the language itself? Does the balance of compositionality and item-specificity within the language itself influence the language processor's recruitment of generative and item-specific knowledge?

Big questions in language structure

Once again, we can ask a comparable set of questions about language structure. If we want to describe the patterns and structures we see in natural language data, to what extent should we attribute to them systematic compositionality, and to what extent should we think of various structures as idiosyncratic? Languages clearly contain a great deal of structure, yet just as clearly contain idiosyncrasies in the form of irregular inflections, idioms, etc. Can we draw a clear distinction between items that behave compositionally versus idiosyncratically, or is there a compositionality-idiosyncrasy spectrum, with some items (perhaps such as *bread and butter* and *I don't know*) lying in the middle?

Once we know how to define idiosyncrasy, what factors influence how likely an item is to behave idiosyncratically? In particular, are higher frequency items more likely to be idiosyncratic? We know that higher frequency verbs are more likely to be irregular (Bybee, 1985; Lieberman et al., 2007); does this relationship hold more broadly, including at other levels of linguistic structure such as multi-word expressions?

Whatever the exact balance of compositionality and idiosyncrasy in the language, how is it preserved over time? We know that individual items can change whether or not they are idiosyncratic: irregular verbs become regular, but new idiosyncrasies also get introduced (such as loan words that violate a language's phonology, or conscious language play resulting in novel constructions like *because* in "because reasons"). How does synchronic language processing interact with processes of cultural transmission to effect language change over time that allows for variation in which items are idiosyncratic, but maintains a structure that includes both compositionality and idiosyncrasy?

Finally, from a functional perspective, we can ask why there are idiosyncrasies in language in the first place. Are they merely hold-over historical accidents (such as stem-changing verbs that follow what were once productive patterns, or expressions such as *hang up the phone* that were once literal but no longer apply in the age of cellphones)? Or do they serve a functional purpose, such as increasing efficiency of communication? For example, a stem-changing irregular verb is the same length in the past tense as in its base form, compared to a regular *-ed* past tense, which will be longer than its base form. The savings in length might be particularly useful for high frequency verbs, which irregular verbs generally are (Bybee, 1985; Lieberman et al., 2007).

These big questions set the stage for the questions I ask in this dissertation. In the remainder of this section, I provide a brief review of previous work and then give my own definitions for key terms that will be used throughout the dissertation. In Section 1.2 of the Introduction, I present the fact that speakers have *finite linguistic experience* and that this motivates a rational solution in which generative and item-specific knowledge trade off gradiently in language processing, which in turns predicts a gradient trade-off between compositionality and idiosyncrasy in language structure. In Section 1.3 I present two test cases, briefly presenting the case of verb inflection before turning to *binomial expressions* (e.g. *bread and butter, salt and pepper*), which are the topic of this dissertation. Finally in Section 1.4, I summarize chapter-by-chapter the main claims made in the dissertation.

1.1.2 A brief history

Beginning with Chomsky's *Syntactic Structures* (1957) and *Aspects* (1965) (see also Hauser et al., 2002), one school of thought in linguistics takes the generative function of language to be **the** hallmark property of language, motivating what Jackendoff (2002) calls the "syntactocentric" approach to linguistics, in which the primary objective is to characterize with as much precision as possible the form of humans' generative linguistic knowledge. In particular, this school holds that when studying syntax and semantics, the goal is to characterize the broadly applicable combinatorial rules and the principles of Universal Grammar that allow us to learn them, while the study of non-compositional structures such as idioms is merely a secondary concern. Likewise, when approaching other levels of linguistic structure such as morphology or phonology, this school still emphasizes the discovery of syntax-like generative rules (see, for example, Everaert et al., 2015). In describing verb inflections, for example, their primary interest lies in characterizing how affixes attach to stems in the "regular" case, while irregular verbs are assumed to be memorized and thus not given much serious attention (except in cases where the so-called "irregularities" are themselves susceptible to a rule-based description; Chomsky and Halle, 1968). Exceptions are consigned to the "periphery" of "phenomena that result from historical accident, dialect mixture, personal idiosyncrasies, and the like" (Chomsky and Lasnik, 1993 as cited in Jackendoff and Pinker, 2005). In other words, under this paradigm, generative computation is given a primary status, with item-specific storage reserved only for those phenomena that cannot be accounted for otherwise.

However, the wisdom of this intense focus on the generative component of linguistic knowledge has long been called into question. In one of the most direct challenges, *exemplar-* or *usage-based* grammatical theories (e.g. Langacker, 1987; Johnson, 1997, 2006; Pierrehumbert, 2000; Bybee, 2001, 2006; Goldberg, 2003; Gahl and Yu, 2006; van den Bosch and Daelemans, 2013) claim that people **store** tokens of linguistic experience at all levels of analysis—from phonemes to multi-word utterances—and that language processing relies upon online generalization from these exemplars. Some of these theories go so far as to avoid abstract generative rules entirely in favor of online generalizations over concrete exemplars (Bybee, 2001, 2009). Others (Langacker, 1987; Pollard and Sag, 1994; Jackendoff, 2002; Goldberg, 2003) take a more moderate position, not excluding abstract rules entirely, but advocating for theories in which the traditional lexicon is replaced with a much broader storage network of constructions of different shapes and sizes (e.g. Goldberg's *construct-i-con*, which contains such diverse elements as words—both monomorphemic and complex, idioms, grammatical constructions such as the double object construction, and more.) What these theories all have in common is their emphasis on the storage and direct reuse of item-specific representations as a central element of the grammar.

For structures which could potentially be either stored or generated (e.g. inflected regular verbs), the primary diagnostic used to argue whether a given structure is stored or not is whether its frequency of occurrence is predictive of its behavior in language processing or language change. Such frequency effects are well documented at the level of individual words: more frequent words are faster to read (Inhoff and Rayner, 1986; Rayner and Duffy, 1986; Rayner et al., 1996), more likely to be skipped in reading (Rayner et al., 1996; Rayner and Well, 1996), and more susceptible to phonetic reduction (Bybee, 1999; Gregory et al., 1999). Analogously, if other structures, such as multimorphemic words or multi-word expressions, exhibit frequency effects—even when the frequency of their component parts is controlled for—this is taken as evidence for the existence of a chunked mental representation of the structure as a whole. Under this diagnostic, exemplar-based models have received support from demonstrations of frequency effects in online processing of inflected verbs (see Section 1.3.1); in online processing of multi-word expressions (Arnon and Snider, 2010); and in corpus-based measures such as phonetic reduction of multi-word expressions (Bybee, 2006).

A conceptually similar (although architecturally distinct) challenge to the Chomskyan syntactocentric approach comes from the connectionist modeling tradition (Rumelhart and McClelland, 1986; Hare et al., 1995; Elman, 2003): in eschewing abstract symbolic rules, connectionist models instead emphasize how online generalizations are directly shaped by one's previous linguistic experience. One of the most famous test cases for the applicability of connectionist models to language modeling has also been one of the mostly hotly debated test cases regarding storage and computation: regular and irregular verb inflection. This debate will be explored in more detail in Section 1.3.1. In short, Rumelhart and McClelland's claim that a single architecture could both apply a default (regular) rule and learn its exceptions poses a direct challenge to the traditional Chomskyan framework, in which language is characterized in terms of a relatively small and parsimonious set of generative rules and lexical items.

A final challenge to the Chomskyan tradition comes from the literature on *prefabricated* or *formulaic language* in the tradition of Sinclair (1991) (see also Barnbrook, 2007). This tradition, much like the exemplar-based approaches, emphasizes the need for including larger-than-single-word units in analyses of language. Based largely on an analysis of corpus data, Sinclair (1991) posits that the choice of each word in language production follows one of two principles: the "principle of open choice," corresponding to insertion of words following standard generative principles, or the "principle of idiom," corresponding to the more constrained selection of a word as part of a known chunk or *prefab.* Sinclair argues that the principle of idiom is used much more prevalently than assumed in traditional Chomskyan analyses, with Erman and Warren (2000) estimating that more than 50% of words (in both spoken and written texts) are parts of prefabs rather than being freely chosen. Wray (2008) takes this claim even further, arguing that the principle of idiom is the default mode of processing, with the principle of open choice (i.e. standard generative analysis) used only when no analysis via prefabricated chunks is possible.

Similar to proponents of exemplar-based theories, scholars in this tradition have emphasized how wide a scope one can cast when looking for prefabricated chunks in language, including not just entirely non-compositional examples such as idioms, but any case in which a collocation occurs more frequently than would be expected under a purely generative theory, including verb-preposition and adjective-preposition collocations (e.g. *dealing with, suitable for*), discourse markers (e.g. *for instance, in the end*), and colloquial expressions (e.g. *I can't see a thing, get the hang of*) (Erman and Warren, 2000). However, their insistence upon a binary classification of all word choices into one of two principles disallows the possibility of acknowledging cases such as verb-preposition collocations that on the one hand clearly exhibit compositional structure, but on the other hand are clearly restricted by known stochastic relationships between words. This disadvantage is highlighted by Wray (2002) as she reviews proposed diagnostics for, or ways to automatically detect, prefabs. Despite considering an impressively wide range of options—from native speaker intuition to corpus frequency measures to psycholinguistic diagnostics such as pauses in spontaneous speech or patterns of code-switching—she finds them all lacking due to the wide range of what potentially counts as a prefab: each of the measures she considers draws different black-and-white boundaries through the fuzzy grey space of potentially-compositional-potentially-prefabricated constructions.

In short, there remains significant room for debate over the roles of, and the boundaries between, storage and computation in language. In order to ask these questions more precisely, I introduce some definitions.

1.1.3 Definitions

Language processing: generative and item-specific knowledge

To ask questions about language processing, I follow the previous literature rather closely in drawing a distinction between two types of knowledge potentially used in processing. On the one hand is knowledge that allows language users to generate composite linguistic units from their component parts (e.g. expressions from words, or words from morphemes). I refer to this knowledge as *generative* or *compositional*, or sometimes—when I want to highlight the potential use of non-linguistic real-world knowledge as well—*abstract*. This knowledge is defined by the fact that it is not specific to particular lexical items, but rather is generally applicable: in particular, it can apply to novel items.[1] Such knowledge may take the form of categorical rules, such as the compositional rules used in traditional syntactic analysis, or alternately may take the form of violable constraints or probabilistic preferences. For example,

[1] Such generative knowledge could have been learned and generalized from particular lexical items originally, but we refer here to knowledge that, in an adult grammar, applies broadly to a class or classes of lexical items.

in the case of the dative alternation, violable constraints referencing the definiteness, animacy, etc. of the two arguments probabilistically determine preferences for use of one alternate over the other (Bresnan et al., 2007). As these violable constraints apply to never-before-seen utterances, they too constitute a form of generative knowledge.

The other type of knowledge language users can draw upon in processing is *item-specific* knowledge learned from *direct experience* with the item in question. In particular, processing of frequent utterances may employ stored holistic knowledge of complex linguistic units, such as multi-word expressions or multimorphemic words. Such knowledge is most evident in the case of idiosyncratic items that categorically violate generative rules, such as irregular verbs, which violate inflectional rules, or idioms, which violate the usual rules of semantic composition. But item-specific knowledge need not only apply to items that deviate from what generative knowledge would predict. Even an expression that is fully compositional could in principle be processed using item-specific rather than generative knowledge, if it has been previously experienced and stored. For example, Bybee (2006) argues that the extreme phonetic reduction of the expression *I don't know* is evidence of its holistic storage and reuse, despite the fact that it seemingly follows the generative rules of English and could in principle be processed via generative knowledge.

The distinction between generative and item-specific knowledge in language processing is an empirical one, in that we can ask empirical questions about how much processing relies upon these two knowledge sources; in fact even the existence of each type of knowledge is in principle an empirical matter (although the existence of at least *some* storage and *some* computation is uncontroversial).

Language structure: compositionality and idiosyncrasy

In order to discuss which linguistic items might be more amenable to processing via either generative or item-specific knowledge, I introduce a related duality in language structure. Wray's (2002) difficulty in finding a satisfactory empirical diagnostic for prefabricated language suggests we take a somewhat different approach, so I will instead draw a **theoretical** distinction: given a theory of the generative structure of language, I will identify utterances that conform to the theory as *compositional* and utterances that aren't predicted by the theory as *idiosyncratic*. This distinction sidesteps Wray's issue with the heterogeneity of prefabricated language, because our idiosyncrasies can be as heterogeneous as our generative linguistic theories. In other words, for anything for which we can develop a generative theory, we can likewise identify the idiosyncratic exceptions. As a simple example, if we have a theory of English past tense formation that consists of adding *-ed*, then we can identify idiosyncratic verbs as those that don't follow this rule. Or if we have a (presumably more elaborate) theory of syntactic and semantic combinatorial rules, then we can identify idioms as those expressions that violate the compositional principles.

A further benefit of this definition is that under this theory, if our compositional theory is probabilistic or gradient, then our idiosyncrasies can likewise be graded. For example, returning to the case of the dative alternation, we find that even after the violable constraints on definiteness, animacy, and so on have been taken into account, the verb *tell* is disproportionately likely to take a double object construction, while *deliver* is disproportionately likely to take a prepositional dative (Baayen, 2011). We can attribute gradient idiosyncrasy to these verbs to the extent that they deviate from our probabilistic predictions.

Foreshadowing what's to come, one might suppose that, to the extent that a linguistic item is compositional, it is amenable to processing via generative knowledge (although, as mentioned previously, processing via item-specific knowledge is also possible), and to the extent that an item is idiosyncratic, processing via item-specific knowledge will be preferred.

We have thus set up an empirical distinction between generative and item-specific knowledge in language processing, and a related theoretical distinction between compositionality and idiosyncrasy in language structure. This way of setting up our definitions is logical given that language processing within an individual speaker is itself an empirical phenomenon, whereas when we talk about language structure, we are making a generalization about a community of speakers and their linguistic behavior.

1.2 The finiteness of the input

What factors might influence the balance of generative and item-specific knowledge in language processing, or the degree of compositionality and idiosyncrasy in language structure? We first introduce the current predominant explanation—a speed-memory trade-off—then propose a novel explanation centered on the finiteness of one's linguistic experience.

1.2.1 A speed–memory trade-off

In the domain of language processing, the trade-off between generative and item-specific knowledge has generally been conceived as a speed-memory trade-off (Nooteboom et al., 2002; Wiechmann et al., 2013). On the one hand, performing compositional operations takes time, and so the more computations are required, the longer processing will take; conversely, if larger units can be precomputed and stored, less computational time will be required, at the expense of more memory dedicated to increase storage capacity. Under the assumption of this guiding principle, the degree to which processing relies upon generative and item-specific knowledge should be motivated by both the brain's online processing speed and its long-term storage capacity. While computational speed and finite memory are both real limitations on human language processing, the precise bounds of each are not known. Thus while the idea of a speed-memory trade-off can predict that there should be a trade-off between generative and item-specific knowledge in language processing, it is difficult to form more specific predictions without knowing more about humans' speed and memory limitations.

1.2.2 The finiteness of the input

Instead of conceiving of generative and item-specific knowledge as instantiating a speed–memory trade-off, I propose that we view them in terms of a different, environmental limitation: *the finiteness of the input.* Even if we completely ignore limitations on both storage and processing speed, people are exposed to a finite amount of linguistic input in their lifetimes. This fact has long been used to motivate the need for generative knowledge to produce and comprehend never-before-seen utterances. I extend this argument further, arguing that the fact that we have *differential* amounts of experience with different linguistic items motivates a *gradient trade-off* between generative and item-specific knowledge in language processing, and between compositionality and idiosyncrasy in language structure.

In considering the domain of language processing, let us take the goal of the processor to be coming up with the best possible form-meaning mapping, where either form or meaning is given, depending on whether the current task is comprehension or production: in other words, in comprehension, the goal of the processor is to determine the meaning that was most likely intended by the given linguistic form, and in production, the goal is to determine a form that will mostly likely convey the intended meaning. Of course, one of the hallmarks of human linguistic competence is the ability to generate novel form-meaning mappings from generative knowledge, and so using this strategy is one possibility. But what of cases when one has used a particular form-meaning mapping before (e.g. conjugating a known verb like *walked* or constructing a ditransitive utterance like *Pass me the salt* that one has direct previous experience with)? Given no limitations on processing time, the processor could simply compute a new form-meaning mapping from scratch using generative knowledge, regardless of previous experience. At the same time, given unlimited storage, the processor could store all previously used mappings, and thus could potentially reuse any previous mapping directly, with no reference to generative knowledge. However, I argue that, given the finiteness of the input, neither of these solutions is optimal.

In particular, the finiteness of the input means that language users can never be 100% certain about the form-meaning mappings they have learned, either on the basis of generative or item-specific knowledge. Given this uncertainty, a gradient trade-off is the optimal way to estimate these mappings. To rely solely on generative knowledge would not be rational as long as one believes the language contains some idiosyncrasies, as reliance solely upon generative knowledge would preclude one from ever using irregular verbs, idioms, or any other idiosyncratic items. Thus if one has direct evidence for a given form-meaning mapping, regardless of whether it agrees with generative knowledge or not, it is rational to take this direct evidence into account. At the same time, relying *solely* on a previously stored mapping for which one has finite previous evidence is also not rational, when one could additionally be drawing upon generative knowledge as a method of hedging one's bets. Intuitively, if one has used a form-meaning mapping only once before, one should not trust it entirely: it could have been produced as a result of an error, or only due to some unknown and unusual property of the context, or due to other unappreciated circumstances. Its previous use provides some evidence of its suitability for future use, but the evidence is relatively weak. A second, third, and so on use of this mapping would provide increasingly strong evidence for its appropriate use. The rational solution, then, is to rely gradiently upon both generative and item-specific knowledge as a function of how often one has previously experienced the form-meaning mapping in question, relying increasingly upon item-specific knowledge as the frequency of one's direct experience increases. And this, I will argue, is indeed what the language processor does. (Other theories that make convergent predictions are discussed in Chapter 2.) This claim will be made more precise using the language of probability theory in Section 1.2.3.

The finiteness of the input, and the resulting gradient trade-off between generative and item-specific knowledge in language processing, likewise predicts a balance of compositionality and idiosyncrasy in language structure. For infrequent items—items for which direct evidence for a specific form-meaning mapping is limited—each production must rely primarily upon generative knowledge. Thus infrequent items must by and large conform to the compositional structure of the language. The more frequent an item, the more people's productions can rely upon their previous direct experience with it. Thus these productions have more ability to deviate from generative knowledge, and, crucially, it is possible for such deviations to be transmitted reliably across generations so that they can persist in the language. The finiteness of the input thus also predicts a balance between compositionality and idiosyncrasy in language structure, with higher frequency items more likely to be idiosyncratic. I will demonstrate that this prediction is confirmed in corpus data.

1.2.3 Rational Analysis and probability theory

The above line of argumentation falls naturally within the framework of Rational Analysis (Anderson, 1990): rather than focusing on assumed limitations of cognition (such as computational speed or memory), we instead focus on the constraints of the environment (specifically, the finite input), and ask what a rational solution to the problem is given these constraints. We can use probability theory to define the *rational* solution as the solution that optimizes a desired outcome measure—in this case, the probability of choosing the correct form for a given meaning, or vice versa. Under this paradigm, one benefit to focusing on the environmental constraint of finite input (rather than the assumed cognitive limitations of speed and memory) is that we already have reasonable estimates of the amount of linguistic exposure people have in their lifetimes, both in terms of total word counts (e.g. Levy et al., 2012) and in terms of the relative frequencies of different words and structures as seen in large corpora (e.g. Marcus et al., 1999; Lin et al., 2012), whereas cognitive limitations on computational speed and memory are much more difficult to quantify.

The language of probability theory gives us a way to express this rational argument more precisely. In particular, we turn to Bayes' Rule, which relates the probability of data to the probability of hypotheses:

$$P(h|d) \propto P(d|h)P(h) \tag{1.1}$$

This rule tells us how to evaluate the probability of a hypothesis h given some data d. In particular, suppose we have a variety of hypotheses h for how to map between linguistic forms and meanings, and we need to choose between them on the basis of our previous linguistic experience d. (For example, for a given verb, one hypothesis might be that the verb has a regular past tense, while another hypothesis might consist of an irregular past tense form. Alternately, hypotheses might be gradient: for a ditransitive verb, a hypothesis might indicate how often it should appear in a double object versus a prepositional dative construction.) Intuitively, to choose between hypotheses, we want to compare the probability of different hypotheses given the data and choose a hypothesis with a high probability. Bayes' Rule tells us that this probability $P(h|d)$ is proportional to the product of the *data likelihood* $P(d|h)$ and the *prior probability* of the hypothesis $P(h)$.

The data likelihood $P(d|h)$ tells us how likely the known data d is to occur under the hypothesis h. Intuitively, for us to consider a hypothesis plausible, it should be a hypothesis that gives the known data a high likelihood. (For example, if one's hypothesis is that a given verb is regular, then occurrences of the predicted regular form will have high likelihood, whereas occurrences of any irregular form will have low likelihood.)

The prior probability $P(h)$ reflects our belief in how likely a given hypothesis is before any data has been observed. Where might we get priors on our linguistic hypotheses? If we consider hypotheses at a fine level of granularity (in particular, preferences for individual linguistic items, such as individual verbs or expressions), *generative knowledge* is the natural source of prior beliefs, as this is the type of knowledge we bring to bear on processing linguistic items with which we have no previous experience.

In other words, if we take generative knowledge as providing prior probabilities on hypotheses, and our previous linguistic experience as providing data, then Bayes' rule tells us how to optimally combine these two knowledge sources in order to put a probability distribution on (and hence choose between) hypotheses about how an item is used in the language.

When we compare across hypotheses, how strongly $P(d|h)$ favors one hypothesis over another will be determined in part by how much data we have. If we have relatively little data, the data likelihood across different hypotheses will be not as strongly differentiated as if we have a large amount of data. To see this intuitively, imagine trying to decide if a coin is fair by flipping it repeatedly: i.e. we are trying to decide between hypotheses $h \in [0, 1]$ for how often a coin comes up heads on the basis of coin flip data d, with $h = 0.5$ being a perfectly fair coin. If in four flips, we obtain two heads and two tails, this small amount of data provides relatively weak evidence: many possible coin weightings h could have produced this data with reasonably high probability. (See Figure 1.1.) But if in ten flips we obtain five heads and five tails—or better yet, if in one hundred flips we obtain fifty heads and fifty tails—these increasingly large amounts of data provide increasingly strong evidence that the coin is fair, as the probability of generating this data under hypotheses far from 0.5 drops off more and more steeply. Likewise, for an unfair coin, more data would provide stronger evidence for $h \neq 0.5$.

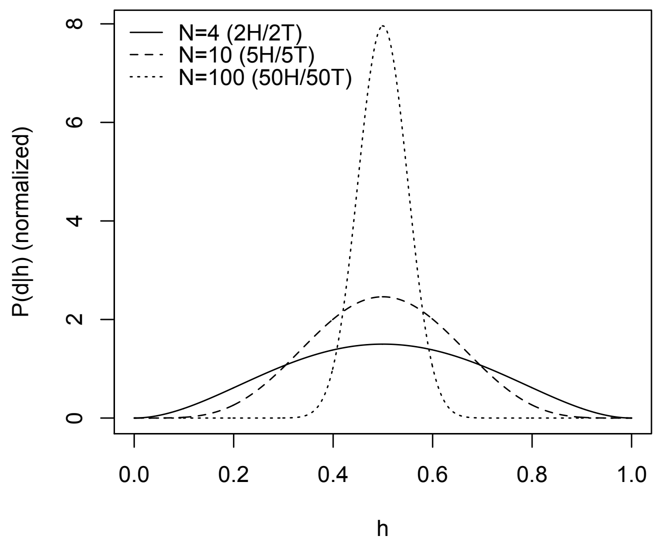


Figure 1.1: Data likelihoods $P(d|h)$ become more diagnostic with more data.

In contrast, the prior $P(h)$ is independent of the data, and therefore how strongly it favors one hypothesis over another does not change depending on the amount of data. Since $P(d|h)$ becomes more diagnostic with increasing data, and $P(h)$ does not change, the end result is that the more data one has, the more strongly $P(h|d)$ is determined by the data likelihood. (In fact, with arbitrarily large amounts of data, the influence of the prior becomes vanishingly small.) Translating this back into the terms described above, we see that Bayes' rule makes the same prediction as we made intuitively earlier: the lower an item's frequency, the more its linguistic behavior should be determined by generative knowledge; the higher an item's frequency, the more its behavior should be determined by item-specific experience.

1.2.4 A chicken-and-egg proposal

The theory proposed above has a chicken-and-egg-like structure: it is rational for language processing to rely gradiently upon generative and item-specific knowledge as a function of item frequency **if** languages contain idiosyncrasies, and language is predicted to contain frequency-dependent idiosyncrasies **if** processing relies upon both generative and item-specific knowledge. This circular structure is of course reflective of the fact that language processing and language structure truly are mutually constraining: language processing must accommodate the existing language structure, but language processing is also the driving mechanism of language change and is therefore what ultimately shapes language structure.

1.3 Test cases

I describe two test cases for the questions raised above. First, I take up the question of regular and irregular verb inflection, one of the most famous test cases for generativity versus item-specificity in language processing as well as a famous case of frequency-dependent idiosyncrasy in language structure. While reviewing results that are convergent with the predictions I make above, I also note ways in which this case does not provide the strongest test for some of these claims. I then turn to the case of *binomial expressions* of the form “X and Y” (e.g. *bread and butter, salt and pepper*), which is the test case taken up for the remainder of this dissertation.

1.3.1 Regular and irregular verb inflection

Regular and irregular verbs in language processing

What are the roles of generative and item-specific knowledge in the processing of inflected regular and irregular verbs? Inflected irregular verbs (e.g. *sang, wrote*) must be processed using item-specific knowledge, as generative knowledge does not apply.[2] But what of inflected regular verbs (e.g. *walked, laughed*)? Such verbs could be processed using generative knowledge, but could also in principle be processed using item-specific knowledge—in other words, the inflected form could be stored and retrieved holistically.

As described in Section 1.1.2, the primary diagnostic for whether a given item is stored and reused directly in processing is whether it exhibits frequency effects specific to its composite form, independent of frequency effects for its component parts. In the case of regular verbs, if inflected forms exhibit effects of the frequency *of the inflected form* (as opposed to of the stem)—called Surface Frequency Effects—this is evidence that the inflected forms themselves are being stored. In contrast, if regular forms are always computed generatively, we would expect to find effects of frequency of the stem, or Base Frequency Effects, reflecting the effect of retrieval of this component of the generated form, but we would predict that the further step of computing the inflected form takes the same amount of time regardless of verb stem frequency. Both Surface and Base Frequency Effects have been demonstrated for regular English (Taft, 1979) and Italian verbs (Burani et al., 1984)—as well as for regular inflected nouns in a variety of languages (Baayen et al., 1997, 2002; Sereno and Jongman, 1997; Alegre and

[2] Although some subregularities hold within the so-called irregular verbs, most theories still assume item-specific knowledge is required to know exactly which verbs these subregularites apply to, e.g. *bring/brought, think/thought*, but *drink/*drought* (Pinker, 2000).

Gordon, 1999; Bertram et al., 2000)—suggesting that both generative and item-specific knowledge play a role in processing regular inflected verbs.

The case of inflected regular verb processing thus confirms the basic prediction that both generative and item-specific knowledge play a role in processing, even for forms that are in principle compositional. However, it remains unknown whether and how this trade-off is affected by verb frequency. Further, we note that verb inflection is not the strongest test of item-specific storage because, particularly in English, the number of possible inflections for any given verb is relatively small. Storing all possible inflected forms for a verb may have a small multiplicative effect on the number of verb forms one must store. In contrast, when one considers storing possible multi-word expressions of varying lengths, as proposed by exemplar- and usage-based theories, the item-specific storage requirements grow polynomially. Thus for a stronger test of the storage claims made by these theories, we should consider a test case at the phrasal rather than the single word level.

Regular and irregular verbs in language structure

How are irregular verbs distributed in language structure? Lieberman et al. (2007) demonstrate using a historical corpus that higher frequency verbs are more likely to be irregular and, moreover, that the higher their frequency, the more likely they are to stay irregular for longer (rather than switching to the regular *-ed* pattern). Their findings confirm the prediction that higher frequency items are more likely to be idiosyncratic. (See also Bybee, 1985.)

How is a balance of compositionality and idiosyncrasy in verb inflection preserved over time? Lieberman et al. address the process by which verbs switch from being irregular to being regular, but not the reverse; they predict that more and more English verbs will become regular over time, with new irregularities presumably only emerging if the default rule changes in the future, leaving some of our currently-regular verbs behind in its wake to become irregular in the new paradigm. While it may be true of verb inflection in particular that new irregular forms do not arise spontaneously without major changes in the default paradigm(s), this fact does not generalize to the language as a whole. As discussed in Section 1.1.1, new idiosyncrasies are constantly being introduced into the language. Verb inflection is thus limited in its ability to serve as a test case for the dynamic preservation of a balance of compositionality and idiosyncrasy.

A further downside of verb inflection as a test case for idiosyncrasies is that English verbs are, to a first approximation, categorically regular or irregular, limiting our ability to test for the possibility of gradient idiosyncrasies as a function of item frequency.

1.3.2 Binomial expressions

To rectify some of the limitations discussed above, we turn to a test case involving multi-word expressions with gradient properties. Specifically, in this dissertation I will take up the test case of word order preferences in *binomial expressions* of the form "X and Y". Ordering preferences for these expressions are gradient: for example, "radio and television" is preferred to "television and radio" in a 63 to 37 ratio, while "bread and butter" is preferred to "butter and bread" 99 to 1 (Lin et al., 2012). We can ask what factors influence binomial ordering preferences from both a synchronic and diachronic perspective. In particular, one possibility is that preferences arise from violable generative constraints that reference phonological, semantic, or other lexical properties of the elements in a binomial (e.g. the shorter word should come first). An alternate possibility is that preferences are driven by direct experience with the specific binomials in question: an order is preferred because it has been experienced more often. Thus in the domain of language processing, we can ask whether ordering preferences in online processing are driven by generative knowledge of ordering constraints or by item-specific direct experience. Likewise in the domain of language structure, we can ask whether the distribution of binomial preferences seen in the language as a whole is shaped by generative constraints versus to what extent it exhibits item-specific idiosyncrasies. Moreover, in both processing and structure, we can ask whether the influence of the two knowledge sources changes as a function of the frequency of an expression.

To ask these questions, there are three properties of binomials we crucially must be able to quantify:

- For a word pair (A, B), the first property we consider is the *overall (unordered) frequency* of binomial expressions containing these elements—in other words, the combined frequency of the expressions *"A and B"* and *"B and A"*. To estimate the overall frequency of people's experience with these expressions, we can obtain frequency estimates from large corpora (generally measured in occurrences per million words).

- Next we consider the *relative frequency* or *observed preference* of a given order. The relative frequency of *"A and B"* is the number of occurrences of *"A and B"* divided by the overall frequency of (A, B) binomial expressions. It is thus a real number between 0 and 1, inclusive. The relative frequency of *"B and A"* is one minus the relative frequency of *"A and B"*. Again, we can estimate this from corpus frequencies.

- Finally, we consider the *compositional* or *abstract-knowledge-based preference* for a given order, i.e. the preference derived purely from people's generative knowledge of binomial ordering constraints. For a given order *"A and B"*, we want a value between 0 and 1 corresponding to the probability of someone producing that order based on their knowledge of the compositional constraints governing binomial ordering, independent of their actual experience with a given item. Unlike the previous two variables, we cannot directly estimate people's abstract knowledge from corpus frequencies. Instead, we will build probabilistic models to give us these estimates.

Of these variables, the two that directly compete to explain binomial ordering preferences in online processing are relative frequency and compositional preference (although we predict that their weights will change as a function of overall frequency). Crucially, although these two variables may be correlated, we assume that they are not equivalent, as relative frequency can be influenced by factors beyond compositional knowledge such as conventionalization and idiomaticity, famous quotations, or language change that interacts with abstract ordering constraints (e.g. changes in word meaning or pronunciation). I discuss this more below, and throughout the dissertation.

Why binomial expressions?

Binomial expressions are a good test case for the interaction of generativity and item-specificity because they are at once theoretically interesting as well as computationally and experimentally tractable. As discussed above, binomial expressions provide a test case concerning gradient preferences in multi-word expressions. In particular, not only can we draw a theoretical dissociation between compositional preferences, relative frequency, and overall frequency, but we can in fact obtain reliable estimates for each factor independently, through a combination of statistical modeling and modern corpora, allowing us to both ask and answer questions about how these factors interact in determining binomial ordering preferences. Moreover, because binomial ordering is a binary choice, we have a variety of well-understood, tractable statistical models (in particular, logistic regression and beta-binomial models) at our disposal to model their preferences. From an experimental perspective, these expressions are convenient because we can manipulate their ordering ("X and Y" versus "Y and X") without changing the formal syntactic or semantic properties of the expression, providing an experimental manipulation that minimizes confounds. Finally, binomial expressions achieve all these aims while being a common construction in English (and other languages). We thus have plentiful naturalistic data about their use, and we can use them in experimental materials without them standing out as unusual.

Predicting distributions

As I will argue over the course of this dissertation, binomial expression preferences are determined both by generative knowledge and by item-specific idiosyncrasies. In particular, while expressions on average conform to compositional preferences, they also acquire idiosyncrasies as a function of their frequency, with more frequent expressions being more idiosyncratic. While some aspects of these idiosyncrasies are predictable (e.g. as we will show, higher frequency expressions have more polarized preferences), some idiosyncrasies are due to unpredictable quirks of history. For example, *men and mice* would be compositionally preferred to *mice and men*—as described in Section 2.2, a Perceptual Markedness constraint favors putting humans before non-humans—but *mice and men* is much more frequently attested due to its use in a Robert Burns poem, a Steinbeck novel, and other popular media (Burns, 1785; Steinbeck, 1937). We don't expect or even particularly want our linguistic theories to predict that this exact binomial (rather than *penguins and people*, *wombats and women*, etc.) would be used in a famous poem. However, we *do* want our theories to predict probabilistic relationships between frequency and idiosyncrasy, such as how likely an expression of a given frequency is to become frozen opposite its compositionally preferred direction.

For this reason, in the parts of this dissertation that focus on language structure and change, our focus will be on predicting *distributions* of binomial expression preferences, rather than on predicting the exact preferences of individual items. While we cannot reasonably expect to predict all the idiosyncratic quirks of individual items, we can aim to predict how these idiosyncrasies will behave in aggregate over a large number of expressions.

1.4 Summary of main claims

In this dissertation, I test the rational theory laid out in Section 1.2 using the test case of binomial expressions. Specifically, I argue as follows:

In Chapter 2, I ask how generative and item-specific knowledge contribute to the processing of both novel and highly frequent binomial expressions. I demonstrate that while processing of novel expressions relies upon generative knowledge, processing of highly frequent expressions relies in large part upon item-specific knowledge—even though these items are in principle fully compositional. I thus show that item-specific knowledge has a larger role to play in language processing than has often been assumed, and moreover that reliance upon the two types of knowledge varies as a function of expression frequency. I propose a more specific form for this trade-off, namely that reliance upon generative versus item-specific knowledge varies *gradiently* with expression frequency. (This proposal will be tested in Chapter 6.)

Why would processing of binomial expressions rely upon item-specific knowledge even when generative knowledge would suffice? As I argue above in Section 1.2, reliance upon item-specific knowledge is rational if the language potentially contains idiosyncrasies. In this case, one should rely upon one's item-specific knowledge to the extent that one believes an item to be idiosyncratic. To test this proposal, I first set out to quantify how idiosyncratic binomial expressions are.

In Chapter 3, I describe the collection and annotation of a new corpus of binomial expressions that provides the data used in the remainder of the dissertation.

In Chapter 4, I explore the degree and manner of idiosyncrasy in binomial expressions through probabilistic models of the corpus data. I begin by showing that observed preferences in the corpus differ substantially from compositional preferences predicted by a basic model, suggesting that there is indeed a substantial amount of idiosyncrasy in binomial expression preferences. Of many further models considered, the only one which correctly predicts the distribution of observed preferences is one that explicitly allows the *regularization* of expressions—i.e. how consistently they are preferred in a given order—to vary as a function of expression frequency.[3] In

[3]This use of *regularization* is different from the notion of "conforming to compositional rules/preferences". Here we are specifically concerned with how *consistent* a preference is, independent of whether or not that preference conforms to what would be predicted compositionally. For further discussion of different types of regularization, see Section 5.1.1.

particular, I find that more frequent expressions are more regularized. In other words, I find that a) idiosyncrasy varies gradiently as a function of expression frequency, with more frequent expressions tending to be more idiosyncratic, and b) this idiosyncrasy tends to manifest in a particular way, namely for higher frequency expressions to be more extreme in their preferences. These findings support my rational motivation for the finding from Chapter 2. The language processor should rely on item-specific as well as generative knowledge because binomial ordering preferences are not entirely compositionally determined: as higher frequency items are more idiosyncratic, it is rational for the processor to rely more heavily upon item-specific knowledge for these items.

The results of Chapter 4 raise a further question: where does the tendency towards regularization come from? Chapter 5 takes up this question, using iterated learning models to simulate how language evolves over generations. Although previous iterated learning models have encoded across-the-board regularization biases, no previous models have demonstrated the *frequency-dependent* regularization seen in Chapter 4. I demonstrate that introducing a frequency-*independent* regularization bias—independently motivated by regularization behavior previously observed in a variety of statistical learning tasks—into the data-generation stage of a 2-Alternative Iterated Learning Model yields frequency-*dependent* regularization in the long term. This model demonstrates how the balance of compositionality and idiosyncrasy seen in corpus data can be achieved and preserved over generations through a combination of individuals' cognitive biases and the bottleneck effect of cultural transmission. Individuals' cognitive architectures include the competing forces of generative knowledge and an idiosyncrasy-promoting regularization bias. The cultural transmission bottleneck regulates how strongly these two forces apply to a given expression as a function of its frequency: for lower frequency items, the bottleneck is tighter, favoring generative knowledge; for higher frequency items, a wider bottleneck allows the effects of the regularization bias to multiply over time, promoting idiosyncrasy. I thus demonstrate how language processing and cultural transmission interact over the course of generations to preserve the balance of compositionality and idiosyncrasy seen in corpus data.

Finally, in Chapter 6, I return to explicitly testing the *gradience* of the trade-off between generative and item-specific knowledge in language processing. Modifying the experiments from Chapter 2 to look at expressions that continuously span the spectrum of overall frequencies from novel to highly frequent, we see a gradient trade-off in reliance upon generative and item-specific knowledge as a function of expression frequency, thus confirming the rational prediction made in Chapter 2.

In summary, the language processor flexibly recruits generative and item-specific knowledge to process binomial expressions, as a function of their frequency, which is a rational solution given that the expressions themselves are gradiently idiosyncratic as a function of frequency. The fact that synchronic preferences depend on both generative and item-specific knowledge in turn gives rise—in conjunction with a regularization bias—to the preservation of frequency-dependent idiosyncrasy in language structure.

Chapter 2

Abstract knowledge versus direct experience in processing of binomial expressions

2.1 Introduction

When we encounter common expressions like *I don't know* or *bread and butter*, do we process them word-by-word or do we treat them as holistic chunks? Research on sentence processing has largely focused on how single words are combined into larger utterances, but intuitively it seems that high frequency multi-word expressions might be processed holistically, even if they could in principle be treated compositionally. Recent research has thus questioned what possible sizes of combinatory units should be considered as the building blocks of sentence processing: Must all multi-word expressions be generated compositionally each time they are used, or can the mental lexicon contain holistic representations of some multi-word units?

The primary diagnostic for this question is whether the frequency of occurrence of multi-word expressions is predictive of their behavior in language processing. Such frequency effects are well documented at the level of individual words: more frequent words are faster to read (Inhoff and Rayner, 1986; Rayner and Duffy, 1986; Rayner et al., 1996), more likely to be skipped in reading (Rayner et al., 1996; Rayner and Well, 1996), and more susceptible to phonetic reduction (Bybee, 1999; Gregory et al., 1999). But do comparable frequency effects exist for multi-word expressions, when the frequency of their component words is controlled for? If the frequency of a given expression is being mentally stored, this implies that there is a mental representation of the expression as a whole. In contrast, if there are no frequency effects at the level of multi-word expressions, this is evidence against them having holistic representations akin to those of individual words.

A traditional view of grammar does not include holistic representations of multi-word expressions. According to this view, as described by Pinker (2000) under the title "traditional words-and-rules theory" (distinct from the modified words-and-rules theory he ultimately argues for), there is a strict separation between the individual words of a language and the rules for combining them. (See also Ullman, 2001; Ullman et al., 2005.) One tenet of this theory is that forms which can be generated compositionally are not stored: for instance, in the case of the English past tense, irregular forms are stored, while regular forms are generated anew using the *-ed* suffix each time they are used (Pinker, 1991). It remains possible within this theory that some regular forms—particular extremely high frequency ones—may be stored as well, but this is not the general method for dealing with such forms. As Pinker (2000) explains, one key motivation for this theory is memory constraints on the representation of language knowledge: it is more efficient to store a single, widely applicable rule than to store each regular form individually.

In a similar vein, this theory predicts that multi-word expressions should not be stored holistically because they can be generated compositionally, except in the case of non-compositional exceptions such as idioms (Swinney and Cutler, 1979). Again, as with regularly inflected wordforms, some exceptions may exist, but the exponentially larger number of multi-word expressions with which people have experience makes it even less likely that these expressions would be stored holistically, given the motivating concern with storage efficiency. The traditional words-and-rules theory thus does not predict that the processing of a multi-word expression will be affected by the frequency of the expression as a whole, though it can be affected by the frequencies of the individual words making up the expression.[1]

In contrast, there exists a growing movement of grammatical theories that do not draw a sharp distinction between the lexicon and the combinatory rules (e.g. Langacker, 1987; Johnson, 1997, 2006; Bybee, 2001, 2006; Goldberg, 2003; Gahl and Yu, 2006; Baayen et al., 2011; van den Bosch and Daelemans, 2013). Instead these approaches claim that people mentally store *exemplars*, or tokens of linguistic experience, which can be larger than single words. Language users form generalizations from exemplars at multiple levels of granularity (e.g. morpheme, word, or phrase) simultaneously, and the resulting network of generalizations constitutes our grammatical knowledge. Single words and multi-word expressions are thus on an equal footing: both are possible units that can be inferred from exemplars, and frequencies of multi-word expressions

[1] It may be possible to accommodate frequency effects for multi-word expressions under this theory, depending upon further details of the parser. In particular, processing of later words in an expression could be conditioned upon earlier words, thus creating an overall frequency difference. But this is not a direct prediction of the traditional words-and-rules theory.

are predicted to be stored and tracked just as frequencies of single words are.

Similar claims are made by exemplar-based computational models, which, like the exemplar-based grammatical theories, can incorporate combinatorial units of varying sizes from morphemes to sentences (e.g. Bod, 1998; Bod et al., 2003; Bod, 2008; Pierrehumbert, 2000; Johnson et al., 2007; O'Donnell et al., 2011; Post and Gildea, 2013). Within these models, the process of learning a grammar is explicitly one of deciding what sizes of units are most applicable or probable to explain the available language data. Under the learned grammars, many utterances can be parsed in multiple ways, either as combinations of individual words, or as holistic expressions, or various combinations thereof.

The development of these exemplar-based (or *usage-based*) theories is due in large part to previous demonstrations of frequency effects for multi-word expressions. Bybee (2006) reviews numerous corpus analyses demonstrating that the frequency of multi-word expressions is predictive of phonological reduction, grammaticalization, and other properties of usage, with a focus on highly frequent expressions such as *I don't know* or *going to*. Frequency effects for multi-word expressions have also been demonstrated in a controlled experimental setting: in a phrasal-decision task (analogous to a lexical decision task), Arnon and Snider (2010) found that more frequent phrases—e.g. *Don't have to worry*—were judged to be sensible phrases of English faster than less frequent phrases matched for word and substring frequencies—e.g. *Don't have to wait*. They further demonstrate that these effects exist across a wide range of frequencies, not just at the highest end of the frequency spectrum. (For a comparable finding using phonetic duration in corpus data, see Arnon and Cohen Priva, 2013.)

The exemplar-based approach also accords with more recent work on idioms, which challenges the traditional notion of idioms as strictly non-compositional. Gibbs (1990) and Nunberg et al. (1994) argue that many idioms can be seen as conventionalized metaphoric extensions of their literal meanings, and thus need not be treated as exceptions to the prevailing rules. (Similarly, see Holsinger, 2013.) On the whole, we thus see a broad shift towards recognizing that many expressions reside in a grey zone between entirely compositional and entirely non-compositional, and furthermore that an expression may be conventionalized while still being at least somewhat compositional.

But there remain open questions regarding these exemplar-based approaches and the interpretation of frequency effects for multi-word expressions. One limitation in the work to date is that it is difficult to differentiate the effects of language experience per se from the effects of real-world knowledge. Bybee (2006), for example, stresses the importance of language experience:

> As is shown here, certain facets of linguistic experience, such as the frequency of use of particular instances of constructions, have an impact on representation that we can see evidenced in various ways…

However, much of her cited evidence conflates linguistic experience with real-world experience. For example, in the phonological reduction of extremely frequent phrases such as *I don't know*, is this reduction due to the frequency of the linguistic expression per se, or is it due to the frequency of the event of not knowing something? Similarly, in the case of Arnon and Snider's contrast between phrases such as *Don't have to worry* and *Don't have to wait*, there could be a difference in the real-world likelihood of the events described by these expressions, which causes faster processing due to the difference in conceptual predictability, as opposed to linguistic predictability.[2] In general, this confound between linguistic experience and real-world knowledge exists whenever one compares expressions describing different real-world events.

Another outstanding question is how to empirically measure the trade-off between the reuse of stored multi-word expressions and the compositional generation of expressions. In the case of novel or infrequently attested expressions, we assume that such expressions must be processed compositionally using abstract linguistic knowledge—that is, generalized knowledge that is not bound to specific lexical items or expressions. In the case of frequently attested expressions, two potential processing strategies exist: compositional generation or reuse of stored holistic representations. Previous experimental work has primarily focused on the question of whether there is *any* reuse of stored multi-word expressions, and has suggested that there is at least some, but it remains possible that even very frequent and conventionalized multi-word expressions could in part or at times also be generated anew using abstract knowledge. Thus the major question now is *to what extent* both holistic reuse and compositional generation play a role in language processing (Wiechmann et al., 2013). As mentioned above, computational models have attempted to address this question by simulating what combination of linguistic units of varying sizes most parsimoniously predict corpus data (Bod et al., 2003; O'Donnell et al., 2011; Post and Gildea, 2013). But there has been no attempt so far to directly measure the competing influences of reuse and generation via behavioral experimentation.

Our work here does just that: we will quantify the extent to which people's

[2] Arnon and Snider did attempt to control for this real-world likelihood difference by collecting plausibility ratings for their materials, which they demonstrated did not differ in plausibility between conditions. However, plausibility in all conditions was very high, so extent differences may not have been detected due to ceiling effects.

processing of attested expressions is influenced by their frequency of direct experience with those specific expressions versus by the abstract linguistic knowledge that allows them to generate such expressions compositionally. To do so, we need to investigate a linguistic construction for which we can independently estimate people's frequency of direct experience and their abstract knowledge of its composition. Moreover, we want a construction with wide variation in how frequently attested specific instances of the construction are, so that we can measure how the influence of these competing explanations changes as a function of the overall frequency of an expression. For these reasons, an ideal construction is *binomial expressions*.

2.1.1 Binomial expressions

In this paper, we will address the generation and reuse of multi-word expressions by focusing on *binomial expressions* of the form *A and B*, such as *bread and butter* or *sweet and sour*. We include in our definition of binomial expressions all potential items with this form, including unattested expressions (e.g. *bishops and seamstresses*). Although binomial expressions are sometimes taken to include expressions with other conjunctions (e.g. *or*), here for simplicity we consider only expressions joined with *and*. Many binomial expressions have a preferred order (e.g. not *butter and bread* or *sour and sweet*), but binomials vary in how strong these ordering preferences are: some binomials are entirely fixed in order, or *frozen* (e.g. *safe and sound*/**sound and safe*), while others are quite free (e.g. *television and radio*/*radio and television*). Binomial expressions are thus a case of multi-word expressions that vary along two dimensions: how frequent they are, and how conventionalized their order is.

What causes binomial ordering preferences? One possibility is that preferences arise from abstract linguistic constraints that reference phonological, semantic, or other lexical properties of the elements in a binomial (e.g. the shorter word should come first). An alternate possibility is that preferences are driven by direct experience with the specific binomials in question: an order is preferred because it has been experienced more often.

Binomial expressions thus allow us to study the trade-off between abstract knowledge and direct experience. Specifically, we ask whether ordering preferences for binomials expressions are driven by direct experience with these expressions or by abstract constraints on the order of their elements. Moreover, we ask whether the influence of these two knowledge sources changes as a function of the frequency of an expression.

Additionally, binomial expressions are particularly suitable for studying effects of language experience per se, as opposed to real-world knowledge or other confounds, because both the syntax and semantics of these expressions are preserved regardless of ordering. We can thus study the effects of direct linguistic experience on these expressions by manipulating binomial ordering while minimizing confounds.

Previous work on binomial ordering preferences

Siyanova-Chanturia et al. (2011) demonstrated online effects of binomial ordering preferences: In an eye-tracking study, participants read common binomial expressions in either their preferred or dispreferred order, embedded in sentence contexts, e.g.:

(2.1) John showed me pictures of the *bride and groom* both dressed in blue.

(2.2) John showed me pictures of the *groom and bride* both dressed in blue.[3]

Expressions were read faster in their preferred order. Is this reading time difference due to the frequency of people's direct experience with these specific expressions or to their abstract knowledge of constraints on binomial ordering?

It has long been known that at least in certain contexts, binomial ordering preferences are sensitive to a variety of semantic, phonological, and lexical constraints, but the degree to which these constraints apply in online processing remains unclear. Early work portrayed these constraints as contributing to the diachronic longevity of expressions, while more recent work has suggested, albeit inconclusively, that such constraints play a role online as well.

Much of the existing work on binomial ordering preferences relies upon corpus analyses or analyses of hand-selected examples. Malkiel (1959) was the first to propose that the relationship between words in a binomial could contribute to the prominence or longevity of the expression. Based on hand-selected examples of frozen binomials, he proposes a number of constraints on ordering, both semantic and phonological, as well as discussing other possible relationships between words (e.g. rhyming and alliteration). A more extensive study of binomial ordering preferences was carried out by Cooper and Ross (1975), whose work focuses on demonstrating a *Me First* constraint, which posits that "first conjuncts refer to those factors which describe the prototypical speaker." (This prototypical speaker is later described as "Here, Now, Adult, Male, Positive, Singular, Living, Friendly, Solid, Agentive, Powerful, At Home, and Patriotic, among other things.") They further introduce a number of phonological constraints on ordering, noting that the various constraints seem to differ

[3]Binomial expressions are italicized here for clarity but were not italicized in the experiment.

in strength and may interact with each other, but they do not attempt to quantify these strengths or their interactions. Their investigation is based on a hand-selected sample of common binomial expressions, and they explicitly frame their discussion in terms of constraints that contribute to the diachronic longevity of an expression. Fenk-Oczlon (1989) introduced the idea that these constraints might apply to online processing as well as diachronic language change, arguing that most of Cooper and Ross's proposed constraints could be subsumed under the constraint that "the more frequent and therefore informationally poorer elements tend to occupy initial position" and that this new constraint is motivated by cognitive principles. His argument is supported by corpus data, but he does not provide any evidence from online processing measures. Similarly, Sobkowiak (1993), again based on corpus data, suggests that most of the previously proposed constraints can be subsumed under a principle of "unmarked-before-marked", which he relates to the information structure principle of "given before new".

More recent work has stopped attempting to unify disparate constraints and has instead focused on determining the relative rankings or weights of different constraints. In particular, Benor and Levy (2006) surveyed a large number of proposed constraints on ordering preferences from the previous literature, and considered a variety of probabilistic modeling frameworks for combining them. They found that a logistic regression model best predicts ordering preferences for a large selection of binomial expressions randomly selected from a corpus. Similarly, Mollin (2012) inferred a hierarchy of constraints from corpus data and found comparable rankings to those found by Benor and Levy.

While the existence of binomial ordering constraints in corpus data is well demonstrated, it is unclear whether these constraints apply only diachronically or whether they have synchronic cognitive status. Offline experimental tasks have suggested the synchronic cognitive reality of some constraints, mostly phonological. Using a forced-choice preference task in which subjects choose between possible orders of a binomial expressions, Bolinger (1962) demonstrated a preference to avoid having two stressed syllables in a row, comparable to findings in other domains of grammatical encoding (Jaeger, 2006; Lee and Gibbons, 2007). Pinker and Birdsong (1979) used a rating task with nonce words to argue for four phonological constraints, including "Panini's Law" (the shorter word, measured in syllables, should come first; named after the 4th Century B.C. Sanskrit linguist), as well as constraints on vowel quality, vowel length, and initial consonant obstruency. Wright et al. (2005) used a forced-choice preference task to demonstrate that male names preferentially precede female names, even when phonology and frequency are controlled for. Moreover, they showed that male names tend to have "first-position" phonological properties and are on average more frequent than female names. These offline tasks demonstrate that at least some abstract constraints on ordering are synchronically cognitively active, but they do not demonstrate whether these constraints are available during real-time language processing or whether they are available only upon later reflection.

Prior to Siyanova-Chanturia et al.'s work, a small number of online investigations used recall tasks to simulate language production, with mixed results regarding whether abstract ordering constraints are active in online production. Bock and Warren (1985) did not find effects of concreteness in ordering preferences, although the number of subjects and items in their task is small relative to the numbers we will use. Kelly et al. (1986) and Onishi et al. (2008) did find effects of prototypicality. McDonald et al. (1993) found effects of animacy and prosody, but—in contrast to Pinker and Birdsong—not word length. Thus the previous work provides weak evidence for some effects of abstract ordering constraints in production. The existence of such effects in comprehension has yet to be tested.

So based on our current knowledge, it is unclear whether to attribute the processing differences found by Siyanova-Chanturia et al. to the frequency of people's direct experience with these specific expressions or to their abstract knowledge of constraints on binomial ordering. Here we adopt a two-pronged approach to address this question. We look for effects of abstract ordering constraints on novel binomial expressions, thus establishing a baseline for such effects in the absence of direct experience with the binomials in question. Additionally, we compare the processing of these novel expressions with Siyanova-Chanturia et al.'s frequently attested expressions, allowing us to assess the relative roles of abstract knowledge and direct experience in the processing of attested expressions.

2.1.2 Our approach and its predictions

In this section, we describe in more detail the theoretical and methodological approach that we will take to studying binomial expressions. We begin by identifying three variables whose potential effects on processing we want to consider and determining how to quantify each one.

Independent variables of interest

For a word pair (A, B), the first variable we consider is the *overall frequency* of binomial expressions containing these elements—in other words, the combined frequency of the expressions *"A and B"* and *"B and A"*. To estimate the overall frequency of people's experience with these expressions, we can obtain frequency estimates from large corpora. Frequency can thus be analyzed as a continuous variable (generally measured in occurrences per million words). In this chapter we will treat frequency dichotomously (unattested versus frequently attested), but we return to the case of frequency as a continuous predictor in Chapter 6.

The next variable we consider is the *relative frequency*, or proportion of occurrences, of each order. Again, we can estimate this from corpus frequencies. The relative frequency of *"A and B"* is the number of occurrences of *"A and B"* divided by the overall frequency of (A, B) binomial expressions. It is thus a real number between 0 and 1, inclusive. The relative frequency of *"B and A"* is one minus the relative frequency of *"A and B"*.

The final variable we consider is the ordering preference due to people's *abstract knowledge* of binomial ordering constraints. For a given order *"A and B"*, we want a value between 0 and 1 corresponding to the probability of someone producing that order based on their knowledge of the abstract constraints governing binomial ordering. Unlike the previous two variables, we cannot directly estimate people's abstract knowledge from corpus frequencies. Instead, we will build a probabilistic model based on that of Benor and Levy (2006) to give us these estimates. In this paper, we make the simplifying assumption that abstract ordering preferences are fixed for a given expression; that is, they do not depend on the local context, linguistic or otherwise. This assumption would not always hold in a more naturalistic setting: in the corpus presented in Chapter 3, we find that ordering preferences for 4% of tokens are directly influenced by the local linguistic context, e.g. because one element in the pair was previously mentioned. However, our experimental materials (described in Section 2.3) will as much as possible avoid local contexts that would influence expression order, so we consider this a reasonable simplification for the present work.

Of these variables, the two that directly compete to explain binomial ordering preferences in online processing are relative frequency and abstract knowledge. Crucially, although these two variables may be correlated, we assume that they are not equivalent, as relative frequency can be influenced by factors beyond abstract knowledge such as conventionalization and idiomaticity, famous quotations, or language change that interacts with abstract ordering constraints (e.g. changes in word meaning or pronunciation). For example, although abstract knowledge includes a strong constraint to put men before women, *ladies and gentlemen* is strongly preferred to *gentlemen and ladies* due to its conventionalized use in formal addresses. Discrepancies between abstract knowledge and relative frequency are not necessarily limited to such extreme cases as *ladies and gentlemen* but may exist in subtler ways for many expressions in the language.

We further note that the roles of relative frequency and abstract knowledge in determining ordering preferences may change depending on the overall frequency of an expression: in the most extreme case, a never-before-encountered binomial by definition cannot be influenced by its relative frequency in previous experience. Our goal is therefore to measure the relative contributions of abstract knowledge and relative frequency to binomial ordering preferences, and to determine whether and how these change as a function of overall frequency.

Dependent variables of interest

In this chapter, we consider two measures of people's processing of binomial expressions. First, we carry out a forced-choice preference experiment in which people see both possible orders of a binomial expression and choose which they prefer. For each expression, we can then calculate the proportion of people who prefer a given order. Next, we measure reading times for expressions in each order as an online measure of processing difficulty. We thus obtain two measures indexing degree of human preference for one order over other. We can then test which combination of our proposed independent variables—overall frequency, relative frequency, and abstract knowledge—best predict the human data.

Predictions

Let us consider possible combinations of independent variables and what effects they might have on the behavioral data.

Abstract knowledge only One possibility is that only abstract knowledge of ordering constraints influences processing. This would be the case if a) there are no effects of direct experience with specific binomial orders (in line with the traditional words-and-rules theory of language processing), and b) there *are* online effects of ordering constraints. In this case, we predict that abstract knowledge but not relative frequency will have predictive power. More specifically, this theory predicts that abstract knowledge will be the best predictor of the behavioral data, and that its predictive power should not change as a function of relative or overall frequency.

Relative frequency only If, as predicted by exemplar-based theories, there are effects of direct experience with specific binomial orders, then relative frequency should influence behavior for expressions that people have experience with, i.e. expressions with nonzero overall frequency. If, furthermore, abstract ordering constraints are *not* active in online processing, then only relative frequency should play a role. In this case we predict that novel binomial expressions will show no ordering preferences because people have no experience with them, but that relative frequency will be predictive of the behavioral data for all attested binomials. Under such a theory, relative frequency may improve as a predictor with increased overall frequency, but this would be due to having more robust estimates of relative frequency with increased overall frequency, not due to any change in the role of abstract knowledge.

Both abstract knowledge and relative frequency If exemplar-based theories are correct that there are effects of direct experience, and moreover if abstract ordering constraints are active in online processing, then we predict that both relative frequency and abstract knowledge will be predictive of the behavioral data. For novel binomial expressions, with which people lack direct experience, abstract knowledge will be predictive. For attested expressions, some combination of abstract knowledge and relative frequency will be the best predictor (as predicted by Bod et al., 2003; O'Donnell et al., 2011; Post and Gildea, 2013).

To summarize, we investigate the roles of abstract knowledge and direct linguistic experience in the processing of both novel and frequently attested binomial expressions. We estimate people's direct experience with expressions in each possible order using corpus frequencies, and we estimate their abstract knowledge of ordering preferences using a probabilistic model. We evaluate which combination of these best predicts behavioral data in a forced-choice preference task and a self-paced reading task.

The organization of the remainder of this chapter is as follows: In Section 2.2, we introduce the probabilistic model used to estimate abstract knowledge of binomial ordering preferences. In Section 2.3, we describe the experimental materials used in our behavioral experiments. In Sections 2.4 and 2.5, we discuss two experiments. Section 2.6 gives a general discussion.

2.2 Probabilistic model of ordering preferences

We begin by developing a probabilistic model of binomial ordering preferences. This model integrates the constraints on ordering that have been discussed in the previous literature (as summarized by Benor and Levy, 2006), allowing us to approximate a native English speaker's abstract of knowledge of ordering preferences for a given binomial expression, independent of their direct experience with tokens of the expression.

We develop a logistic regression model following Benor and Levy. For a given word pair (A, B), this model predicts the probability that a binomial expression will be realized as *A and B*. We train our model on Benor and Levy's dataset, a random selection of binomial expressions drawn from a collection of corpora.[4] As Benor and Levy note, conclusions drawn from token counts rather than type counts may be skewed by the presence of a small number of very frequently attested frozen expressions (e.g. *back and forth*, with a token count of 49). We thus train our model on binomial

[4] For reasons that could not be determined, the version of the dataset we had access to contained 689 binomial tokens, three tokens fewer than stated in Benor and Levy.

types rather than tokens. This necessitated excluding expressions that appeared in both orders (15 word pairs), leaving us with 379 binomial expression types.

Benor and Levy coded their dataset for twenty potential constraints on ordering based on a thorough review of the previous literature. A constraint is said to be *active* for a given word pair if it favors one order over another; not all constraints are active for all word pairs. When constraints are active, they are binary-valued, favoring either word A first or word B first. Specifically, constraints are coded as 1 when they favor alphabetic order, -1 when they favor non-alphabetic order, and 0 when they are inactive. Outcomes are coded as 1 if the binomial expression appears in alphabetical order and 0 otherwise.

Benor and Levy did not do any model selection to determine which of their constraints were good predictors, although their results show that some, particularly among the nonmetrical phonological constraints, are very poor predictors. For our model, we use a subset of their constraints. Our goal is to develop the best possible model of binomial expression preferences that is nonetheless reasonably parsimonious (in particular, does not include those constraints that are clearly poor predictors), but it is not our goal to conclusively demonstrate that particular constraints are significant predictors of preferences: rather, our goal is to develop an effective predictive model that can be used to investigate the link between abstract knowledge of binomial ordering preferences and behavioral responses in offline and online processing tasks. We thus adopt relatively lenient criteria for inclusion of constraints in our final model. From Benor and Levy's twenty constraints, we begin by excluding two constraints that are rarely active in the dataset, and all expressions in which they are active: the Absolute Formal Markedness constraint (the two elements do not share a derivation, but one element is structurally more simple—i.e. contains fewer morphemes; active once) and the Pragmatic constraint (ordering is directly influence by the local linguistic context; active thrice). With the remaining constraints, we fit a logistic regression model using the `glm` function in R (R Core Team, 2014). Each constraint was entered as a predictor, with no interactions between constraints. We performed backwards model selection, excluding constraints one at a time based on their Wald z statistic, until all remaining constraints had $p < 0.15$.[5] (Backwards model selection is anti-conservative (Harrell, 2001), but this is not a problem in light of the desire for leniency discussed above.)

Our final model contains seven constraints. All affected the model's predicted ordering preference in the direction expected by Benor and Levy or by the sources who first proposed the constraint. See Table 2.1 for details of the constraint weightings. The constraints included in our final model are (with examples of binomials that satisfy each constraint drawn from the training data):

Formal markedness The word with more general meaning or broader distribution comes first. For example, in *boards and two-by-fours*, *boards* are a broader class of which *two-by-fours* is one member.

Perceptual markedness Elements that are more closely connected to the speaker come first. This constraint encompasses Cooper and Ross's (1975) 'Me First' constraint and includes numerous subconstraints, e.g.: animates precede inanimates; concrete words precede abstract words. For example, in *deer and trees*,

[5]We made one exception by keeping the Iconic Sequencing constraint in our model, although it had a high p value. This constraint was never violated in our dataset, and estimation of the Wald z statistic is unreliable in cases such as this with large estimated coefficients, due to inflated standard error estimates (Agresti, 2002; Menard, 2002). A likelihood ratio test supports our keeping this constraint in the model. (See Table 2.1.)

deer is animate while *trees* is inanimate.

Power The more powerful or culturally prioritized word comes first. For example, in *clergymen and parishioners*, *clergymen* have higher rank within the church.

Iconic/scalar sequencing Elements that exist in sequence should be ordered in sequence. For example, in *achieved and maintained*, a state must be *achieved* before it can then be *maintained*.

No final stress The final syllable of the second word should not be stressed. For example, in *abused and neglected*, *abused* has final stress and should therefore not be in the second position.

Frequency The more frequent word comes first, e.g. *bride and groom*.

Length The shorter word (measured in syllables) comes first, e.g. *abused and neglected*.

2.2.1 Model validation

We validate the model by testing its predictions on the training data and on the 42 attested binomials used by Siyanova-Chanturia et al. (2011).[6] Constraint values for the Siyanova-Chanturia et al. binomials were hand-coded as described in Section 2.3. The model correctly predicts the ordering preferences for 287/372 (77%) of the training data and 30/42 (71%) of Siyanova-Chanturia et al.'s items, both significantly greater than chance (50%) in a one-tailed binomial test ($p < 0.001$ and $p < 0.01$).

[6] The dataset on which we originally trained our model contained seven binomial expressions that were also included in Siyanova-Chanturia et al.'s (2011) items. Therefore, after doing model selection on the original dataset, we retrained our model, excluding these seven items from the training data. All results, beginning with Table 2.1, are reported based on the retrained model.

Table 2.1: Constraint weights in our probabilistic model. In addition to reporting the Wald z statistic and p-values based on it (columns 3–4), we report results of a likelihood-ratio test comparing versions of the model differing only in whether they include the constraint in question (and containing all other constraints; columns 5–6).

Constraint	*Regression coeff.*	*Std. Error*	*z value*	*p value (z)*	*Log-lik ratio*	*p value (χ^2)*
Formal Markedness	1.39	0.56	2.49	0.01	3.85	0.006
Perceptual Markedness	1.72	0.51	3.40	0.0007	7.77	0.00008
Power	1.03	0.57	1.81	0.07	1.81	0.06
Iconic Sequencing	18.62[a]	709.22	0.026	0.98	53.47	$< 2x10^{-16}$
No Final Stress	0.50	0.33	1.50	0.13	1.16	0.13
Frequency	0.32	0.14	2.35	0.02	2.76	0.02
Length	0.43	0.21	2.07	0.04	2.18	0.04

[a]This coefficient is effectively infinity, as this constraint is never violated in the training data. See Footnote 5 regarding the standard error and z statistic in this case.

2.3 Experimental materials

Using our probabilistic model, we develop the linguistic stimuli used in both experiments. Our stimuli consisted of 84 word pairs, with each pair producing two possible binomial expressions (*A and B* or *B and A*). 42 of our items, taken directly from Siyanova-Chanturia et al. (2011), are frequently attested. They range from almost completely frozen (e.g. *bread and butter*) to relatively flexible (e.g. *radio and television/television and radio*).

We further created 42 novel items which our model predicts to have strong ordering preferences (e.g. *bishops and seamstresses/seamstresses and bishops*). To ensure that speakers have no prior experience with these expressions, we consult the nearly 500-billion-word Google books n-gram corpus (Lin et al., 2012). Our novel binomials are not included in this corpus in either order.[7]

Our probabilistic model gives us an estimate of the direction and strength of ordering preference for each item based on abstract ordering constraints. To generate model predictions for these items, we must code them for the seven constraints described in Section 2.2. Final Stress and Length were coded by either the first author or a trained research assistant, both native speakers of American English. Frequency estimates were obtained from the HAL database via the English Lexicon Project (Balota et al., 2007).[8] Coding the remaining four constraints requires real-world

[7]Levy et al. (2012) estimate that college-age English speakers have been exposed to no more than 350 million words of English in their lifetimes. To be included in the Google books corpus, an n-gram must have appeared at least 40 times in their 468,491,999,592 word corpus. Thus our binomials can have appeared at most 39 times in this corpus, and there is at most a roughly 3% chance that a college-age speaker would have heard any given one of these expressions. Although our participants are on average slightly older than college-age, we believe there is still an exceedingly small chance that they will have substantial experience with any of these expressions.

[8]On three occasions, one word in a pair was not in the English Lexicon Project database (*groundskeeper, ninety-eighth*, and *wildfires*). In these cases, the non-included word was assumed to be the less frequent.

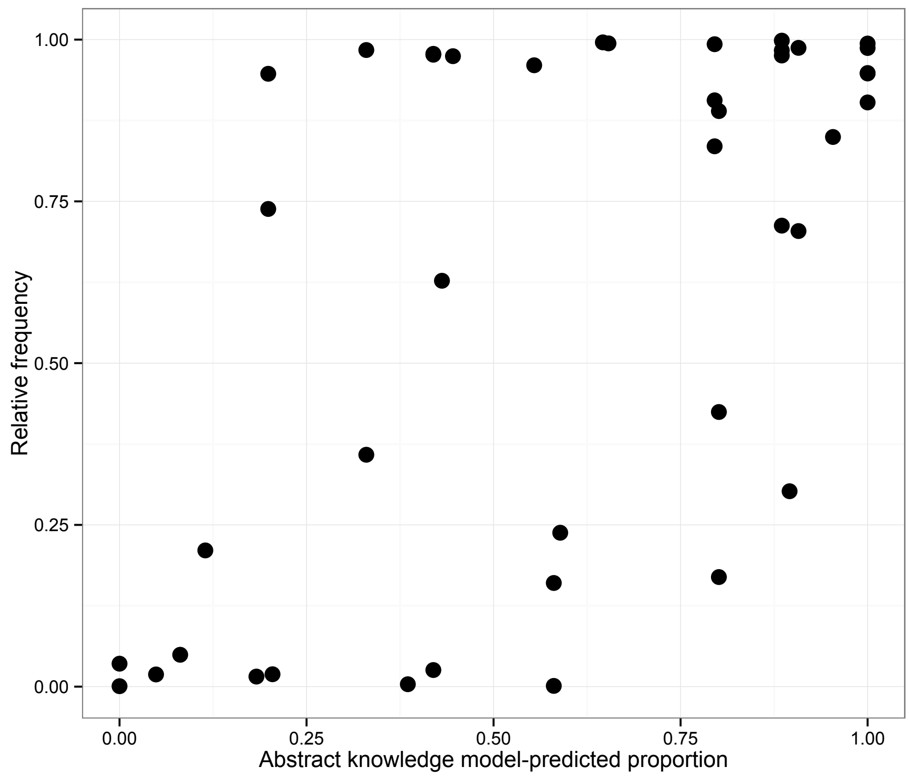


Figure 2.1: Abstract knowledge model-predicted proportion and empirical relative frequency of each attested binomial appearing in alphabetical order. Abstract knowledge and relative frequency are significantly but not perfectly correlated.

knowledge, and so they were coded twice, independently, by the first author and a trained research assistant. Conflicting judgments were resolved through discussion; with discussion, the two coders were always able to reach agreement.

As predicted in Section 2.1.2, our attested items show a significant but not perfect correlation between model-predicted abstract ordering preference and relative frequencies (computed from the Google n-grams corpus; Brants and Franz, 2006): $r(40) = 0.59; p < 0.0001$. This relationship is visualized in Figure 2.1.

For our novel binomials, we chose expressions that our model predicts to have strong ordering preferences, with values less than 0.3 or greater than 0.7. As much as possible, we chose expressions that minimized the correlations between constraints (e.g. to dissociate length and frequency). A comparison of the profiles of constraint activity for novel and attested items is given in A.1.3.

For all items, both novel and attested, we constructed a sentence context for the binomial expression, e.g.:

(2.3) There were many bishops and seamstresses in the small town where I grew up.

(2.4) There were many seamstresses and bishops in the small town where I grew up.

Sentence structure was unrestricted, but the binomial expression was never in the first two or the last four words of the sentence. Sentences were designed not to introduce pragmatic constraints on binomial ordering: in particular, neither binomial element (nor any word related exclusively or primarily to only one of the elements) was mentioned in the sentence before the binomial occurred.

With these materials, we carried out two behavioral experiments, a forced-choice preference experiment and a self-paced reading experiment.

2.4 Experiment 1: Forced-choice preference

2.4.1 Method

Participants

75 native English speakers (mean age=36 years; sd=14) participated. Participants were recruited through Amazon Mechanical Turk, restricted to people connecting to the website from within the United States, and were paid 50 cents. Participants

were asked to report their "Native language (what you learned to speak with your mother as a child)". Those who did not report English among their native languages were excluded.

Procedure

The Amazon Mechanical Turk instructions directed participants to an external website, where our experiment was presented using WebExp (Keller et al., 2009). Participants first filled out a demographic questionnaire, then continued to the main experiment. On each trial, participants saw one item embedded in sentence context, in both possible orders, e.g.:

- There were many bishops and seamstresses in the small town where I grew up.
- There were many seamstresses and bishops in the small town where I grew up.

Participants were asked to choose which order "sounds more natural". Each participant saw all 84 items. Which expression order was listed first was counterbalanced across participants. Order of item presentation was randomized separately for each participant. The experiment typically took 10-15 minutes.

2.4.2 Results

Before proceeding with multivariate analysis of the effects of abstract knowledge and direct experience on ordering preference, we present a striking overall difference between the distributions of preference strengths for attested versus novel binomials. Figure 2.2 shows that ordering preferences are more polarized for attested than for novel binomials (despite the fact that we selected our novel binomials to have extreme preferences); in other words, preferences are more consistent across subjects for the attested expressions. We define a measure of extremity for each item as the difference between its experimentally determined preference strength (i.e. proportion of times preferred in alphabetical order) and 0.5. In a t-test, the attested items are significantly more extreme than the novel ($t = 8.31, p < 0.001$). We discuss this issue further in Sections 2.4.3 and 2.6.3.

Multivariate analysis

Next we analyze our data using mixed-effects logistic regression (Jaeger, 2008). Our dependent variable is the preferred order, coded as alphabetical or non-alphabetical: alphabetical order is used as a neutral order because results of our initial model selection—see Section 2.2—indicate that alphabetical order is not a significant predictor of ordering preference. Our independent (fixed-effect) predictors are:

- **Type** (attested/novel) is treatment coded with "attested" as the reference level, i.e. the `Intercept` value applies to attested items, and this value is adjusted by the `Type:novel` value for novel binomials. We predict no significant intercept (i.e. attested binomials are not significantly more likely to be preferred in alphabetical or non-alphabetical order, absent other factors), and no significant effect of type (i.e. novel binomials are not significantly more or less likely to be preferred in alphabetical order than attested binomials).

- **Abstract knowledge** is operationalized as our model's predicted probability (between 0 and 1) of the expression occurring in alphabetical order. We center this predictor around 0.5. We nest the abstract knowledge predictor within type,

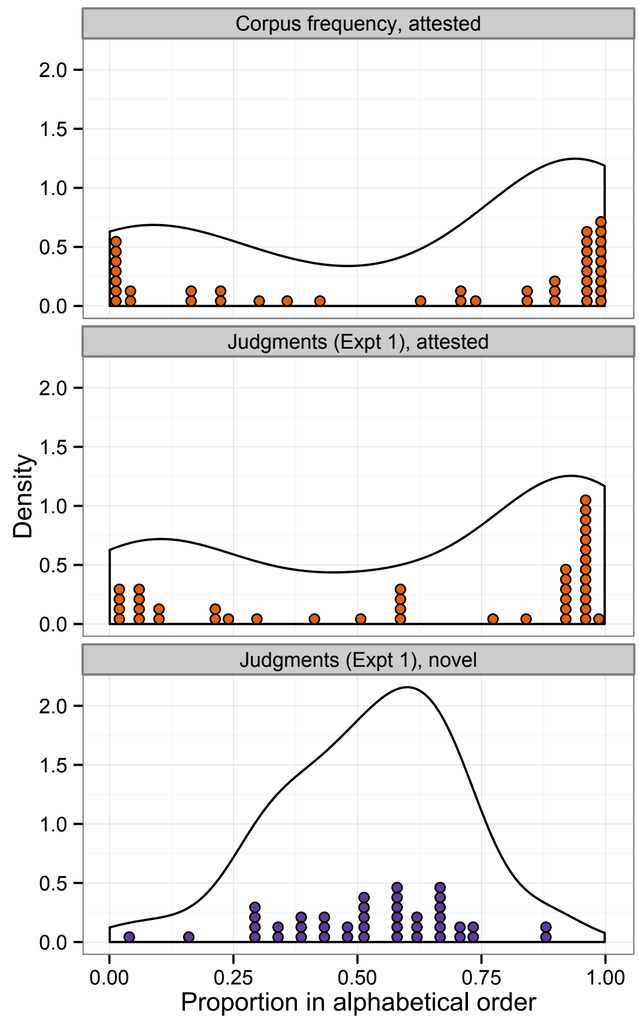


Figure 2.2: Results of Experiment 1: Proportion of binomials occurring in alphabetical order in Google n-grams corpus frequency (top) and subjects' forced-choice preference judgments (middle/bottom). Dots show individual binomial types, while lines show density estimates. In judgments, attested binomials have more extreme preferences (i.e. more consistent across subjects) than novel binomials, demonstrating a qualitatively similar distribution to corpus frequencies.

i.e. we fit separate parameters for the effect of abstract knowledge for novel and attested binomials, allowing us to consider the effects of abstract knowledge on each type independently. For each type, if abstract ordering constraints are active in influencing offline judgments, then we predict a significant effect of abstract knowledge.

- **Relative frequency** estimates are computed for attested binomials using the Google n-grams corpus (Brants and Franz, 2006) as the frequency of *"A and B"* divided by the frequency of *"A and B"* plus *"B and A"* (resulting in a value between 0 and 1), and centered around 0.5. Relative frequency for all novel binomials is set to 0 after centering. (Thus no interaction of relative frequency with type is necessary.) If direct experience with attested expressions influences offline judgments, then we predict a significant effect of relative frequency.

Following Barr et al. (2013), we use the maximal random effects structures for subjects and items justified by the experimental design: by-subject and by-item intercepts, and by-subject slopes for type, abstract knowledge, their interaction, and frequency.

Model results are given in Table 2.2.[9] Significance levels for effects are reported using the Wald z statistic and are confirmed using likelihood ratio tests. We see a significant effect of abstract knowledge for both novel and attested expressions, demonstrating that abstract ordering constraints are active in determining forced-choice preferences for both binomial types. In a likelihood ratio test comparing this model to a model with only an additive (non-nested) fixed effect of abstract knowledge, we find no significant difference ($\chi^2(1) = 1.63, p = 0.20$); in other words, the effect of abstract knowledge does not differ significantly between novel and attested expressions.

Table 2.2: Model fit for results of Experiment 1. All VIF < 1.2.

	Estimate	*Std. Error*	*z value*	*p value*
Intercept	-0.14	0.15	-0.98	0.33
Type: novel	0.25	0.19	1.32	0.19
Abs know (Type: attested)	2.32	0.56	4.12	.00004***
Abs know (Type: novel)	1.45	0.35	4.11	.00004***
Rel freq	6.18	0.49	12.55	$< 2x10^{-16***}$

The effect of abstract knowledge for novel binomials is displayed in Figure 2.3.

We also see a significant effect of relative frequency, demonstrating that direct experience also plays a role in determining preferences for attested expressions. We note that relative frequency is a stronger predictor than abstract knowledge, measured in terms of larger regression coefficient estimate, larger z value, and larger change in likelihood when removed from the model. The strong predictive power of relative frequency is displayed in Figure 2.4.

2.4.3 Discussion

In this experiment, we set out to test whether abstract knowledge and direct experience (specifically, relative frequency) predict ordering preferences in a forced-choice preference task for both novel and frequently attested binomial expressions.

[9]The model presented here includes all the fixed-effect predictors and interactions that are of crucial theoretical interest for the hypotheses we set out to test. In order to explore possible further interactions between predictors, as well as possible changes in behavior over the course of the experiment, we fit a mixed-effects logistic regression including as predictors all the previous predictors, a trial order predictor, and all two-way interactions, using the `MCMCglmm` package in `R` (Hadfield, 2010). (The trial order predictor was not included in the original model presented here because a main effect of trial order is implausible, as it would indicate a changing probability of prefering binomials in alphabetical order over the course of the experiment. However, its interaction with other predictors—in particular, abstract knowledge and relative frequency—is potentially of interest.) No further interactions (beyond the type x abstract knowledge interaction included in the original model) reached significance.

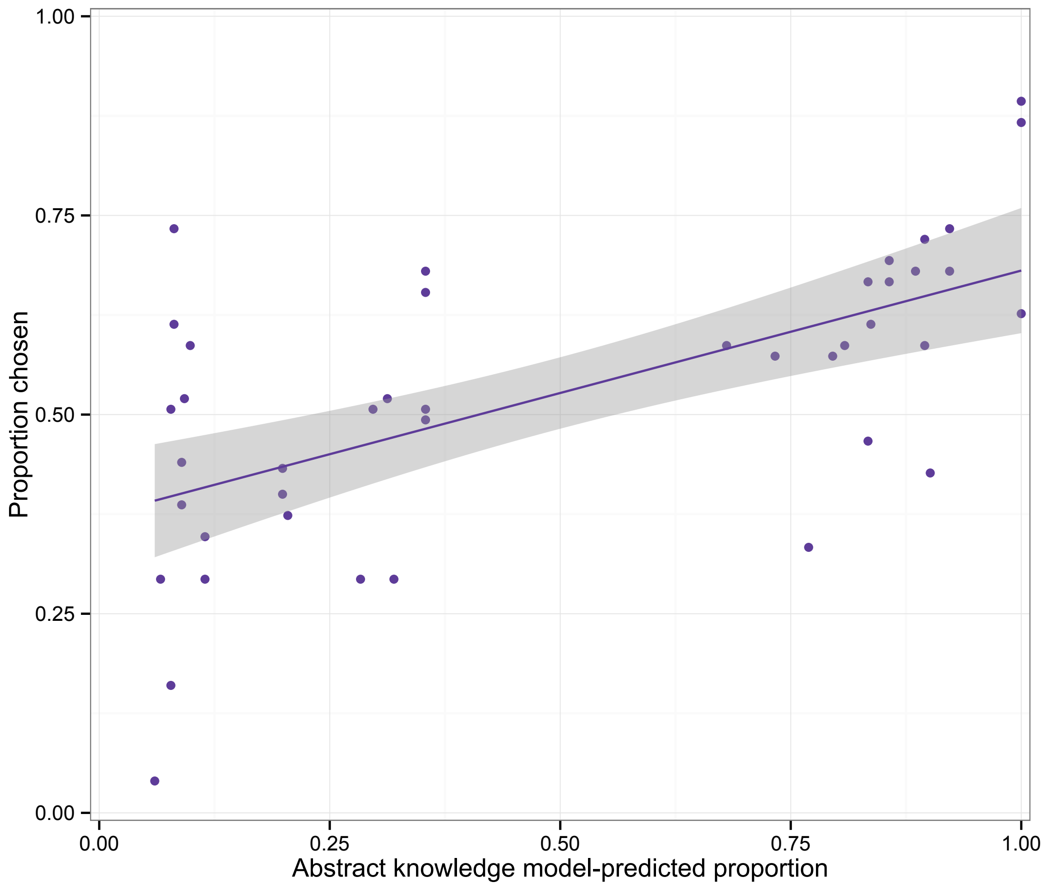


Figure 2.3: Results of Experiment 1 (novel items): Ordering preferences for novel binomials by model-predicted abstract knowledge. Each point represents an item. x values are the abstract knowledge model's prediction for how often the item will appear in alphabetical order. y values are how often the item was preferred in that order. Line shows best linear fit on the by-items aggregated data. Abstract knowledge is a significant predictor of preferences for novel expressions.

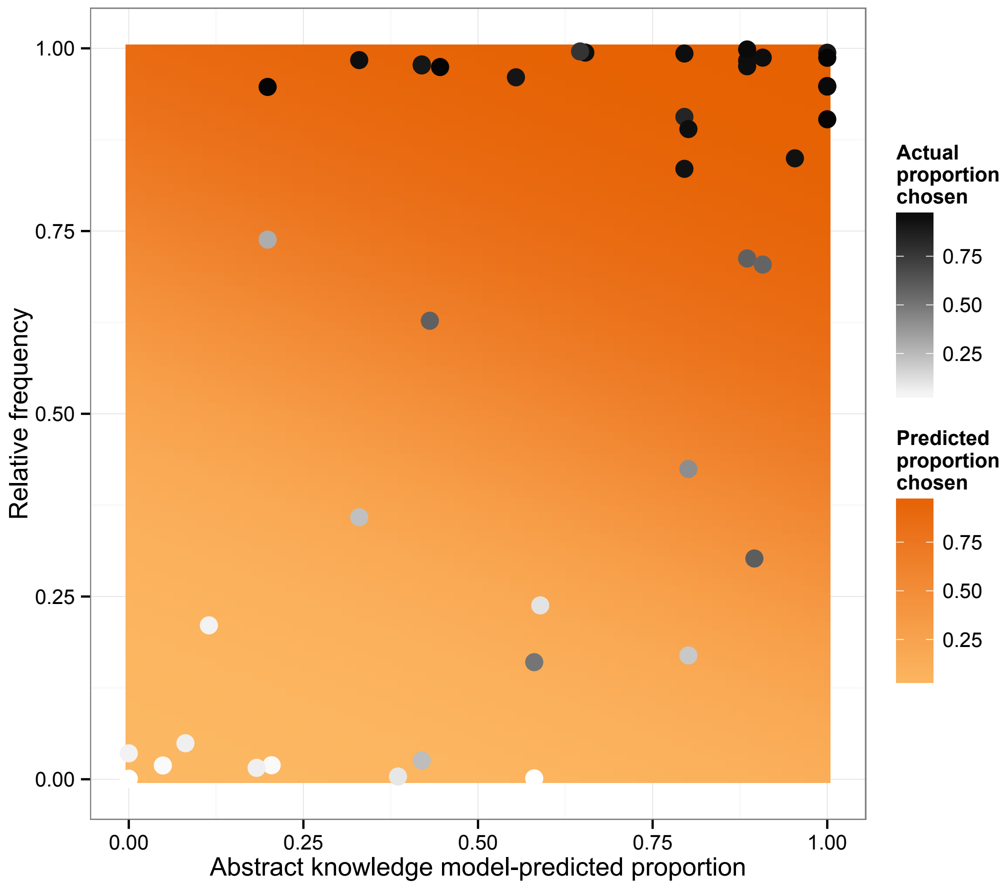


Figure 2.4: Results of Experiment 1 (attested items), visualized as colors overlaid on Figure 2.1. Each point represents an item. x values are the abstract knowledge model's prediction for how often the item will appear in alphabetical order. y values are the item's relative frequency of appearing in that order. Points' shading (white to black) shows often the item was preferred in that order. Background shading (light to dark orange) shows the best-fit model (Table 2.2) prediction for how often the item was preferred in that order. Both relative frequency and abstract knowledge predict true preferences, as depicted by the diagonal background gradient but relative frequency is the stronger predictor, as depicted by the stronger vertical than horizontal gradient.

We demonstrate that preferences of both attested and novel expressions are affected by abstract knowledge and that preferences of attested expressions are also strongly predicted by relative frequency. This pattern of results supports a theory wherein both abstract knowledge and direct experience play a role in processing. Moreover, for attested expressions, we find that relatively frequency is a stronger predictor of preferences than abstract knowledge, suggesting that processing of these expressions relies more heavily upon direct experience than upon abstract knowledge.

Although the effect of abstract knowledge does not differ significantly across binomial types, we do not think it is justified to draw strong theoretical conclusions from this null result. As we will see in Section 2.5.2, abstract knowledge does interact significantly with binomial type in Experiment 2. We defer further discussion of this issue until Section 2.5.3.

We additionally find that forced-choice preferences are more extreme for attested than for novel expressions; that is, attested expressions are more consistently preferred in one direction than novel expressions. Taken at face value, this finding suggests that increased overall frequency of an expression exaggerates or solidifies people's preferences. Another possibility, however, is that preferences for novel expressions are underlying equally as extreme as those of the attested expressions, but that the forced-choice judgement process for these items is noisier,[10] making the resulting preferences for novel expressions appear less extreme than they truly are. We will return to this question in the general discussion.

One potential confound mentioned earlier is the role of local sentence context on binomial order preferences. Although we tried to avoid biasing contexts in designing

[10]There are many reasons why this could be the case. For instance, when judging attested items, participants may believe that there is a "right" answer and take care to give that answer, whereas when judging novel items, they may put in less effort.

our materials, it is always possible that some bias unintentionally slipped through. Thus we hasten to point out that even if such bias does exist within individual sentences—i.e. the sentence context favors one order more than another, relative to the binomials' intrinsic ordering preference in a hypothetical neutral context—it would not confound the results presented here. Specifically, because our dependent variable is an alphabetical versus non-alphabetical preference, in order to bias our results the local context biases would need to be systematically correlated with the alphabetical/non-alphabetical preferences as given by our predictors of interest (abstract knowledge and relative frequency). Since we have no reason to expect this to be the case, any unintentional effects of local context will merely add noise to our estimates of ordering preferences.

In the next experiment, we ask whether the patterns found in our forced-choice preference experiment likewise hold in an online reading experiment.

2.5 Experiment 2: Self-paced reading

2.5.1 Method

Participants

400 native English speakers (mean age=34 years; sd=12) participated.[11] Participant recruitment was identical to Experiment 1, except that participants were paid $1.00.

[11]Experiment 2 required substantially more participants than Experiment 1 because the self-paced reading data is noisier than the forced-choice data, and because each subject saw approximately half the items in Experiment 2 (compared to all the items in Experiment 1).

Procedure

The experiment was presented within Amazon Mechanical Turk using flexspr (Tily, 2012; previously used by Bergen et al., 2012; Singh et al., 2015). Using this method online allows for collection of more data than would be possible in a laboratory setting, and previous work has replicated multiple in-the-lab results with web-based self-paced reading (Enochson and Culbertson, 2015). Participants first filled out a demographic questionnaire, then read sentences in a self-paced reading paradigm: sentences were presented one word at a time, and participants pressed a button to advance to the next word. Reading times for each word were recorded. Participants read three practice sentences, then continued to the main experiment.

Our materials consisted of the same 84 binomial expressions in sentence context as used in Experiment 1, plus 84 unrelated filler sentences. Two stimulus lists were constructed with items rotated and counterbalanced between lists so that each participant only saw a given binomial in one of its two possible orders. Due to a programming error, out of the 168 items in each list, each participant saw a random selection of 80 items. Order of presentation was randomized separately for each participant.

Presentation of each sentence was followed by a yes/no comprehension question. Answers did not depend on the order of the binomial expression. The experiment typically took about 30 minutes.

Table 2.3: Comprehension question accuracy for Experiment 2. Novel and attested items are divided into preferred/non-preferred order according to abstract knowledge model predictions.

	novel	attested
preferred	0.97	0.97
non-preferred	0.97	0.96

2.5.2 Results

Comprehension question accuracy

Comprehension question accuracy is extremely high across all conditions. See Table 2.3.

Multivariate analysis

We use regression analysis to compare abstract knowledge and relative frequency as predictors of reading times, analagous to our analysis in Experiment 1.

We divide our experimental items into regions of analysis as shown below:

Prelim	*Word1*	*And*	*Word2*	*Spill1*	*Spill2*	*Spill3*	
	bishops		seamstresses				
There were many		and		in	the	small	...
	seamstresses		bishops				

The *Prelim* region encompasses the entire beginning of the sentence up to the binomial expression; all further regions are a single word. We analyze reading time data for each trial summed over a six-word region spanning from Word1 through Spill3. By summing across reading times for these regions, we take advantage of the controlled properties of our stimuli: regardless of order of binomial presentation across conditions, participants will have read the same group of words within the region being analyzed.

(For more direct comparison with the previous literature, we present word-by-word analyses of reading times in B.1.)

Specifically, we computed a summed reading time measure for each trial as follows: we excluded all trials in which the reading time for any word was less than 100ms or greater than 5000ms (following Fine et al., 2013). To account for influences of word length, as described by Ferreira and Clifton (1986), we then computed subject-specific residualized reading times (regressed against word length) for each word from the Word1 through Spill3 regions, using data from all non-sentence-final words in non-practice trials.[12] Summing the residuals for this six-word region gives us a residual reading time for each trial. We performed outlier removal without regard to item type or condition: we computed a grand mean and standard deviation and exclude trials with summed times more than 2.5 standard deviations above or below the mean (following Garnsey et al., 1997), resulting in a loss of 1.7% of data.

We analyze the data using a mixed-effects linear regression similar to that used in Experiment 1. Our dependent variable is summed residual reading time. Our independent (fixed-effect) predictors and their interpretations are identical to those used in Experiment 1 (Section 2.4.2) with one addition:

- **Trial order** is the position in the experiment in which the given trial occurred. As is common in reading experiments (e.g. Hofmeister et al., 2011; Fine et al., 2013 and many others), we expect that subjects will read faster later in the experiment due to practice effects.

In addition to our hypotheses regarding possible influences of abstract knowledge and direct experience on reading times (which are the same as in Experiment 1), we

[12]For analyses using raw reading times, see Appendix B.2.

additionally anticipate a possible statistically significant but theoretically uninteresting main effect of binomial type because the two types contain different words in different sentence frames, and thus one type may be read faster than the other. Following Barr et al. (2013), we use the maximal random effects structure for subjects as justified by the experimental design, namely an intercept and slopes for type, abstract knowledge, their interaction, and relative frequency. We also include a by-subjects random slope for trial order. For items, defined as unordered word pairs, we include a random intercept, a random slope for trial order, and (in place of random slopes for both abstract knowledge and relative frequency) a random slope for a binary alphabetical/non-alphabetical factor, thus allowing for arbitrary item-specific ordering preferences.

Model results are given in Table 2.4.[13] Effects with $t \geq 2$ are taken to be significant. Positive coefficients indicate slower reading. We see a significant main effect of type with novel expressions read slower, which we attribute to these expressions containing less frequent words on average, in addition to being less frequent expressions overall.

We do not find a significant effect of abstract knowledge for attested expressions, suggesting that abstract ordering constraints are not active in the online processing of these expressions. However, we do find a significant effect of abstract knowledge for novel expressions. In a likelihood ratio test comparing this model to a model with only an additive (non-nested) effect of abstract knowledge, we find a significant

[13]The model presented here includes all the fixed effect predictors and interactions that are of crucial theoretical interest for the hypotheses we set out to test. In order to explore possible further interactions between predictors, we fit a mixed-effects linear regression including as predictors all these fixed-effect predictors and all two-way interactions using the `MCMCglmm` package in `R` (Hadfield, 2010). No further interactions (beyond the type x abstract knowledge interaction included in the original model) reached significance.

Table 2.4: Model fit for results of Experiment 2. Effects with $t > 2$ are taken to be significant. All VIF < 1.7.

	Estimate	*Std. Error*	*t value*
Intercept	196.34	26.04	7.54
Type: novel	195.17	25.77	7.57
Abs know (Type: attested)	13.88	23.14	0.60
Abs know (Type: novel)	-48.73	18.02	-2.70
Rel freq	-59.25	18.42	-3.22
Trial order	-8.35	0.39	-21.24

difference ($\chi^2(1) = 4.24, p < 0.04$); in other words, the effect of abstract knowledge differs significantly between novel and attested expressions, playing a significant role in online processing for novel expressions only. We additionally find a significant effect of relative frequency, demonstrating that higher relative frequency leads to faster reading in the online processing of attested expressions.

Finally, we find a significant effect of trial order, with faster reading later in the experiment. Results are visualized in Figures 2.5 and 2.6.

2.5.3 Discussion

We demonstrate for the first time that novel binomial expressions show online effects of abstract ordering preferences. In contrast, reading times for frequently attested binomial expressions are only influenced by relative frequency. These findings imply a trade-off in online processing between reliance on abstract knowledge and direct experience, where novel expressions must be processed on the basis of abstract knowledge only, but highly frequent attested expressions can be processed primarily with reference to previous direct experience.

Here we found a significant interaction of abstract knowledge with binomial

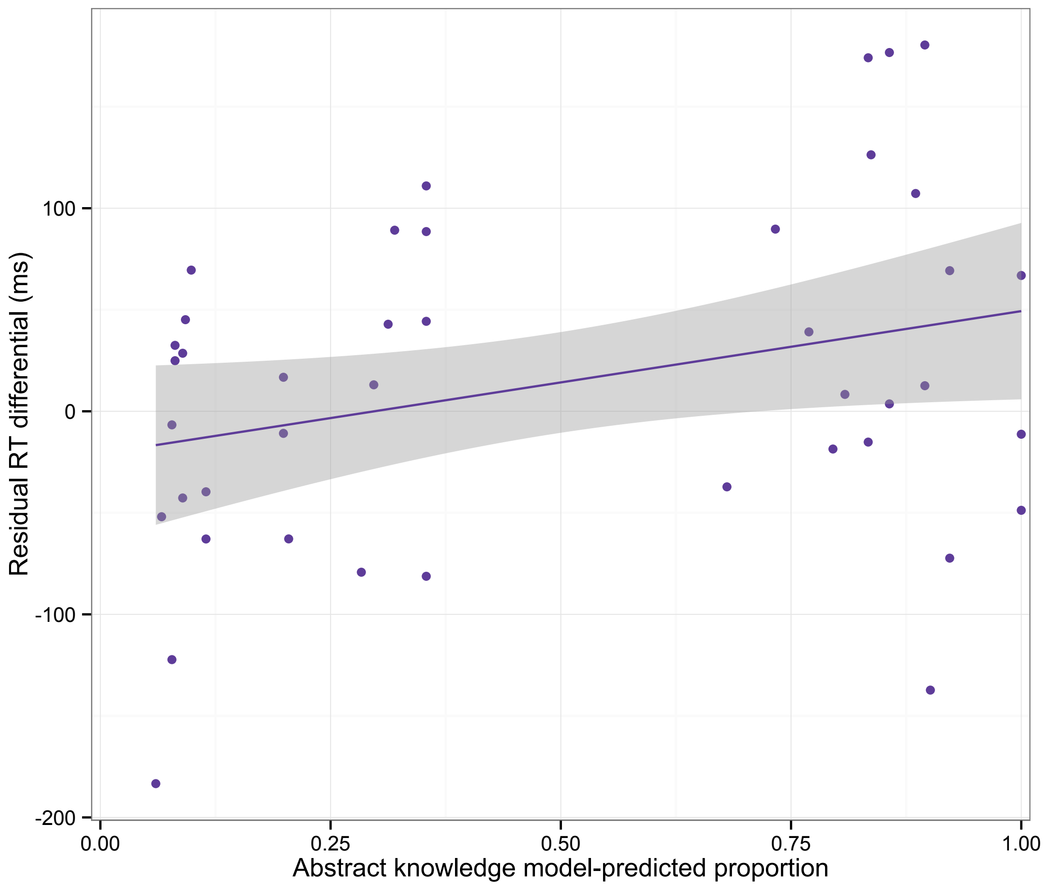


Figure 2.5: Results of Experiment 2 (novel items): Reading time differentials for novel binomials by model-predicted abstract knowledge. Each point represents an item. x values are abstract knowledge model's predictions for how often the item will appear in alphabetical order. y values are the differences between average summed residual reading times for the non-alphabetical and alphabetical orders. Line shows best linear fit on the by-items aggregated data. Abstract knowledge is a significant predictor of reading times.

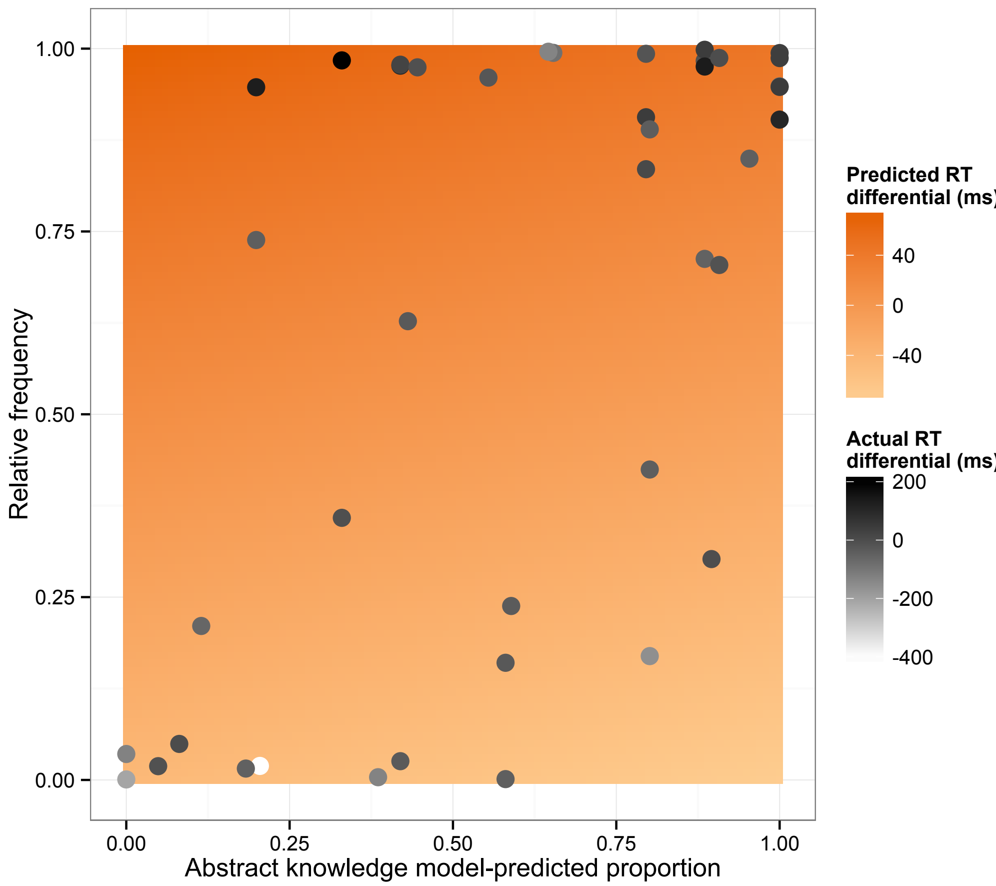


Figure 2.6: Results of Experiment 2 (attested items), visualized as colors overlaid on Figure 2.1. Each point represents an item. x values are the abstract knowledge model's prediction for how often the item will appear in alphabetical order. y values are the item's relative frequency of appearing in that order. Points' shading (white to black) shows the item's true average RT differential (the differences between average summed residual reading times for the non-alphabetical and alphabetical orders). Background shading (light to dark orange) shows the best-fit model (Table 2.4) prediction for RT differential. Only relative frequency is a significant predictor of reading times, as depicted by the strong vertical background gradient.

type, such that abstract knowledge was significantly less active in determining reading times for attested binomials than for novel binomials. In contrast, in Experiment 1, we found no such significant interaction. What is consistent across these two experiments is that processing of attested expressions is more strongly influenced by direct experience than by abstract knowledge. However, given the inconsistent results concerning the interaction of abstract knowledge and binomial type, we cannot state with confidence whether abstract knowledge is differentially active between novel and attested binomials.

2.6 General discussion

We set out to investigate the roles of abstract knowledge and direct experience in the processing of binomial expressions, asking whether binomial ordering preferences are driven by constraints on the semantic, phonological, and lexical properties of words in an expression, or by prior experience with the specific expression in question. Our key findings are as follows. First, we demonstrated that abstract ordering constraints are active in the comprehension of novel expressions in both an offline forced-choice task and a online self-paced reading task. Second, we demonstrated that for frequently attested expressions, effects of direct experience largely overwhelm abstract knowledge in predicting behavioral data, both in the offline task and especially in the online task.

Our results support exemplar- or usage-based theories of language, which allow for the storage and reuse of multi-word expressions. Specifically, our finding that ordering preferences for attested binomial expressions are primarily driven by relative frequency is evidence that the processing of these expressions makes use of holistic multi-word mental representations. In contrast, a traditional words-and-rules theory would predict that these expressions are generated compositionally each time they are encountered, and that the ordering preferences of attested expressions, like those of novel expressions, should stem from abstract ordering constraints rather than relative frequency of direct experience.

Of the predictions made in Section 2.1.2, our results indicate that both abstract knowledge and relative frequency play a role in the processing of binomial expressions. Many patterns are possible for the manner in which these two knowledge sources trade off as a function of the overall frequency of an expression: In one extreme, abstract knowledge could apply only for expressions that have never before been encountered, with relative frequency taking over as soon as any direct experience exists. In the other extreme, abstract knowledge could apply in the vast majority of cases, with relative frequency limited to playing a role only for the highest frequency items, such as those used in our experiments. A middle ground position proposes a gradual switch from reliance on abstract knowledge to reliance on relative frequency as overall frequency increases.

We propose that both extremes are unlikely and that the middle position of a gradual trade-off is the most likely. The first extreme is counterintuitive, since a single encounter with an expression seems insufficient to thoroughly trump abstract knowledge. The second extreme has been argued against by Arnon and Snider (2010), who found frequency effects for multi-word expressions across a wide range of frequencies. Their finding of frequency effects for low-to-medium frequency items would not be predicted by a theory in which direct experience applies only to the processing of extremely high frequency items. The gradual trade-off theory, on the other hand, is supported by a wide variety of computational models.

2.6.1 Convergent evidence from computational models

Connectionist models

A similar trade-off has been demonstrated in connectionist models of language learning in domains such as past-tense formation (Rumelhart and McClelland, 1986) and grammatical structure (Elman, 2003), which learn both generalized patterns and specific exceptions. These models learn to predict patterns within their training data (e.g. Form the past tense by adding *-ed*). When new items are introduced, they are at first treated accorded to the general patterns, but with further training, the model can learn to treat certain items as exceptions.

These models have primarily been conceived as models of early language acquisition and tested on frequent items (e.g. common verbs), where it can assumed that by adulthood, most native speakers will have extensive experience with all the items in questions, and will thus consistently recognize certain words as exceptions to the general rules. However, their behavior on new items straightforwardly generalizes to low frequency items that even adult native speakers would have relatively little direct experience with, such as attested but low frequency binomial expressions, making the prediction that these items could occupy a middle ground of partial reliance upon both general patterns (i.e. abstract knowledge) and direct experience, even in a fully developed adult grammar.

Exemplar-based computational models

A gradual trade-off is also predicted by exemplar-based computational models of language (e.g. Bod, 1998; Bod et al., 2003; Bod, 2008; Johnson et al., 2007; Demberg, 2010; O'Donnell et al., 2011). These models—which primarily focus on adult-like language representation, rather than acquisition—incorporate representations of sentence fragments of varying sizes, thus allowing for the representation of holistic multi-word expressions as well as the individual words and rules used to generate such expressions compositionally. Within these models, multi-word expressions can thus be parsed both through direct reuse and through compositional generation. The probabilities assigned to these units—the holistic expressions, the individual words, and the compositional rules—will collectively determine the relative likelihoods of reuse versus regeneration. For more frequent expressions, the probability of reusing a holistic unit will be higher, while for less frequent expressions, the probability of compositional generation will be higher. These probabilities change gradually depending on the frequency of a given expression as well as the frequencies of similar expressions. These models thus also predict a gradual trade-off between reliance on abstract knowledge for infrequent items and reliance upon direct experience for frequent items.

Nonparametric Bayesian models

The gradual trade-off theory is also supported by a nonparametric Bayesian perspective (e.g. Goldwater et al., 2009; Xu and Tenenbaum, 2007), in which expectations are influenced by both a prior probability and the incoming data. In a Bayesian model, when little data has been seen, expectations are driven by the prior probability. As more data is seen, the data becomes increasingly influential, asymptotically approaching complete dominance. For binomial expressions, abstract knowledge can be thought of as a prior probability for ordering preferences, absent any direct experience with a given expression, and each direct encounter with an expression constitutes further data. Under the Bayesian perspective, when one has little experience with an expression, expectations will be governed by abstract knowledge, but with increasing experience, the relative frequency of ordering within the experienced data will be increasingly dominant in determining expectations.

2.6.2 Advantages of our approach

While numerous models support our conclusions, the experiments presented here crucially advance the state of our understanding beyond what was previously known by providing a novel approach for using *behavioral evidence*, in conjunction with modern corpora and multivariate statistics, to quantify the contributions of abstract knowledge and direct experience. Our probabilistic model provides quantitative estimates for the effects of abstract knowledge, while corpus frequencies provide estimates for direct experience. Using multivariate regression modeling, we can directly compare the predictive strength of these two influences on behavioral data such as the results of our forced-choice and self-paced reading tasks. This approach allows us to move beyond the previous modeling-based approaches, which focused on predicting corpus data or language-wide trends. We can now investigate the trade-off between abstract knowledge and direct experience using behavioral evidence.

Additionally, the statistical techniques employed here allow us to make quantitative claims about the strength of reliance on both abstract knowledge and direct knowledge. We have seen this in a limited way so far, as we demonstrated that processing of frequently attested binomials is driven primarily by relative frequency, and only to a lesser degree by abstract knowledge. We have also predicted that there should be a gradual shift from reliance upon abstract knowledge to reliance upon relative frequency estimates as overall frequency increases; however, we cannot conclude this directly from our current data because overall frequency has only been explored as a dichotomous variable: either entirely novel or very frequent. In Chapter 6, we will look at an in-between zone of attested but not highly frequent expressions, e.g. *sunglasses and sunscreen/sunscreen and sunglasses* (1/1000th the frequency of the average attested expression in the current study). We predict that these expressions should show noticeable effects of both abstract constraints and relative frequency. Moreover, looking over a range of overall frequencies, we predict that we will see a quantitative trade-off between reliance on abstract knowledge and reliance on direct experience.

This approach to studying the trade-off between abstract knowledge and direct experience generalizes beyond the study of binomial expression ordering preferences. The cornerstone of this approach is that we are able to independently quantify the contributions of direct experience with specific expressions and abstract knowledge in the absence of direct experience. We propose that a combination of corpus frequencies and probabilistic modeling can provide such estimates for a wide range of linguistic constructions (e.g. the dative alternation [Bresnan et al., 2007] and adjective ordering [Dixon, 1982; Truswell, 2009]) allowing us to ask broad questions about the trade-off between compositional generation and the reuse of stored expressions in linguistic processing. For example, to what extent are adjective ordering preferences due to abstract rules (e.g. shape before color) versus to known collocations of highly frequent adjective sequences? The methods we have developed here make these questions accessible for future research.

2.6.3 Further predictions about language structure

Our results additionally lead to predictions about language structure. Our gradual trade-off theory predicts that items with higher overall frequency will be more likely to have relative frequency preferences that contradict abstract knowledge preferences. This prediction is analogous to the finding that more frequent verbs are more likely to be irregular (Lieberman et al., 2007): in the case of high overall-frequency items, people have enough exposure to learn idiosyncratic or abstract-knowledge-violating preferences, but in the case of low overall-frequency items, people have insufficient exposure to overcome their abstract knowledge. A further prediction follows from the results of Experiment 1, in which we found that preferences for attested items were more extreme, or polarized, than preferences for novel items. Assuming that preferences for attested items are driven primarily by relative frequency, this result predicts that as overall frequency increases, relative frequencies will become more polarized.[14]

In Chapters 3 and 4, we will see these predictions are borne out in a corpus analysis, which demonstrates that binomial expressions with higher overall frequency have relative frequencies that deviate more from abstract knowledge—in particular, by being more polarized. This finding in turn leads to further questions about the historical trajectories of binomial expression ordering preferences, and the dual roles of individuals' language processing and cultural transmission in shaping language structure (Kirby et al., 2007), which we address in Chapter 5. Thus the results presented here additionally open the door to further investigation of the mutually

[14]We did not see the analog of this result in Experiment 2: reading time were not slower in the dispreferred order and faster in the preferred order for attested than for novel expressions. Based on the results of Chapter 4, we conclude that this is due either to noise or to floor/ceiling effects on reading times.

constraining processes of synchronic language processing and diachronic language change.

Chapter 2, in full, contains material being prepared for publication in Morgan, E., & Levy, R. (2015). Abstract knowledge versus direct experience in processing of binomial expressions. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Creating a Corpus

To provide data for experiments in the remainder of this dissertation, we created a new corpus of naturally occurring binomial expressions.

3.1 Identifying binomial expressions

To populate our corpus, we extracted all *Noun-and-Noun* binomials from the parsed section of the Brown corpus—a 1 million word, hand-parsed, mixed-genre corpus of English (Marcus et al., 1999)—using the following `tregex` (Levy and Andrew, 2006) search pattern:

```
/^N/=top < (/^NN/ !$, (/,/ > =top) . ((CC <: and > =top) .
(/^NN/ > =top)))
```

The pattern extracts all sequences of a noun followed by "and" followed by a noun, with the entire sequence dominated by a noun phrase, except sequences preceded by a comma (to rule out those conjunctions at the end of a list not using an Oxford comma—it was manually verified that excluding sequences preceded by a comma excluded only items at the end of lists). This search pattern is conservative in that it also identifies some additional sequences beyond those that meet our definition of a binomial expression (e.g. sequences of the form “X and Y and Z”). However, as described below, each binomial expression is manually inspected during coding for semantic constraints, and so additional sequences can be excluded later. Proper names were also excluded during coding.

3.2 Coding for generative constraints

Binomials were coded for generative constraints as described by Benor and Levy (2006) but restricted to a subset of constraints based on the model selection described in Chapter 2. Here we describe these constraints and how they were coded, including example binomials in which each constraint applies, taken from the corpus. Appendix C contains the full constraint coding guide used by the coders.

Three metrical constraints were considered:

Length The shorter word comes first, coded as difference in number of syllables, e.g. *food and shelter*.

No final stress The final syllable of the second word should not be stressed, coded as a categorical predictor. For example, in *Japan and Holland*, *Japan* has final stress and should therefore not be in the second position.

Minimize Lapse Minimize unstressed syllables in a row, coded as difference in number of unstressed syllables surrounding “and”, e.g. *FARMS and HAYfields* preferred over *HAYfields and FARMS*.

Stress patterns were automatically extracted from the CMU Pronouncing Dictionary (The Carnegie Mellon Speech Group, 2014). When multiple pronunciations were available for a given word, a native English speaking research assistant selected the appropriate one. When no pronunciation was available, the research assistant manually determined the stress pattern. Constraint values were computed automatically from the given stress patterns.

The next constraint we consider is Frequency:

Frequency The more frequent word comes first, coded as difference in log frequency, e.g. *bride and groom*.

Frequency was calculated from the Google Books corpus, counting occurrences of each word from 1900 or later in all lowercase, all uppercase, and first letter only capitalized (Lin et al., 2012).

Finally, the corpus was hand coded for semantic constraints. Semantic constraints were coded by two independent coders (drawing from the author and two trained research assistants), and discrepancies were resolved through discussion, with the opinion of a third coder if necessary. The constraints coded for were:

Formal markedness The word with more general meaning or broader distribution comes first. For example, in *boards and two-by-fours*, *boards* are a broader class of which *two-by-fours* is one member.

Perceptual markedness Elements that are more closely connected to the speaker come first. This constraint encompasses Cooper and Ross's (1975) 'Me First' constraint and includes numerous subconstraints, e.g.: animates precede inanimates; positive words precede negative words; concrete words precede abstract words. For example, in *attractions and repulsions*, *attractions* are positive while *repulsions* are negative.

Power The more powerful or prioritized word comes first. For example, in *clergymen and parishioners*, *clergymen* have higher rank within the church.

Iconic/scalar sequencing Elements that exist in sequence should be ordered in sequence. For example, *breakfast and dinner* are generally eaten in that order.

Cultural centrality The more culturally central element should come first, e.g. *oranges and grapefruit*.

Intensity The more intense element should come first, e.g. *fear and relief*.

Pragmatic The ordering is directly influence by the local context; for example, one element has been mentioned previously. (Items for which this constraint is active are excluded from further analysis.)

All semantic constraints were coded categorically. The constraints Cultural Centrality and Intensity are additions to the repertoire of constraints proposed by Benor and Levy (2006). They describe concepts included in Benor and Levy's definitions of other constraints, but it was decided to move these concepts into their own constraints (rather than including them within the definitions of other constraints) based on conflicting definitions discovered during the coders' early experience, specifically cases where the same type of semantic relationship could be classified under multiple constraint definitions. Cultural centrality could previously fall under either Perceptual markedness or Power. Intensity could fall under Power (predicting more intense first), Iconic/scalar sequencing (in cases where elements exist on a scale of intensity, predicting more intense second; e.g. *sparks and flames*), or Formal Markedness (in cases where one item is a more intense subset of the other, also predicting more intense second; e.g. *apprehension and fear*). For the type of corpus analyses we will be performing, it is better to err on the side of a proliferation of redundant constraints rather than collapsing constraints that should be separate into one. In particular, in our predictive models, it will always be possible for two constraints to be weighted equally (e.g. if it turns out that we have separated what is underlyingly one constraint into two), but to collapse two theoretically distinct constraints into one would prevent us from ever distinguishing between them in our analyses.

Items for the which the Pragmatic constraint is active are excluded from all further analyses. This is because our goal in future corpus analyses will be to predict language-wide trends in binomial ordering preferences, but cases where the Pragmatic constraint is active are cases in which the given ordering are due to highly local influences, which would not generalize to other instances of the same binomial expression. Including these instances in our future modeling would add unnecessary noise.

3.3 Frequency counts

For each binomial type, we obtain the number of occurrences in each order ("X and Y" and "Y and X") in the Google Books ngrams corpus from 1900 and later, counting all possible combinations of lowercase and capitalized versions of each word. These values allow us to calculate both the overall (unordered) frequency of the binomial type and the observed preference/relative frequency for each order.

The Google Books ngrams corpus only contains ngrams which have appeared at least 40 times in the Google Books database. Because we are crucially interested in the relative frequencies of a binomial's two orders, it is important that we use expressions with sufficient frequencies that we can be confident in these estimates.

For example, if an item appeared in the Google Books database 41 times in one order and 39 times in the other, only the order that appeared 41 times would be included, falsely implying that this binomial is entirely frozen, when in fact it is almost perfectly balanced. To avoid this problem, for most of our analyses we consider only items with at least 1000 total instances post-1900.

3.4 Totals

Our initial corpus contained 1280 binomial tokens. After exclusions, we were left with 889 binomial expression types, of which 594 meet the criteria of having at least 1000 total occurrences.

Chapter 4

Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure

Abstract

Most models of choice in language focus on broadly applicable generative knowledge, treating item-specific variation as noise. Focusing on word order preferences in *binomial expressions* (e.g. *bread and butter*), we find meaning in the item-specific variation: more frequent expressions have more polarized (i.e. frozen) preferences. Of many models considered, only one that takes expression frequency into account can predict the language-wide distribution of preference strengths seen in corpus data. Our results support a gradient trade-off in language processing between generative

knowledge and item-specific knowledge as a function of frequency.

4.1 Introduction

A pervasive question in language processing research is how we reconcile generative knowledge with idiosyncratic properties of specific lexical items. In many cases, the generative knowledge is the primary object of study, while item-specific idiosyncrasies are treated as noise. For instance, in modeling the dative alternation, Bresnan et al. (2007) take care to demonstrate that effects of animacy, givenness, etc. on structure choice hold even after accounting for biases of individual verbs. But the verb biases themselves are not subject to any serious investigation. Here we present evidence that patterns within the item-specific variation are meaningful, and that by modeling this variation, we not only obtain better models of the phenomenon of interest, we also learn more about language structure and its cognitive representation.

Specifically, we will develop a model of word order preferences for *binomial expressions* of the form *X and Y* (i.e. *bread and butter* preferred over *butter and bread*). Binomial ordering preferences are in part determined by generative knowledge of violable constraints which reference the semantic, phonological, and lexical properties of the constituent words (e.g. short-before-long; Cooper and Ross, 1975; McDonald et al., 1993), but speakers also have idiosyncratic preferences for known expressions (Siyanova-Chanturia et al., 2011; Morgan and Levy, 2015a). Binomial expressions are a useful test case for modeling idiosyncracies because their frequencies can be robustly estimated from the Google Books n-grams corpus (Lin et al., 2012). Here we will demonstrate that explicitly modeling these expressions' idiosyncrasies both produces a better predictive model for novel expressions and also constrains our theory of these expressions' cognitive representations.

Specifically, we identify two reasons why such a model is advantageous:

1. Models identify both rules and exceptions.

One intrinsic reason that modeling idiosyncrasies is advantageous is because identifying exceptions can help identify rules. In a traditional linguistic setting (e.g. identifying rules for past tense formation), we rely upon intuition to determine what is the grammatical rule and which verbs should be treated as exceptions. In the case of binomial expressions, we likewise expect there to be exceptions to the rules, particularly for frequent expressions. For example, there is in general a strong constraint to put men before women; however, *ladies and gentlemen* is preferred over the reverse due to its conventionalized formal use. But compared with past tense formation, the rules that determine binomial ordering are far more complex and gradient, such that using traditional linguistic analysis to determine the full set of rules is not viable. In this case, we require our model not only to identify what the rules are but simultaneously to determine which expressions must be treated as exceptions. Having such a model is useful for empirical cognitive science, e.g. for disentangling the effects of people's generative knowledge from effects of their item-specific linguistic experience on language processing (Morgan and Levy, 2015a).

2. Models relate cognitive representations to language-wide structure.

As a further benefit, models can help us understand how structural properties of the language relate to people's cognitive linguistic representations. In particular, let us look at the distribution of preferences for binomial expressions taken from a subset of the Google Books corpus (described later in Creating the Corpus.) Each binomial can be assigned a preference strength corresponding to how frequently it appears in alphabetical order, from 0 (always in non-alphabetical order) to 0.5 (perfectly balanced) to 1 (always alphabetical). Binomials which always or nearly always appear in one order are said to be *frozen*. The distribution of preference strengths is shown in Figure 4.1. Preferences have a multimodal distribution with modes at the extremes as well as around 0.5. This distribution poses a challenge to standard models of binomial preferences. As we will show later, standard models predict only a single mode around 0.5. In other words, the true distribution of binomial expressions includes more frozen binomials than standard models predict. As we develop a model that accounts for this multimodal distribution, we will see that this language-structural fact puts constraints on our theories of individuals' cognitive representations of binomial expressions.

In the remainder of this paper, we first describe how we developed a new corpus of binomial expressions. We then explore a variety of models with differing levels of ability to model item-specific idiosyncrasies. Finally, we return to the issue of how these models inform us about cognitive representations of language.

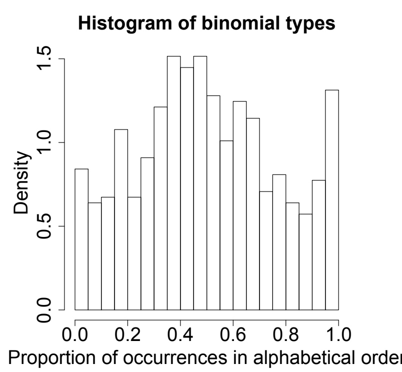


Figure 4.1: Binomial preferences are multimodally distributed in corpus data

4.2 Creating the Corpus

We extracted all *Noun-and-Noun* binomials from the parsed section of the Brown corpus (Marcus et al., 1999) using the following Tregex (Levy and Andrew, 2006) search pattern:

```
/^N/=top < (/^NN/ !$, (/,/ > =top) .
((CC <: and > =top) . (/^NN/ > =top)))
```

This pattern finds all Noun-*and*-Noun sequences dominated by a Noun Phrase which are not preceded by a comma (to exclude the final pair in lists of more than two elements), a total of 1280 tokens.

Binomials were coded for a variety of constraints, originally described by Benor and Levy (2006) but restricted to the subset determined to be most relevant for predicting ordering preferences by Morgan and Levy (2015a):

Length The shorter word (in syllables) comes first, e.g. *abused and neglected.*

No final stress The final syllable of the second word should not be stressed, e.g. *abused and neglected.*

Lapse Avoid unstressed syllables in a row, e.g. *FARMS and HAY-fields vs HAY-fields and FARMS*

Frequency The more frequent word comes first, e.g. *bride and groom.*

Formal markedness The word with more general meaning or broader distribution comes first, e.g. *boards and two-by-fours.*

Perceptual markedness Elements that are more closely connected to the speaker come first. This constraint encompasses Cooper and Ross's (1975) 'Me First' constraint and includes numerous subconstraints, e.g.: animates precede inanimates; concrete words precede abstract words; e.g. *deer and trees.*

Power The more powerful or culturally prioritized word comes first, e.g. *clergymen and parishioners.*

Iconic/scalar sequencing Elements that exist in sequence should be ordered in sequence, e.g. *achieved and maintained.*

Cultural Centrality The more culturally central or common element should come first, e.g. *oranges and grapefruits.*

Intensity The element with more intensity appears first, e.g. *war and peace.*

The metrical constraints, Length and No final stress, were automatically extracted from the CMU Pronouncing Dictionary (2014), augmented by manual annotations when necessary. Word frequency was taken from the Google Books corpus, counting occurrences from 1900 or later. Semantic constraints were hand coded by two independent coders (drawing from the first author and two trained research assistants). Discrepancies were resolved through discussion.

For each binomial, we obtained the number of occurrences in both possible orders in the Google Books corpus from 1900 or later. Items containing proper names, those with errors in the given parses, those whose order was directly affected by the local context (e.g. one element had been mentioned previously), and those with less than 1000 total occurrences across both orders were excluded from analysis, leaving 594 binomial expression types.

4.3 Models

We will develop four models of binomial ordering preferences: a standard logistic regression, a mixed-effects logistic regression, and two hierarchical Bayesian beta-binomial models. All are based on the idea of using logistic regression to combine the constraints described above in a weighted fashion to produce an initial preference estimate for each binomial. The models differ in whether and how they explicitly model the fact that true preferences will be distributed idiosyncratically around these estimates. The standard logistic regression includes no explicit representation of item-specific idiosyncrasies. The mixed-effect logistic regression includes random intercepts which account for item-specific idiosyncrasies, but which are constrained to be distributed normally around the initial prediction. The two Bayesian models assume that item-specific preferences are drawn from a beta distribution whose mean is determined by the initial prediction. In the first of these models, the concentration of the beta distribution is fixed, while in the second, it varies with the frequency of the binomial in question.

4.3.1 Evaluation

One obvious criterion for evaluating a model is how well it predicts known binomial preferences (i.e. the corpus data). For this, we report $R^2(X, \hat{X})$ as well as mean L1 error, $\frac{1}{N}\Sigma_{i=1}^{N} |\hat{x}_i - x_i|$, where $\hat{x}_i$ is the model prediction for how often binomial i occurs in a given order, and x_i is the true corpus proportion.

In addition to considering model predictions for each individual item, we want to consider the overall distribution of preferences within the language. As we will see, a model can provide good predictions for individual items without correctly capturing the language-wide multimodal distribution of these expressions' preference strengths. Thus our second desideratum will be the shape of the histogram of expression preferences.

4.3.2 Logistic regression

Logistic regression is the standard for modeling syntactic alternations, both for binomial expressions specifically (e.g. Benor and Levy, 2006; Morgan and Levy, 2015a) as well as other syntactic alternations (e.g. Bresnan et al., 2007; Jaeger, 2010). Thus we begin by constructing a baseline logistic regression model. Benor and Levy have argued that one should train such a model on binomial types rather than binomial tokens because otherwise a large number of tokens for a small number of overrepresented types can skew the results. While agreeing with this logic, we note that to train only a single instance of each type is to ignore a vast amount of data about the gradient nature of binomial preferences. As a compromise, we instead train a model on binomial tokens, using token counts from the Google Books corpus, with each token weighted in inverse proportion to how many tokens there are for that binomial type, i.e. a type with 1000 tokens will have each token weighted at 1/1000. In this way, we preserve the gradient information about ordering preferences (via the diversity of outcomes among tokens) while still weighting each type equally. The constraints described above are used as predictors. Outcomes are coded as whether or not the binomial token is in alphabetical order.

For this and all future models, predictions are generated for all training items using 20-fold cross validation. Results for all models can be seen in Figure 4.2. While the logistic regression model does a reasonable job of predicting preferences for individual items, it does not capture the multimodal distribution of preference strengths seen in the corpus data. We proceed to consider models in which item-specific idiosyncrasies are modeled explicitly.

4.3.3 Mixed-effects regression

By far the most common method in language modeling for accounting for item-specific idiosyncrasies is mixed-effects regression models (Jaeger, 2008). Formally, this model assumes that idiosyncratic preferences are distributed normally (in logit space) around the point estimate given by the fixed-effects components of the regression model.

We train a mixed-effect logistic regression on binomial tokens using the `lme4` package in R. We use as predictors the same fixed effects as before, plus a random intercept for binomial types. As described above, the fitted model now predicts a *distribution*, rather than a single point estimate, for a novel binomial. To make predictions for our (cross-validated) novel data, we sampled 1000 times from this distribution for each item. The histogram in Figure 4.2(c) shows the full sample distribution across all items. In order to generate point estimate predictions for computing L1 and R^2 (shown in Figure 4.2(b)), we take the sample median for each item, which optimizes the L1 error.

Including random intercepts improves neither our point estimates nor our language-wide distribution prediction. Apparently, the normal distribution of the random intercepts is not well suited to capturing the true distribution of binomial preferences. In particular, for a given item, the normality of random effects in *logit* space leads to predictions that are skewed towards the extremities of *probability* space.[1]

[1] An alternative method of prediction for novel items would be to take the median random intercept in logit space, i.e. to set all random intercepts to 0. This method yields results that are very similar to—but all-around slightly worse than—the original regression model.

4.3.4 Hierarchical Bayesian beta-binomial model

Having seen that normally distributed random intercepts do not adequately capture the distribution of item-specific preferences, we introduce the beta distribution as a potentially better way to model this distribution. The beta distribution, defined on the interval $[0, 1]$, has two parameters: one which determines the mean of the draws from the distribution, and one which determines the *concentration*, i.e. whether draws are likely to be clustered around the mean versus distributed towards 0 and 1. For example, for a beta distribution with a mean of 0.7, a high concentration implies that most draws will be close to 0.7, while a low concentration implies that roughly 70% of draws will be close to 1 and 30% of draws will be close to 0. When we treat the output of the beta distribution as a predicted binomial preference, a high concentration corresponds to a pressure to maintain variation while a low concentration corresponds to a pressure to regularize.

In order to incorporate the beta distribution into our model of binomial preferences, we combine the logistic regression and the beta distribution in a hierarchical Bayesian model (Gelman et al., 2013), as shown in Figure 4.3. For each item, the model determines a mean μ via standard logistic regression, using the same predictors as before. The model also fits a concentration parameter ν. These two parameters determine a beta distribution from which the binomial preference π is drawn. Observed data is drawn from a binomial distribution with parameter π.

We fit this model using the `rjags` package in `R` (Plummer, 2003). After a burn-in period of 2000 iterations, we run for 2000 more iterations sampling every 20 iterations. In order to predict novel data, we fix the point estimates for the regression coefficients $\hat{\beta}$ and the concentration parameter ν. We then sample 1000 draws of π

for each item. As with the mixed-effects model, the histogram in Figure 4.2(c) shows the full sample distribution, while point estimates (the sample median) are used to calculate L1 error and R^2 (Figure 4.2(b)).

This model performs better on L1 and R^2 than the mixed-effects model, but still worse than the initial logistic regression. The predicted histogram shows hints of the multimodal distribution seen in corpus data, but is overall too flat.

4.3.5 Beta-binomial with a variable concentration parameter

A crucial fact that we have not taken into account in previous models is the role of frequent reuse in shaping expressions' preferences. In particular, the degree to which an expression takes on a polarized preference may depend upon its frequency. We build upon the beta-binomial model in the previous section by parameterizing the concentration parameter by the frequency of the (unordered) binomial expression:

$$\nu = exp(\alpha + \beta \cdot \log(M_n)) \tag{4.1}$$

where M_n is the total number of occurrences of binomial n in both orders. Training and testing of the model are identical to above.

We find that $\beta = -0.26$ is significantly different from 0 ($t_{99} = -94; p < 2.2 \times 10^{-16}$), indicating that the concentration parameter changes significantly as a function of frequency: less frequent expressions have more dense distributions while more frequent expressions have more polarized distributions, as shown in Figure 4.5. We find that this model generates the best predictions of all our models, producing a marginally significant improvement in both L1 ($t_{593} = 1.86; p = 0.06$) and R^2 (by fold

$t_{19} = 1.76; p = 0.09$) relative to the initial logistic regression. Moreover, it correctly predicts the multimodal distribution of expression preferences.

4.4 Discussion

Overall, we found that all models made approximately similarly good best-guess predictions for binomials they weren't trained on, but the frequency-sensitive beta-binomial model was clearly superior in predicting the language-wide distribution of idiosyncratic binomial-specific ordering preferences. This model also indicates that more frequent binomials are on average more polarized.

This modeling finding supports Morgan and Levy (2015a)'s claim that generative knowledge and item-specific direct experience trade off gradiently in language processing, such that processing of novel or infrequent items relies upon generative knowledge, with reliance upon item-specific experience increasing with increasing frequency of exposure. Morgan and Levy support this claim with behavioral data, showing that empirical preferences for binomials which are completely novel depend on generative constraints while preferences for frequent expressions depend primarily on frequency of experience with each order. Our modeling results augment this argument by demonstrating that this trade-off is likewise necessary in order to predict the language-wide distribution of preference strengths. In particular, we can conceive of generative knowledge as providing a prior for ordering preferences. Under our final model, the logistic regression component serves an estimate of generative knowledge, which generates preferences clustered unimodally around 0.5. The amount of direct experience one has with an expression then modulates whether it conforms to this prior or whether it deviates. Items with low frequency have a high concentration: they maintain their variability and continue to contribute to the mode around 0.5. Items with high frequency have a low concentration: they are more likely to regularize and contribute to the modes at 0 and 1. Crucially, the inclusion of expression frequency as a predictor of the concentration of the beta distribution is necessary in order to achieve this effect in the model, demonstrating that expressions are indeed relying differentially on generative knowledge versus direct experience depending on their frequency.

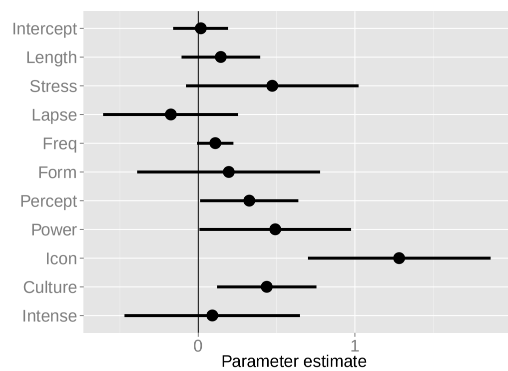
This finding fits with previous models of cultural transmission in which, in general, preferences gravitate towards the prior (Griffiths and Kalish, 2005), but with sufficient exposure, exceptions can be learned (e.g. irregular verbs; Lieberman et al., 2007). However, this raises a question which is not answered by our or others' models: why don't all expressions converge to their prior preferences eventually? We present two possibilities.

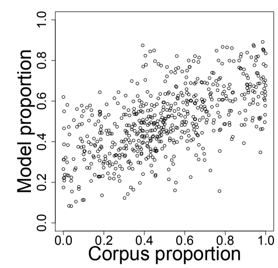
One possibility is that people's probabilistic transmission behavior differs at different frequencies. Convergence to the prior relies upon *probability matching*: people must reproduce variants in approximately the proportion in which they have encountered them. However, this is not the only possible behavior. Another possibility is that people preferentially reproduce the most frequent variant they have encountered, to the exclusion of all other variants, a process known as *regularizing*. If people's tendency to probability match versus regularize is dependent on the frequency of the expression in question (with more regularizing at high frequencies), this could produce the pattern of more polarized expressions at higher frequencies seen in our data. Another possibility is that there is some other unspecified exogenous source of pressure towards regularization, as for instance seems to be the case in child language acquisition (Hudson Kam and Newport, 2009). This pressure might be weak enough that it is overwhelmed by convergence towards the prior at lower frequencies, but can be maintained for items with high enough frequencies to have sufficient exposure to deviate from the prior. Further work is necessary to disentangle these explanations.

In addition to contributing to our understanding of binomial expression processing, we have demonstrated the value of modeling the distribution of idiosyncratic preferences in two ways. First, it has improved our ability to predict preferences for novel items, by better differentiating the rule-following training data from the exceptions. Second, this model turns an observation about language-wide structure (the multimodal distribution of preferences) into a constraint on our theory of the cognitive representation and processing of language (more polarization at higher frequencies).

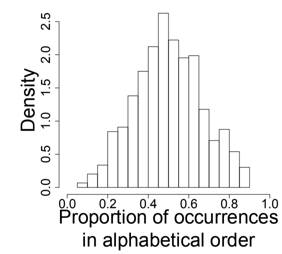
Chapter 4, in full, is a reprint of the material as it appears in Morgan, E., & Levy, R. (2015). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), 37th Annual Meeting of the Cognitive Science Society (pp. 1649-1654). Austin, TX: Cognitive Science Society. The dissertation author was the primary investigator and author of this paper.

Logistic regression

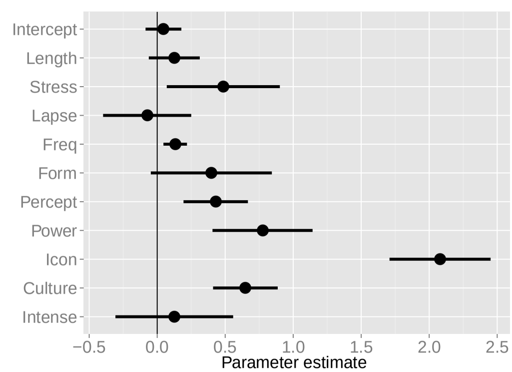


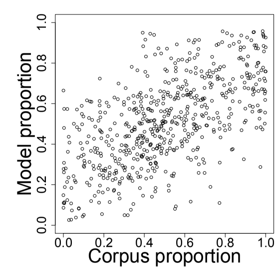


L1= 0.169 (0.006)
$R^2 = 0.368$ (0.021)

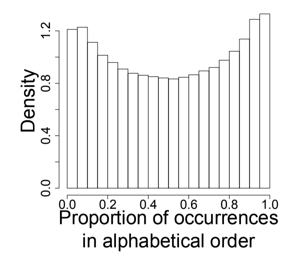


Mixed-effects regression with random intercept





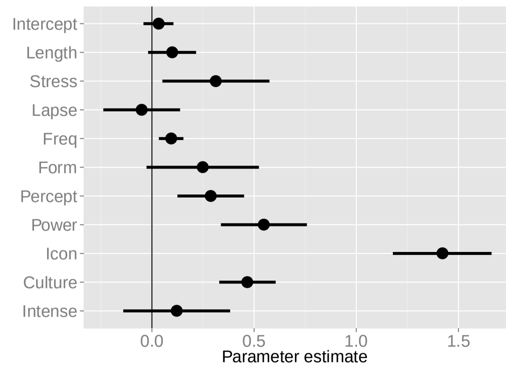
L1= 0.173 (0.006)
$R^2 = 0.355$ (0.022)

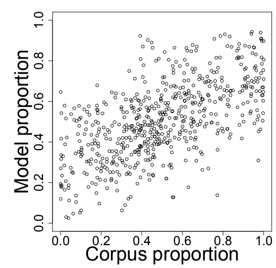


(a) (b) (c)

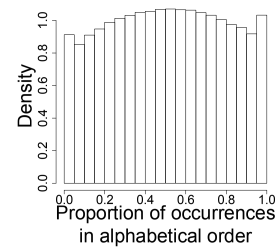
Figure 4.2: For each of our four models, we display: (a) Parameter estimates for the logistic regression component. Dots show point estimates with bars indicating standard errors. (b) Predictions for each item, as well as mean by-type L1 error and R^2 with by-fold standard errors. (c) Language-wide predicted distribution of preference strengths.

Beta-binomial model

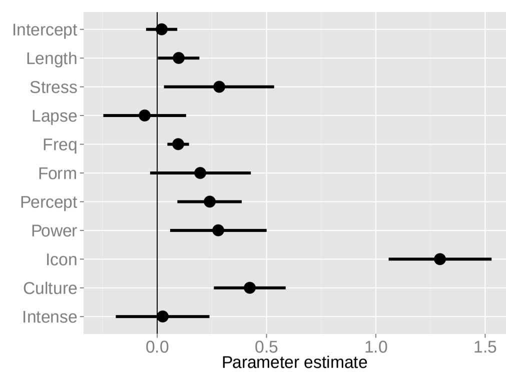


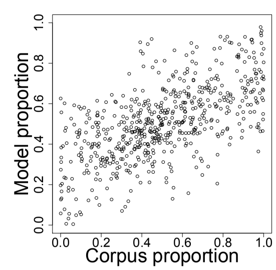


L1= 0.170 (0.003)
$R^2 = 0.367$ (0.020)

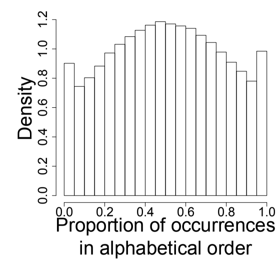


Beta-binomial model with variable concentration



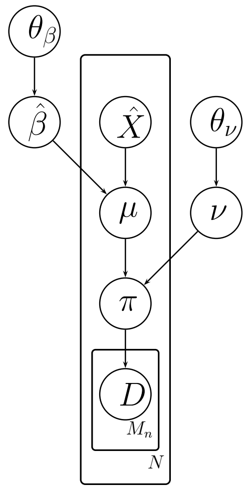


L1= 0.166 (0.003)
$R^2 = 0.381$ (0.021)



(a) (b) (c)

Figure 4.2: For each of our four models, we display. . . (Continued.)



N: # unordered binomial types
M_n: Frequency of binomial n
$\hat{X}$: Predictors (i.e. generative constraints)
$\hat{\beta}$: Regression coefficients
θ: Uninformative priors

$$\mu = \frac{1}{1 + e^{-\hat{X} \cdot \hat{\beta}}}$$

$\nu \sim exp(\theta_\nu)$
$\pi \sim \text{Beta}(\mu, \nu)$
$D \sim \text{Binomial}(\pi, M_n)$

Figure 4.3: Our initial hierarchical Bayesian beta-binomial model. The set of nodes culiminating in μ implements a standard logistic regression. The output of this regression determines the mean of the beta distribution (with ν determining the concentration) from which π and finally the observed data itself is drawn.

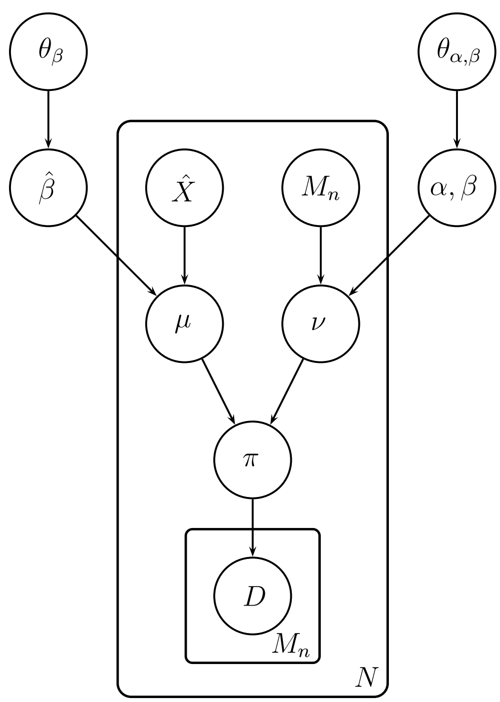


Figure 4.4: Hierarchical Bayesian beta-binomial model with variable concentration parameter

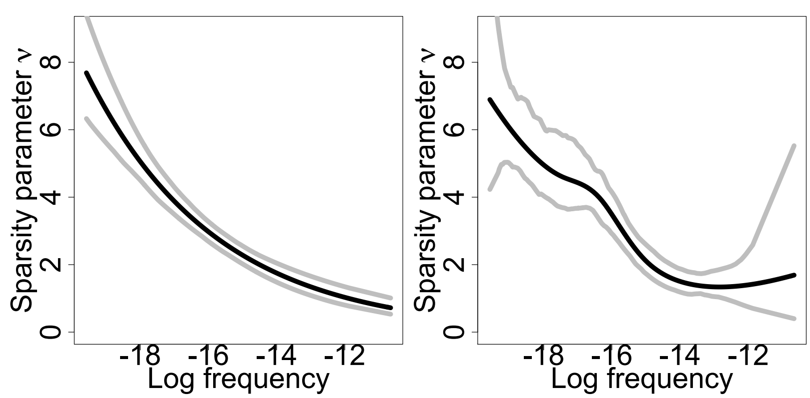


Figure 4.5: Concentration parameter ν as a function of frequency with 95% confidence intervals. (Left) Parameterization given in Eq. 1. (Right) Alternate parameterization with cubic splines, for comparison.

Chapter 5

Frequency-dependent regularization in iterated learning

Abstract

Binomial expressions are more *regularized*—their ordering preferences (e.g. "bread and butter" vs. "butter and bread") are more extreme—the higher their frequency. Although standard iterated-learning models of language evolution can encode overall regularization biases, the stationary distributions in these standard models do not exhibit a relationship between expression frequency and regularization. Here we show that introducing a frequency-*independent* regularization bias into the data-generation stage of a 2-Alternative Iterated Learning Model yields frequency-*dependent* regularization in the stationary distribution. We also show that this model accounts for the distribution of binomial ordering preferences seen in corpus data.

5.1 Introduction

Languages are shaped both by the cognitive architectures of individual speakers and by the process of cultural transmission that acts across generations. In this paper we ask how these two factors jointly contribute to a key dichotomy in language structure: the trade-off between broadly-applicable compositional knowledge and knowledge of item-specific idiosyncrasies. Specifically, we take up the case of frequency dependence in a type of regularization behavior we refer to as *Entropy-Reducing-* or *ER-regularization*, in which the preference for a given form among multiple alternatives becomes more extreme or polarized. Although ER-regularization is a well-attested phenomenon in statistical learning, *frequency-dependent* ER-regularization is not. Here we demonstrate that frequency dependence of ER-regularization can arise as an emergent property of a frequency-*independent* regularization bias in language production, combined with the bottleneck effect of cultural transmission.

Item-specific idiosyncrasies (i.e exceptions to the rules) are well known to be frequency-dependent. For example, more frequent verbs are more likely to have irregular conjugations (Lieberman et al., 2007). More recently, Morgan and Levy (2015b) have demonstrated a different type of frequency-dependent idiosyncrasy at the level of multi-word phrases, specifically *binomial expressions* of the form "X and Y" (Cooper and Ross, 1975; Benor and Levy, 2006). Word order preferences for these expressions are gradient; for example, "radio and television" is preferred to "television and radio" in a 63 to 37 ratio, while "bread and butter" is preferred to "butter and bread" 99 to 1 (Lin et al., 2012). These ordering preferences are partially determined by productive, violable constraints, e.g. a constraint to put shorter words before longer words. But these expressions are also subject to learned item-specific idiosyncrasies, e.g.

despite a generally strong constraint to put men before women, "ladies and gentlemen" is preferred over "gentlemen and ladies". In addition to the possibility of the complete reversal of compositional preferences, item-specific idiosyncrasies can also be gradient, e.g. a binomial whose compositional preference predicts a 60/40 distribution might instead be used in a 90/10 ratio. Morgan and Levy (2015b) showed that, as is the case with irregular verbs, the distribution of idiosyncrasies in binomial ordering preference is frequency-dependent: more frequent binomial expressions deviate more from compositional preferences. In particular, more frequent binomials tend to have more polarized preferences, i.e. they tend to be more ER-regularized.

5.1.1 What is regularization?

The term *regularization* has a complicated history in the literature on language development and language evolution. In general, *regularization* refers to making a language more systematic. However, systematicity can take multiple forms. For example, it can involve making the language more consistent by more consistently using one form over another (i.e. in a case of free variation, converging on a single form); it can involve standarizing variation (i.e. moving from free variation to conditioned variation); or it can involve switching a given item from an idiosyncratic pattern to a pattern that better conforms to other items in the language (e.g. when an irregular verb becomes regular).

While the idea of regularization as a mechanism of language change has a long history, people have often not been consistent or specific with regards to what form(s) of regularization they are discussing. Bickerton (1981) is the first author generally cited with regards to the role of regularization in language change, although he doesn't use the word *regularization* himself. He argues that children's regularization is an important mechanism of language change, particularly for forming consistent languages out of pidgins—in particular, that children exposed to inconsistent input data will draw upon innate knowledge of grammar to standardize the variation they are exposed to. More recently, the term *regularization* has been popularized in large part by Newport and colleagues (e.g. Newport, 1999; Hudson Kam and Newport, 2005). Though disagreeing with Bickerton on the role of innate grammatical knowledge, they broadly concur that children are rampant regularizers and that this regularization is an important driver of language change. Newport and colleagues use the term *regularization* to refer to multiple of the types of regularization described above. For example, Hudson Kam and Newport (2005) give as one example of regularization an adult second-language learner of German who produces the definite and indefinite articles corresponding to one noun class consistently across all noun classes, i.e. this speaker regularizes by using a single set of articles more consistently, but their productions are not "regular" in the sense of conforming to the language's rules. But later on, they also use regularization to refer to an instance of creating conditioned variation. In their Experiment 2, children are exposed to a novel language with nouns that are inconsistently either marked with determiners or not. One child is described as regularizing her noun production by introducing a conditioning factor: she consistently produces nouns with determiners in transitive sentences and without determiners in intransitive sentences. In this case, the relevant notion of regularization is not whether a single a form is used more consistently, but rather whether the variation between two forms is conditioned in a principled way.

In the case of binomial expressions, we see regularization in the form of more consistently using a single form over another (e.g. consistently using "bread and butter" over "butter and bread"). We can formalize this notion of regularization as a reduction in entropy of the probability distribution over alternates. Thus, to distinguish this type of regularization from other uses of the same term, we will henceforth refer to the phenomenon in question as *Entropy-Reducing-* or *ER-regularization*.

ER-regularization is a well-established phenomenon in statistical learning. In a variety of tasks, both linguistic and non-linguistic, in which participants learn and reproduce probability distributions over alternates, both children and adults tend to ER-regularize their productions (Hudson Kam and Newport, 2005; Reali and Griffiths, 2009; Ferdinand et al., 2014). For example, Reali and Griffiths (2009) found that when exposed to two labels for a novel object, subjects on average reproduced the more frequent label *even more frequently* than that label was seen in training. Although this tendency was weak, they demonstrated that even such a small bias towards ER-regularization can have significant long-term impacts, as the bias acts across successive generations to shape language over time.

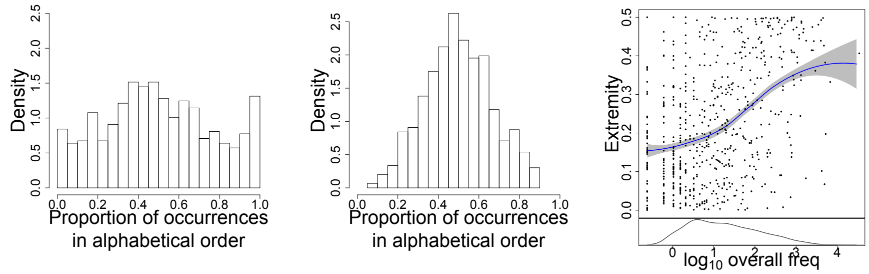
Although some standard iterated-learning theories predict across-the-board ER-regularization (in particular Reali and Griffiths, 2009), they do not predict *frequency-dependent* ER-regularization. Thus Morgan and Levy's finding is unexpected, and poses a challenge to models of language evolution. In this paper, we review the key data (Section 5.2) and show that standard iterated-learning models fail to account for frequency-dependent ER-regularization (Section 5.3). We then show that frequency-dependent ER-regularization emerges when the data-generation stage of a standard iterated learning model is augmented with a frequency-independent regularization bias, and that this augmented model accounts for the empirical distribution of binomial ordering preferences (Section 5.4). Section 5.5 concludes.

5.2 Dataset

We take advantage of a uniquely appropriate real-world data set: Morgan and Levy (2015b)'s corpus of 594 binomial expression types hand-annotated for a range of semantic, phonological, and lexical constraints known to affect binomial ordering preferences, and with frequencies of each ordering extracted from the Google Books corpus (Lin et al., 2012). Morgan and Levy also reported a model estimating the quantitative compositional ordering preference for each binomial expression, as expected on the basis of the above constraints (independent of actual occurrence frequencies). The dataset and model thus give us three key measures for these expressions:

- The *overall (unordered) frequency* of an expression: freq("X and Y")+freq("Y and X")
- The *observed preference* for occurrence in a given order, expressed as a number between 0 and 1: freq("X and Y")/(freq("X and Y")+freq("Y and X"))
- The *compositional preference* for occurrence in a given order, expressed as a number between 0 and 1, given by Morgan and Levy's model.

Observed preferences are multimodally distributed, with modes at the extremes as well as around 0.5 (Fig. 5.1(a)). Crucially, this pattern is not predicted by compositional preferences, which predict only a single mode (Fig. 5.1(b)). This pattern reflects the key generalization to be accounted for in the present paper: that expressions with



(a) (b) (c)

Figure 5.1: Results from Morgan and Levy (2015). (a) Histogram of binomial types' observed preferences. (b) Histogram of binomial types' compositional preferences. (c) We define an expression's *extremity* as the absolute difference between its observed preference and 0.5. More frequent expressions have more extreme/ER-regularized preferences; see Morgan & Levy (2015) for alternative ways to quantify extremity that yield similar conclusions. Lower panel shows density of overall frequency counts (scaled as described in Section 5.4.2). The distribution is non-Zipfian because the corpus is restricted to binomial types with at least 1000 occurrences in the Google Books corpus to ensure accurate observed preference estimates.

higher overall frequency diverge most from compositional preferences, and are more ER-regularized (Fig. 5.1(c)).

5.3 ER-Regularization is Frequency-Independent in Standard Iterated Learning

We use 2-alternative iterated learning (Reali and Griffiths, 2009; Smith, 2009) to simulate the evolution of binomial expressions over generations of speakers. A learner hears N tokens of a binomial expression, with x_1 of them in a given order—we use alphabetical order as a neutral reference order—and then infers a hypothesis $\theta_1 \in [0, 1]$ which is the proportion of time a binomial should be produced in alphabetical order.

The learner then generates new data using θ_1.

The prior probability $P(\theta_1)$ of a binomial being preferred in a given order can be expressed using the beta distribution. We can treat the compositional preference as a form of prior knowledge of ordering preferences for a binomial. To incorporate this prior knowledge, we use a parameterization of the beta distribution with a parameter μ that determines the mean of draws and a concentration parameter ν that determines how tightly clustered around the mean those draws are. (ν can also be thought of as reflecting how confident in the prior we are, e.g. $\nu = 10$ would indicate confidence equivalent to having seen ten instances of a given binomial expression type before.) Under this parameterization,

$$P(\theta_1) = \frac{\theta_1^{\mu\nu-1}(1-\theta_1)^{(1-\mu)\nu-1}}{\mathrm{B}(\mu\nu,(1-\mu)\nu)} \tag{5.1}$$

where B is the beta function. Because μ represents compositional ordering preferences, it varies for each binomial, and is set according to Morgan and Levy's model. All learners are assumed to have the same μ value for a given binomial. ν is constant for all binomial expressions for all learners, and is a free parameter. Given θ_1, data is generated binomially:

$$P(x_1|\theta_1) = \binom{N}{x_1}\theta_1^{x_1}(1-\theta_1)^{N-x_1} \tag{5.2}$$

We define a chain of learners under this model by initializing a single learner with some hypothesis. This first generation produces N utterances according to the distribution defined in Eq. 5.2. The learner in the next generation applies Bayes rule and chooses a hypothesis from the resulting posterior distribution over hypotheses.

This process continues iteratively.

Reali and Griffiths (2009) have demonstrated that ER-regularization occurs in iterated learning models with sparse priors (i.e. those that favor hypothesis close to 0 and 1); given our parameterization of the beta distribution, these are hypothesis with $\nu < 2$. However, this ER-regularization is not dependent on the expression's overall frequency. We demonstrate this by modeling chains of learners with different values of N. We model a single binomial expression with compositional preference $\mu = 0.6$. We explore different values of ν, specifically $\nu = 1$ (a sparse prior) and $\nu = 10$ (a dense prior), and values of $N = 10, 100, 200, 500$. For each combination of ν and N, we approximate the distribution over expression preferences by running 100 chains of learners for 500 generations each and taking the hypothesis of the final generation in each chain, except in the $N = 500$, $\nu = 1$ case where chains are run for 1000 generations each because convergence to the stationary distribution is slower for higher values of N. (For all chains in all simulations in this paper, we initialize $\theta_1 = 0.5$ and use MAP estimation to choose θ_1 in each new generation. Results are qualitatively similar under posterior sampling.) ER-regularization in the resulting distributions does not depend on N (Fig. 5.2, dashed lines; the small apparent sensitivity to N for a given value of ν is due to the finite number of chains used in the simulations.) The number of times an expression is seen in each generation does not affect its ultimate degree of ER-regularization.

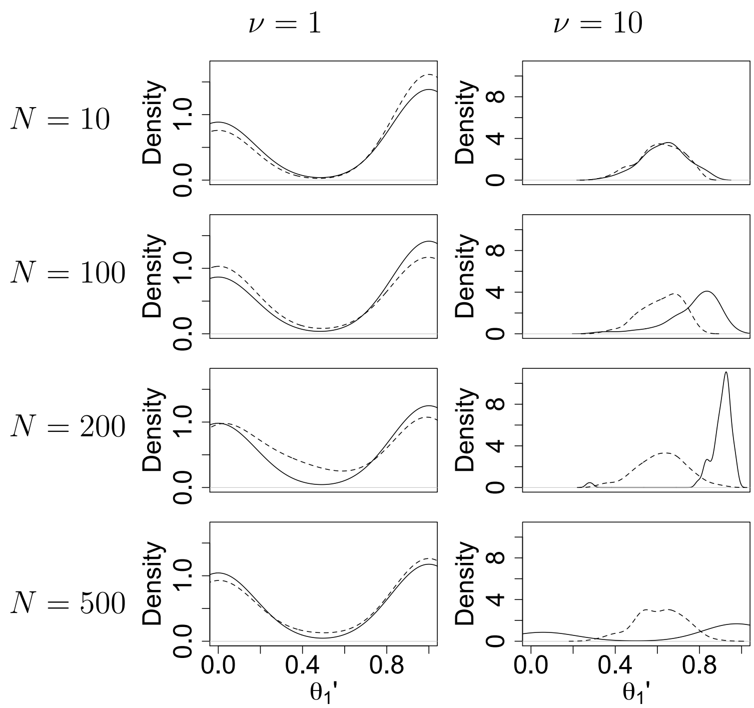


Figure 5.2: Simulated distribution of binomial ordering preferences for a single expression type with $\mu = 0.6$ in a standard 2-Alternative Iterated Learning Model (dotted lines) and one with an explicit regularization bias in data production of $\alpha = 1.1$ (solid lines). Note that $\theta'_1 = \theta_1$ in the standard model. ER-regularization depends upon N only in the model with an explicit regularization bias.

5.4 Emergence of Frequency-Dependent ER-Regularization in Iterated Learning

The standard 2-Alternative Iterated Learning Model does not predict frequency-dependent ER-regularization. We now demonstrate that we can predict frequency-dependent ER-regularization by introducing a frequency-*independent* regularization bias into our model. Under this model, frequency-dependent ER-regularization is an emergent property of the interaction of the frequency-independent regularization bias with the bottleneck effect of cultural transmission.

We augment the learning and transmission process as follows. After hearing data, the learner chooses a hypothesis θ_1 as before, then applies a regularization function to produce a new hypothesis θ_1', then generates data from θ_1'.

To model this process, we need to choose a regularization function $f : [0,1] \to [0,1]$ that captures the notion of ER-regularization. In particular, the values 0, 0.5, and 1 should be fixed points—because 0 and 1 don't permit further regularization, and 0.5 can't be regularized without making an arbitrary choice about which direction to regularize in—while values in $(0, 0.5)$ should be pushed closer to 0 and values in $(0.5, 1)$ should be pushed closer to 1. Moreover, we would like the regularization function to be parameterizable to control the degree of regularization, i.e. how strongly values are pushed towards 0 and 1. To fit these criteria, a mathematically convenient function to choose is the regularized incomplete beta function (equivalently, the cumulative distribution function of the beta distribution), restricted to be symmetric such that it has a single free parameter α:

$$f(x;\alpha) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\alpha-1}\,\mathrm{d}t}{\mathrm{B}(\alpha,\alpha)} \tag{5.3}$$

As shown in Fig. 5.3, the bias parameter α controls strength of regularization. When $\alpha = 1$, this is the identity function, i.e. no explicit regularization is added. As α increases, the regularization bias grows stronger.

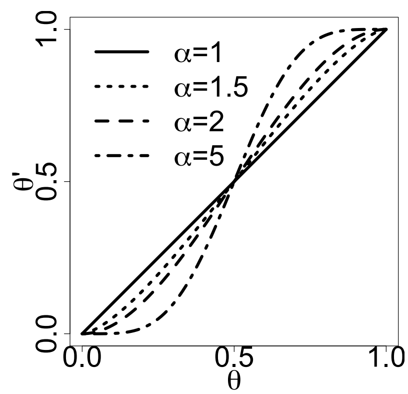


Figure 5.3: Regularization function with different values of α

5.4.1 Results: Frequency-dependent ER-Regularization

We repeat the simulations from above using a non-trivial regularization bias $\alpha = 1.1$. (This value was chosen semi-arbitrarily to demonstrate the effects of a small but non-trivial bias parameter. Section 5.4.2 shows the effect of manipulating this parameter.) In these simulations, we see frequency-dependent ER-regularization in the case with a dense prior (Fig. 5.2). Although the regularization bias itself is frequency-independent, frequency-dependence emerges from the interaction of the regularization bias with the process of cultural transmission: At lower frequencies, there is not sufficient data for the regularization bias to overcome the prior. At higher frequencies, the regularization bias becomes increasingly dominant as there is increasingly enough data for the effects of this bias to be carried across generations. Even a relatively weak bias ($\alpha = 1.1$) can produce noticeable ER-regularization when compounded across generations. However, the prior always continues to exert some influence; thus, even the highest frequency expressions do not become completely ER-regularized (i.e. they maintain some amount of variation).

Another linguistically accurate property of this model is that for sufficiently high values of N, the distribution over hypotheses includes a mode on the opposite side of 0.5 from the prior. Thus the model correctly predicts that at high enough frequencies, an expression can become idiosyncratically preferred in the opposite of its compositionally predicted direction (as in “ladies and gentlemen”).

5.4.2 Results: Simulating corpus data

Having demonstrated that our augmented model produces frequency-dependent ER-regularization, we now show that it additionally predicts the true language-wide distribution of binomial preference strengths seen in corpus data. The target distribution to be accounted for is shown in Fig. 5.1(a).

We take the true properties of each binomial expression in the corpus: its compositional preference determines μ and its overall frequency determines N. We scale overall frequency counts based on estimated lifetime exposure to 300 million total words (Levy et al., 2012, footnote 10). The resulting distribution of values N is shown in Fig. 5.1(c). For each binomial in the corpus, we approximate the stationary distribution by modeling 10 chains of learners for 200 generations each and take the hypothesis θ_1' of the final generation of each chain.

Our model has two free parameters, ν and α. We model the corpus data as described above for a range of values of both of these parameters. As shown in Fig. 5.4, our model displays a trade-off between the prior and the regularization bias as a function of these parameters. At a range of appropriately balanced values (e.g. $\nu = 10, \alpha = 1.1; \nu = 15, \alpha = 1.3; \nu = 20, \alpha = 1.5$), our model correctly predicts the multimodal distribution of corpus data as seen in Fig. 5.1(a).

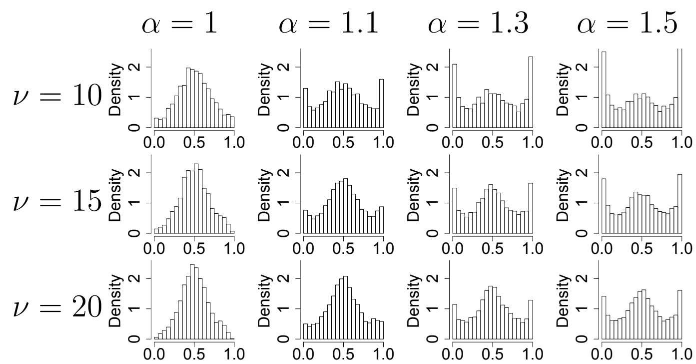


Figure 5.4: Predicted distribution of θ_1'. We see a trade-off between effects of the prior and the regularization bias. When the prior is stronger (high ν, low α), we see a unimodal distribution of preferences, similar to Fig. 5.1(b). When the regularization bias is stronger (low ν, high α), we see too much ER-regularization. At appropriately balanced values of α and ν, we see the correct multimodal distribution of preferences as seen in corpus data (Fig. 5.1(a)).

5.5 Conclusion

We have demonstrated that a frequency-independent regularization bias in data generation, combined with cultural transmission, can produce the pattern of frequency-dependent ER-regularization of binomial ordering preferences seen in corpus data. Cultural transmission creates frequency-dependence by introducing a bottleneck effect (i.e. a limit on the number of tokens of a binomial seen by each generation) that favors prior knowledge at lower frequencies while allowing the regularization bias to be increasingly well transmitted at higher frequencies. This finding sheds light on the origins of linguistic structure in two important ways: one, it confirms earlier demonstrations of a bias to ER-regularize when learning stochastic linguistic items. Second, it shows that this bias can apply equally across all levels of frequency, but that the distribution of idiosyncrasy seen in the language emerges from the interaction of individuals' cognitive biases with the bottleneck effect of cultural transmission. Additionally, we have expanded the empirical coverage of iterated learning models, showing that they can account not only for qualitative generalizations in natural language and data from laboratory experiments, but also detailed patterns of naturalistic corpus data. As we hope to have shown, binomial ordering preferences are a particularly suitable test case for iterated learning models, at once theoretically interesting, data-rich, and computationally tractable.

Chapter 5, in part, contains material being prepared for publication in Morgan, E., & Levy, R. (2016). Frequency-dependent regularization in iterated learning. The 11th International Conference on the Evolution of Language. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Abstract knowledge versus direct experience: Evidence of gradience

6.1 Introduction

As demonstrated in Chapter 2, processing of binomial expressions relies both upon abstract generative knowledge and upon our previous direct experience with particular expressions. In Chapter 2 we demonstrated a trade-off between these knowledge sources as a function of expression frequency: novel binomial expressions' ordering preferences were determined by abstract knowledge, while frequently attested expressions' preferences were determined largely by direct experience (with abstract knowledge potentially still active but playing a smaller role). In the Discussion section, we proposed that reliance upon these two knowledge sources varies gradiently as a function of overall expression frequency; however, the experiments presented there only tested the far ends of the overall frequency spectrum. Here we explicitly test the gradience of this trade-off by looking at expressions that continuously span the spectrum of overall frequency from novel to highly frequent.

6.2 Experimental Materials

To develop experimental materials, we chose binomial expressions from the corpus described in Chapter 3. As before, we can generate estimates of abstract knowledge of ordering preferences for these expressions using probabilistic modeling, and estimates of relative frequency can be collected from corpus frequencies.

Our primary goal in choosing materials for this experiment was to choose items that continuously spanned the possible range of overall frequencies. As a secondary goal, we wanted to choose a set of items for which abstract knowledge model predictions and relative frequencies were as uncorrelated as possible, in order to maximize statistical power for detecting independent effects of these two knowledge sources. (As will be explained later, this second goal was not achieved, but it turned out not to be a problem.) To achieve these goals, we divided the overall frequency space spanned by our corpus items into eight equally sized bins and chose six items from each bin for inclusion in the experiment. The items in each bin were chosen by pseudorandomly sampling items from each bin and selecting a set that minimized the absolute value of the correlation between abstract knowledge and relative frequency, subject to the following constraints:

- No word was included in more than one item.
- We must be able to write a natural-sounding sentence for the item. (This constraint is necessary because in a corpus of naturally occurring exemplars, some items will of course be very context-specific and would sound bizarre in an

experimental context.)

However, after running the experiment, we discovered that the items had been chosen using a version of our abstract knowledge model with a bug that substantially changed the model predictions. The data as presented here have been reanalyzed with a corrected version of the model. The existence of the bug during model selection means that the selection of items was essentially random with respect to the second stated goal (of dissociating abstract knowledge and relative frequency). Fortunately, as we will see below, we appear to nonetheless have had sufficient statistical power to detect the relevant effects. The distribution of these items is visualized in Figure 6.1.

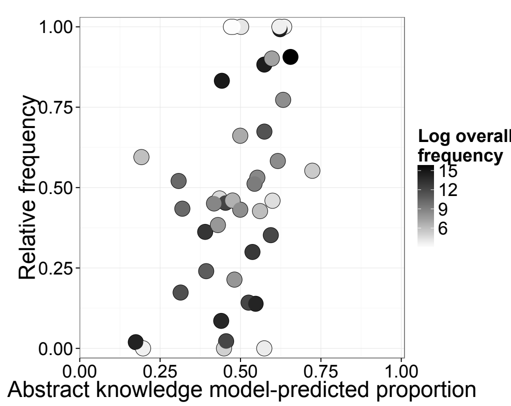


Figure 6.1: Distribution of abstract knowledge model predictions and relative frequencies of experimental items, excluding 6 novel items (for which relative frequencies cannot be calculated). x values are the abstract knowledge modelÕs prediction for how often the item will appear in alphabetical order. y values are how often the item was preferred in that order. Shading shows overall frequencies.

As before, we constructed sentence contexts for each item. Abstract knowledge model predictions were obtained from the final model described in Chapter 4 (specifically, the value π from the beta-binomial model with a variable concentration parameter), and relative frequencies were obtained from the Google Books ngram corpus as described in Chapter 3.

With these materials, we carried out a forced-choice preference experiment, analogous to Experiment 1 in Chapter 2. No filler items were included (again analogous to Experiment 1 in Chapter 2).

6.3 Methods

6.3.1 Participants

88 native English speakers (mean age=32 years; sd=10) participated. Participants were recruited through Amazon Mechanical Turk, restricted to people connecting to the website from within the United States, and were paid 75 cents. Participants were asked to report their "Native language (what you learned to speak with your mother as a child)". Those who did not report English among their native languages were excluded.

6.3.2 Procedure

The procedure was identical to that of Experiment 1 in Chapter 2. The experiment typically took 10-15 minutes.

6.4 Results

We begin by analyzing the data by dividing the experimental items into bins by overall frequency. We predict that looking across the bins, we will see a gradual shift from reliance upon abstract knowledge in the novel and low frequency bins, through reliance upon both abstract knowledge and relative frequency in the medium frequency bin, to reliance primarily upon relative frequency in the high frequency bin. Next we will do an all-items analysis by combining all items in a single regression model in which we test for the interaction of overall frequency with both abstract knowledge and relative frequency. We predict a negative interaction with abstract knowledge (indicating that abstract knowledge is decreasingly important with increasing overall frequency) and a positive interaction with relative frequency (indicated that relative frequency is increasingly important with increasing overall frequency).

6.4.1 Binned analyses

We divide items into bins as follows: The 6 entirely novel items constitute the Novel bin. The remaining items are split evenly into low-, medium-, and high-frequency bins of 14 items each. Within each bin, we analyze the data using mixed-effects logistic regression, analogous to that used in Experiment 1 of Chapter 2. Our dependent variable is the preferred order, coded as alphabetical or non-alphabetical. Our independent (fixed-effects) predictors are abstract knowledge model predictions and relative frequency, both centered around 0.5. (Relative frequency is not included in the analysis of novel items because it is not defined for these items.) We predict that looking across bins from novel through high-frequency, abstract knowledge will decrease in strength as a predictor of preferences, while relative frequency will increase. Following Barr et al. (2013), we use the maximal random effects structures for subjects and items justified by the experimental design: by-subject and by-item intercepts, and by-subject slopes for abstract knowledge and relative frequency.

Model results for the predictors of interest are given in Table 6.1 and visualized in Figures 6.2 and 6.3. For novel items, we see a significant effect of abstract knowledge

Table 6.1: Coefficient estimates for key predictors in binned regression analyses.

	Novel		LF		MF		HF	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
Abs.know.	5.58	0.002	2.42	0.02	1.69	0.39	0.45	0.82
Rel.freq.	NA	NA	0.98	0.01	3.327	0.003	6.48	< 0.0001

on preferences. For low-frequency items, we see significant effects of both abstract knowledge and relative frequency, with abstract knowledge being the stronger predictor, indicated by its larger coefficient size. (Note that we can directly compare coefficient estimates for the two predictors because they are both on the $[0, 1]$ scale.) For medium frequency items, we see a significant effect of relative frequency on preferences. Although the abstract knowledge predictor does not reach significance in this analysis, its coefficient has nontrivial magnitude; we suspect that abstract knowledge continues to play a role in determining preferences for these items but that it does not reach significance in this analysis due to lack of power. For high frequency items, relative frequency is a significant predictors of preferences, with little evidence of any effect of abstract knowledge.

Comparing across bins, we see that as overall frequency increases from novel through high-frequency items, the role of abstract knowledge decreases, indicated by decreasing coefficient estimates and increasing p values. At the same time, the role of relative frequency increases, indicated by increasing coefficient estimates and decreasing p values.

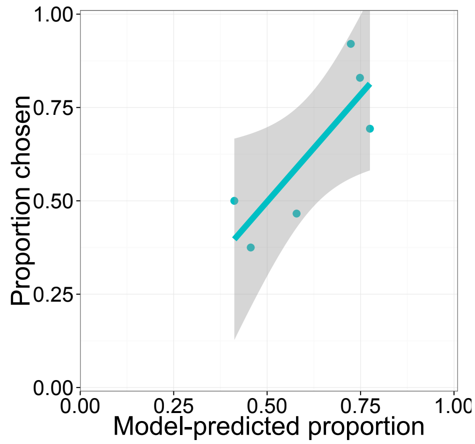


Figure 6.2: Ordering preferences for novel subset of items by abstract knowledge model predictions. x values are the abstract knowledge modelÕs prediction for how often the item will appear in alphabetical order. y values are how often the item was preferred in that order. Line shows best linear fit on the by-items aggregated data. Abstract knowledge is a significant predictor of preferences for these expressions.

6.4.2 All-items analysis

Next we analyze all items together, again using mixed-effects logistic regression. Again, our dependent variable is preferred order. Our independent variables are:

- **Abstract knowledge** (as defined above; centered around 0.5)
- **Relative frequency** (as defined above; centered around 0.5, with values for novel items set to 0 after centering)
- **Overall frequency** is estimated from the Google Books ngrams corpus as described in Chapter 3. The predictor is logged and standardized. We include two-way interactions of overall frequency with both abstract knowledge and relative frequency.

Table 6.2: Coefficient estimates for key predictors in the all-items regression analysis.

Predictor	Coeff.	p
Abs.know.	2.00	0.01
Rel.freq	3.06	< 0.001
Overall freq*Abs know	-0.59	< 0.001
Overall freq*Rel freq	0.61	< 0.001

Following Barr et al. (2013), we use the maximal random effects structures for subjects and items justified by the experimental design: by-subject and by-item intercepts, and by-subject slopes for abstract knowledge, relative frequency, overall frequency, overall frequency by abstract knowledge, and overall frequency by relative frequency.

Model results are given in Table 6.2. We see a significant negative interaction of overall frequency with abstract knowledge— indicating that as overall frequency increases, abstract knowledge plays less of a role in determining preferences—and a significant positive interaction of overall frequency with relative frequency—indicating that as overall frequency increases, relative frequency plays a greater role in determining preferences.

6.5 Discussion

In this chapter we have shown evidence for the gradience of the trade-off between abstract knowledge and direct experience in processing of binomial expressions. Specifically, we have demonstrated using a forced-choice preference task that as overall frequency increases, abstract knowledge gradiently decreases as a predictor of expressions' preferences and relative frequency gradiently increases. Moreover, we have replicated the results of the experiments presented in Chapter 2, again demonstrating that preferences for novel items depend on abstract knowledge, while preferences for highly frequent items depend primarily on direct experience.

We propose that this gradient trade-off provides a rational solution for the needs for both robustness and efficiency in language processing. Abstract knowledge provides robustness, allowing listeners to process never-before- or rarely-seen expressions. But the additional ability to rely upon direct experience allows the language processor to become appropriately specialized for higher frequency inputs, thereby increasing efficiency. In other words, the trade-off between these two knowledge sources allows the language processor to strike a balance between broadly applicable abstract knowledge and highly specific knowledge of individual items.

In future work, we will also test the gradience of this trade-off using self-paced reading.

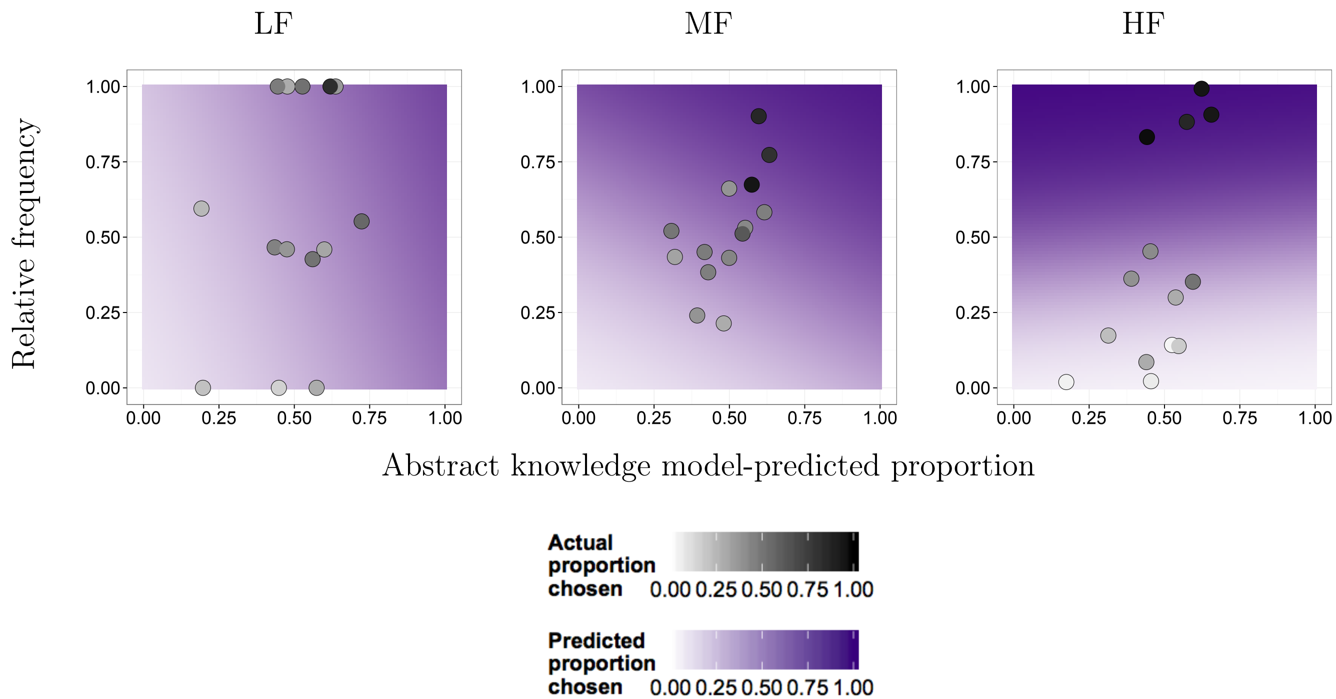


Figure 6.3: Results for low-, medium-, and high-frequency subsets of items. Each point represents an item. x values are the abstract knowledge model's prediction for how often the item will appear in alphabetical order. y values are the item's relative frequency of appearing in that order. Points' shading (white to black) shows often the item was preferred in that order. Background shading (white to purple) shows the best-fit model (Table 6.1) prediction for how often the item was preferred in that order. For low frequency items, abstract knowledge is the stronger predictor (as indicated by the horizontal gradient). For medium frequency items, relative frequency becomes a stronger predictor (diagonal gradient). For high frequency items, relative frequency is the strongest (vertical gradient).

Chapter 7

Conclusion

In this dissertation I have argued that binomial expressions are processed using a combination of generative and item-specific knowledge which varies gradiently depending on the expression's frequency. This processing strategy both is motivated by and serves to perpetuate a frequency-dependent balance of compositionality and idiosyncrasy in binomial ordering preferences, wherein more frequent expressions are gradiently more idiosyncratic—in particular, tending to be more regularized. This gradient frequency-dependent trade-off between generative and item-specific knowledge is rationally predicted as a consequence of the fact that speakers have *finite linguistic experience.*

In addition to informing us about properties of binomial expressions specifically, the results presented here generalize to promote broader claims about the trade-off between generativity and item-specificity in language generally. In particular, using a multi-word construction as a test case provides a stronger test for the use of item-specific knowledge in language processing generally, as compared to a single-word case such as the verb inflections described in Chapter 1. Furthermore, I hope to have demonstrated the utility of conceiving of idiosyncrasy itself as a gradient rather than binary phenomenon. In the case of binomial expressions, a gradient notion of idiosyncrasy allows us to recognize and quantify regularization as a form of idiosyncrasy; in contrast, a binary vision would recognize idiosyncrasy only in expressions whose observed preferences are completely opposite their compositional preferences, and would therefore not acknowledge the frequency-dependence of idiosyncrasies in binomial ordering preferences. Likewise, I propose that this notion of gradience extends to other phenomena we have discussed in passing, such as subcategorization preferences for verbs that participate in the dative alternation, or selection of prepositions in collocations such as *dealing with* and *suitable for*.

In the remainder of this Conclusion chapter I take up two broad questions relating to but extending beyond the current work: why languages have idiosyncrasies, and how generalization from specific items to abstract knowledge occurs. Although these broad questions go beyond the scope of the current work, I speculate about how the work presented in this dissertation suggests new avenues for approaching these questions in the future.

7.1 Why do languages have idiosyncrasies?

I now return to a question raised in the Introduction: why should there be idiosyncrasies in language at all? I have argued, in chicken-and-egg fashion, that the balance of generative and item-specific knowledge in language processing, and the balance of compositionality and idiosyncrasy in language structure, are mutually promoting. However, another stable state of affairs would be for language structure and processing to both be entirely compositional. If language contained no idiosyncrasies, all language processing could be purely generative, and in turn no idiosyncrasies would be introduced. In fact, this state of affairs is predicted by iterated learning models without a regularization bias, in which the basic result is convergence to the prior (i.e. to purely generative knowledge). The argument I have thus far provided against a purely compositional language is the existence of a regularization bias. Here I address two further questions: First, why would there be a regularization bias in language production? And second, what sources beyond this regularization bias might additionally lead to or promote idiosyncrasy in language?

7.1.1 Whence the regularization bias?

ER-regularization is rational for statistical tasks in which the goal is to maximize one's chances of correctly guessing an outcome. For example, if a coin comes up heads 70% of the time, someone who guesses heads on every flip will be right 70% of the time, whereas someone who guesses heads on 70% of flips and tails on 30% will be right only 58% of the time ($0.7 * 0.7 + 0.3 * 0.3 = 0.58$). How does this logic translate to the task of language production? If a speaker believes there is a single best order for a given binomial—for example, an order that will maximize the chance of their meaning being understood—then applying a regularization bias to their productions increases their probability of producing the best order. (For more on this topic, see the literature on probability matching versus maximizing in decision making, e.g. Gaissmaier and Schooler, 2008; Koehler and James, 2009; Shanks et al., 2002; Vulkan, 2000.)

However, the result presented in Chapter 5 relies not only on the fact that speakers apply a regularization bias to their productions, but moreover on the fact that learners do not take this regularization bias into account in learning (i.e. they do not reverse engineer the regularization process to derive what the speaker's hypothesized binomial expression preference was before regularizing). Although listeners are able to reverse engineer aspects of language production in some circumstances (Goodman and Stuhlmüller, 2013; Frank and Goodman, 2014), various factors might prevent this reverse engineering in the case of binomial expression preferences. In particular, because the amount of regularization applied by each individual speaker/generation is very small, this regularization may be undetectable and therefore go unaccounted for by learners. Alternately, if learners are attempting to take into account speakers' regularization but have some uncertainty about how much regularization is occuring, they may be erring on the side of assuming too little regularization, producing qualitatively similar results to if they assumed no regularization at all.

7.1.2 What other sources might create idiosyncrasy?

In addition to the regularization bias discussed above, many other (non-exclusive) explanations are possible for the existence of idiosyncrasies in language, spanning many different aspects and levels of description of language. Mechanistically, we know that some idiosyncrasies arise from historical language change, such as forms that follow what was once a productive rule but is no longer. From a functional perspective, such idiosyncrasies may improve efficiency in language: a stem-changing irregular verb is the same length in the past tense as in its base form, whereas a regular *-ed* past tense is longer than its base form. Given that irregular verbs tend to have high frequency, the preservation of shorter past tense forms for this high frequency verbs may promote efficient communication by assigning shorter forms to higher predictability meanings and longer forms to lower predictability meanings, as predicted by Zipf (1949). Idiosyncrasies may also be introduced by the process of language acquisition: Tomasello (2000), Pine and Lieven (2008) and many others have argued that young children's language processing, even more so than adults', is driven by item-specific knowledge, with productive generative rules emerging later in acquisition. If knowledge of some early acquired items becomes fixed before generative knowledge language is even fully available, and if these items continue to be processed primarily via item-specific knowledge even into adulthood rather than being reanalyzed via generative knowledge later on, then the process of acquisition may introduce idiosyncrasies via these early acquired items. These are just a few possible explanations for idiosyncrasy in language; there are undoubtedly many others. This dissertation makes the prediction that regardless of the source of idiosyncrasy, it will be gradiently apparent in language structure as a function of item frequency: lower frequency items are processed primarily via generative knowledge and thus cannot retain idiosyncrasies across generations, but the more experience one has with an item, the more reliably its idiosyncratic usage can be stored and reproduced.

7.2 How does generalization occur?

Finally, we turn to one of the biggest questions in language research: how does generative linguistic knowledge arise from generalization over knowledge of specific items? Taking for granted that generative knowledge does in fact consist of generalizations from known items, are those generalizations made in advance and stored in abstract symbolic form, or does use of generative knowledge actually consist of making on-the-fly comparisons to large numbers of known related items?

For example, suppose one wants to produce the binomial expression *bread and*

Marmite/Marmite and bread, and let us assume that this is a novel item for the speaker. An abstract symbolic approach to generalization would predict that the speaker has a known repertoire of abstract ordering constraints—much like we have listed in Chapters 2-4—each of which can be applied to the novel binomial and then combined to produce the speaker's generative preference (presumably *bread and Marmite*). An online generalization approach would not posit the existence of such a repertoire of ordering constraints, but rather would claim that a potential utterance such as *bread and Marmite* (or *Marmite and bread*) is generated via on-the-fly comparison to all other known expressions in the language—weighted somehow by similarity—and that *bread and Marmite* will end up being preferred over *Marmite and bread* because it is more similar to known expressions (on dimensions such as those captured by our abstract ordering constraints, but without those abstract constraints having been explicitly represented). A further difference between these two approaches is that the abstract symbolic approach predicts that constraint weights are fixed across items: the "Condiment rule" (i.e. main dishes should come before condiments, a subconstraint of the Cultural Centrality constraint) will apply equally to *bread and Marmite* or *toast and Marmite.* In contrast, the online generalization approach predicts that novel binomial ordering preferences will be more strongly determined by the preferences of other expressions that they most closely resemble, thus *bread and Marmite* might be more strongly preferred over *Marmite and bread* than *toast and Marmite* is over *Marmite and toast* due to its stronger resemblance to high frequency frozen expressions such as *bread and butter* and *bread and jam.*

In the spirit of this dissertation, I propose a gradient position on this question as well. Highly frequent, contextually diverse patterns (e.g. basic syntactic combinatorial operations along the lines of S→NP VP) are likely to be abstracted in advance away from specific items (such that a simple SVO sentence need not be processed via comparison to every other SVO sentence in one's previous experience). But if these abstractions themselves are built on the basis of their frequent use, then rarer constructions, which speakers have less experience generalizing about, are more likely to be processed via online analogy to known items. For example, a question such as "got cookies?" is unlikely to be processed via a pre-existing VO-question abstraction, since such questions are rare, but instead might be processed via direct analogy to the famous "got milk?" quote that it alludes to.[1] In between, we might find examples like *bread and Marmite* which, on the one hand, consists of a relatively high frequency pattern "X and Y" for which some abstract ordering constraints might already be known, but on the other hand shares a high degree of similarity with known high frequency expressions including *bread and butter* to which it can be directly analogized.

Binomial expressions may thus prove to be a valuable test case for this question as well, allowing us to compare the predictive power of different models of generative knowledge of binomial ordering preferences. In particular, we could compare abstract symbolic models like the ones we have used in this dissertation—which rely upon a known battery of abstract constraints with fixed weights—against models that directly analogize a given expression to its close neighbors (such as *bread and Marmite* to *bread and butter*). In this way, the work presented here may serve as a scaffold for future work on language representation and processing.

[1] Thank you to Ryan Lepic for providing this example.

Appendix A

Experimental materials

A.1 Materials used in Chapter 2

Comprehension questions are used only in Experiment 2.

A.1.1 Novel expressions

1. He was **abashed** and **sorry** about his horrible behavior.
   - Did he defend his behavior?
2. This bar is popular among the **actresses** and **lumberjacks** who live in the neighborhood.
   - Do the lumberjacks hate the bar?
3. Because Jim was **allergic** and **unaccustomed** to elderberries, he was careful to avoid them.
   - Did Jim like to eat elderberries?
4. My cousin's new talking and singing toy is **annoying** and **teal** according to my aunt.
   - Does my cousin have a new toy?
5. The dentist told Sally that **bacteria** and **candy** would rot her teeth.
   - Did the dentist recommend eating candy?
6. The elephants at the zoo were **beautiful** and **stinky** so the children loved them.
   - Were there elephants at the zoo?
7. The engineer specialized in making **bicycles** and **robots** when he worked for the company.
   - Did the engineer specialize in destroying things?

8. There were many **bishops** and **seamstresses** in the small town where I grew up.
    - Did I grow up in a small town?
9. The berries were **bitter** and **purple** when I ate them this morning.
    - Did I eat berries this morning?
10. Seth told me that there are **blankets** and **kittens** in that box over there.
    - Were there blankets in the box?
11. The rangers seemed to act like **campfires** and **wildfires** were the same thing.
    - Did I hear about fires from a policeman?
12. At the wizard school, **chanting** and **enchanting** were very common occurrences.
    - Did the wizards ride broomsticks frequently?
13. When I met many **chauffeurs** and **stewardesses** at a party, I started questioning my job.
    - Did I go to a party?
14. The third grade class saw **cherries** and **llamas** at the state fair.
    - Did the class go to the state fair?
15. There was nothing but **chickens** and **fences** in the field behind the house.
    - Was the field behind the house?
16. His uncles were all **coroners** and **senators** in their day jobs, but they all wanted to get into the movie industry.
    - Did he have uncles?
17. The drink flavored with **currant** and **pomegranate** was delicious according to Kim.
    - Did Kim like the drink?
18. The dictator was **deposed** and **murdered** by his military adviser.
    - Did the dictator survive?
19. I talked with my boss about whether to hire the **determined** and **forgettable** job candidate that we interviewed.
    - Did I discuss something with my boss?
20. The doctor said that **discontent** and **tearfulness** are signs of depression.
    - Did the doctor talk about flu symptoms?
21. Luke always looked so **disheveled** and **dreary** but he was my best friend.
    - Was Luke my best friend?
22. The kind minister **donates** and **provides** a lot of food to the charity.
    - Was the minister kind?
23. My favorite animals have been **felines** and **quails** ever since I was a kid.
    - Have I always hated animals?
24. The finalists in the tennis championship were ranked **first** and **ninety-eighth** in the world prior to the tournament.
    - Was there a golf champtionship?
25. In the spring, Julie will plant **flowers** and **zinnias** in her new garden.
    - Does Julie have a garden?
26. The store owner was **fuming** and **mad** when he found out what was stolen.

- Was something stolen?

27. As a vegetarian, **gelatin** and **lard** are difficult to avoid.
    - Do vegetarians have a hard time?
28. Laura heard that the school's **groundskeeper** and **superintendent** got married over the summer.
    - Did Laura hear about a divorce?
29. His mother didn't hear when when Nate **happily** and **rudely** told his sister to shut up.
    - Did his mother hear what Nate said?
30. As Joe carried a tall stack of boxes, he had to **hesitate** and **readjust** before he could go further.
    - Was the worker carrying barrels?
31. At the zoo we saw **horses** and **loons** in their natural habitats.
    - Did we go to the zoo?
32. I need to grab my **jacket** and **phone** before I leave the house.
    - Do I have everything I need in order to leave?
33. Sarah likes to buy **kale** and **vegetables** at the famer's market.
    - Does Sarah only buy meat?
34. My cousins were all **lankier** and **lanky** but were surprisingly strong.
    - Were my cousins weak?
35. The pet store was full of **litter** and **newts** when Martha visited on Saturday.
    - Did Martha go to the pet store?
36. Peter met a man who was **masculine** and **undignified** at the conference he went to last month.
    - Did Peter go to the conference last year?
37. The pirate was **marooned** and **missing** for nearly five months.
    - Was the pirate stranded for a year?
38. My grandparents were all **nurses** and **patriarchs** when they were alive.
    - Were some of my grandparents teachers?
39. In my dream, I had **puppies** and **tigers** that I kept as pets.
    - Was I dreaming?
40. Jenny was interested in **rats** and **sharks** as a young child.
    - Was Jenny interested in kittens?
41. Maria could use **therapy** and **vacations** to feel less stressed.
    - Is Maria stressed?
42. Irena had trouble with **vocabulary** and **vowels** while she was learning English.
    - Did Irena have trouble with vowels?

A.1.2 Attested expressions

1. The clerk asked for Melissa's **address** and **name** in order to complete the form.
    - Did the clerk help Melissa complete the form?
2. Sarah was relieved to find that her friends were **alive** and **well** after the car

crash.
- Were Sarah's friends alright?

3. Most universities have programs in the **arts** and **sciences** in addition to having various professional schools.
   - Do most university have programs about law?
4. Soccer players practice running both **backwards** and **forwards** in order to stay nimble.
   - Do soccer players practice running sideways?
5. Hunter dislikes reading **black** and **white** text off a computer screen so he uses an unusual color scheme.
   - Does Hunter like the standard color scheme?
6. Learning to strengthen your **body** and **mind** is one of main purposes of doing yoga.
   - Does yoga improve your strength?
7. George always brings **bread** and **butter** with him when he goes camping.
   - Does George always bring hot chocolate when he goes camping?
8. John showed me pictures of the **bride** and **groom** both dressed in blue.
   - Did the couple wear green?
9. I always love seeing my **brothers** and **sisters** when I go home for the holidays.
   - Do I enjoy going home?
10. Caleb likes to **buy** and **sell** electronics on eBay as a hobby.
    - Does Caleb work with eBay professionally?
11. I watched the **cat** and **mouse** run frantically around the barn.
    - Was there a dog in the barn?
12. It can be difficult to determine the **cause** and **effect** of weather patterns over the ocean.
    - Are ocean weather patterns hard to predict?
13. Clarissa found the painting of a **child** and **mother** to be very moving.
    - Did Clarissa see a painting?
14. Catherine was not surprised that tensions between **church** and **state** ran high during the election season.
    - Was there tension during the election season?
15. Peter studied the laws concerning **crime** and **punishment** in Ancient Greece and Rome.
    - Did Peter study what happened in Ancient Greece?
16. Jesse felt like he had worked **day** and **night** on the project but he only got a B on it.
    - Did Jesse get an A?
17. The economist became famous for studying the way **demand** and **supply** affect the steel industry.
    - Did the economist study oil companies?
18. Mark finds working on **development** and **research** for the marketing company to be a very satisfying career.

- Does Mark want to change jobs?

19. Although some **drink** and **food** were provided at the reception, there was not enough to go around.
    - Was there something to eat at the reception?
20. Diane wrote a book about her travels **east** and **west** around the globe for a year.
    - Did Diane write a book about living in Paris?
21. Sometimes it feels like **error** and **trial** is the only way to learn.
    - Do you sometimes need to learn by trying things?
22. Heather invited her **family** and **friends** to her annual holiday party.
    - Does Heather have a holiday party every year?
23. It is important to study both the **fauna** and **flora** in a region in order to fully understand the ecosystem.
    - Can studying plant life tell you everything you need to know about an ecosystem?
24. Many children find eating with a **fork** and **knife** to be a difficult skill to learn.
    - Do some children have trouble with eating utensils?
25. Keith marveled at the **gold** and **silver** decorations on the walls of the palace.
    - Were the walls dull?
26. Excercising regularly is important for your **heart** and **soul** according to my mother.
    - Did I receive advice from my aunt?
27. Michelle was surprised to learn that the **husband** and **wife** were getting a divorce.
    - Was the couple celebrating their anniversary?
28. I could not guess the **intents** and **purposes** of the confusing new regulations.
    - Were the regulations confusing?
29. Everyone bowed as the **king** and **queen** entered the throne room.
    - Did a jester enter the room?
30. Learning to forecast **loss** and **profit** was a topic in Brian's business skills class.
    - Did Brian take a class on business skills?
31. Paul primarily got his news through **magazines** and **newspapers** rather than through television.
    - Does Paul read the news?
32. I like to **match** and **mix** my clothing to create new outfits.
    - Do I like to always wear the same thing?
33. Jen thought that the **men** and **women** in her dance class were all very talented.
    - Did Jen think that some of her classmates were untalented?
34. Blake dislikes seeing all the **pain** and **suffering** in the world when he watches the news.
    - Does Blake enjoy watching the news?
35. The anthropologist studied the way different cultures conceived of **peace** and **war** during the Middle Ages.

- Did the anthropologist study dinosaurs?

36. By comparing the **past** and **present** we can learn about universal human tendencies.
    - Does history help us understand humanity?
37. Seth follows both **radio** and **television** broadcasts to stay informed about current events.
    - Does Seth like to follow current events?
38. Some children enjoy learning to **read** and **write** but others dislike it.
    - Do some children enjoy reading more than others?
39. Teaching children what is **right** and **wrong** is a difficult task for parents.
    - Is it easy to teach children morals?
40. After the storm, Haley was glad to hear that her grandparents were **safe** and **sound** in their country home.
    - Was there a storm?
41. The broker bought some risky **shares** and **stocks** without knowing it and only discovered it later.
    - Was the broker originally unaware of what he did?
42. Susan disliked the **sour** and **sweet** soup at the fancy restaurant.
    - Was the restaurant fancy?

A.1.3 Constraint activity profiles

Figure A.1 shows the proportion of items for which each constraint is active (recalling that each constraint can be active or inactive for a given expression). As we can see, constraints are active approximately equally often in each group. Tables A.1 and A.2 show correlations between constraints: constraint activity is coded as 1 if it predicts that an expression should occur in alphabetical order and -1 if it predicts that an expression should occur in non-alphabetical order, or 0 for inactive constraints. We see that, for both novel and attested expressions, most constraints are not highly correlated. One noteworthy exception is Length and Final Stress, which are highly correlated because single-syllable words are as short as possible (hence should come first according to Length) and necessarily have final stress (hence should come first according to Final Stress).

Table A.1: Correlations of constraint activity for attested binomials.

	Form	Power	Icon	Percept	Length	Freq	Stress
Form	1.00	0.22	0.01	0.01	0.01	−0.34	0.02
Power	0.22	1.00	0.01	0.06	−0.05	0.05	−0.06
Icon	0.01	0.01	1.00	−0.11	0.44	0.26	0.30
Percept	0.01	0.06	−0.11	1.00	−0.14	−0.10	−0.17
Length	0.01	−0.05	0.44	−0.14	1.00	0.30	0.83
Freq	−0.34	0.05	0.26	−0.10	0.30	1.00	0.21
Stress	0.02	−0.06	0.30	−0.17	0.83	0.21	1.00

Table A.2: Correlations of constraint activity for novel binomials.

	Form	Power	Icon	Percept	Length	Freq	Stress
Form	1.00	0.09	0.02	−0.01	0.15	0.44	0.18
Power	0.09	1.00	0.05	−0.02	0.10	0.13	0.11
Icon	0.02	0.05	1.00	0.03	0.13	0.03	0.04
Percept	−0.01	−0.02	0.03	1.00	0.24	0.22	0.11
Length	0.15	0.10	0.13	0.24	1.00	0.10	0.50
Freq	0.44	0.13	0.03	0.22	0.10	1.00	−0.28
Stress	0.18	0.11	0.04	0.11	0.50	−0.28	1.00

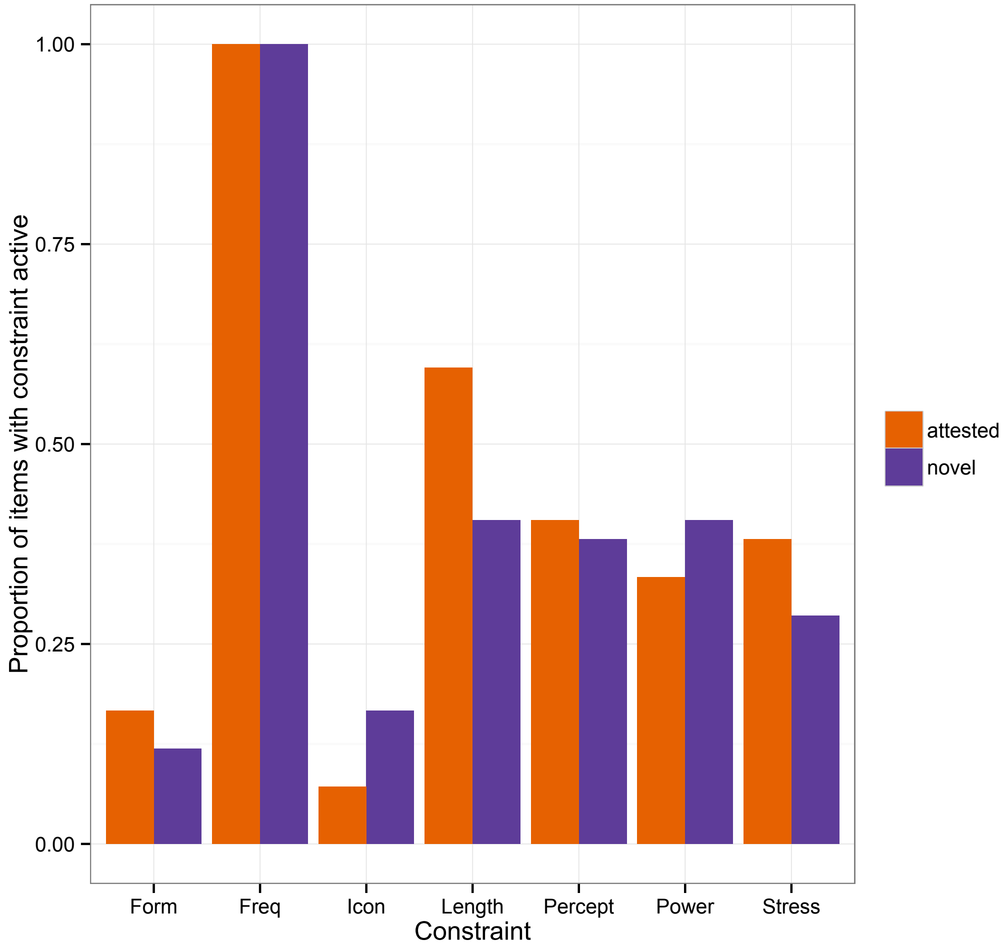


Figure A.1: Proportion of binomial expressions for which each constraint is active.

A.2 Materials used in Chapter 6

Experimental materials in order of increasing overall frequency.

- The workers loaded the **barriers** and **stakes** into the cargo van and drove away.
- Mark measured the **boards** and **two-by-fours** before starting to build the new shed.
- I believe that **Catholics** and **Non-Catholics** can agree on most moral questions.
- Laura watched the **farms** and **hayfields** pass by on the long car ride.
- Emily studied **linguistics** and **psychiatry** before becoming an author.
- Dealing with frequent **nagging** and **stress** is bad for your health.
- The lecturer said that **diffidence** and **gentleness** are important qualities in a nurse.

- The country was ruled by a **madman** and **tyrant** after the failed revolution.
- The salesman said that **appliances** and **cosmetics** were on the third floor.
- It is difficult to deal with the **ambiguities** and **fears** that arise when a loved one is ill.
- The documentary explored the culture of **jazz** and **marijuana** in the early 1930s.
- Judith's favorite colors have been **fuchsia** and **rose** ever since she was a child.
- Ross was praised for his **authority** and **decency** during his performance review.
- The old woman loved her **dogs** and **grandchildren** despite all the noise they made.
- Marissa avoided the **derelicts** and **outcasts** as she walked down the alley.
- The violinist played with **abandon** and **fervor** during the orchestra concert.
- Dennis bought **salt** and **whiskey** in preparation for the party.
- My boss insists on **directness** and **truth** in all of our communication.
- Tony had many **scars** and **welts** from his days as a boxer.
- Mary learned **graces** and **manners** while growing up with her grandmother.
- Todd was shocked by the look of **blame** and **hate** in his girlfriend's eyes.
- I memorized **facts** and **techniques** for my final exam in medical school.
- The economist studied how the **domination** and **influence** of one country affected a whole region.
- The committee discussed the job candidateÕs **activities** and **character** during the hiring meeting.
- Gabe missed the **comfort** and **companionship** of owning a dog.
- The prevalence of online harassment is a **danger** and **threat** to children growing up today.
- Liz likes to make **pies** and **puddings** for parties and special occasions.
- Dan appreciated the **certainty** and **security** of having a five-year job contract.
- Many students experience **boredom** and **loneliness** during their first year of college.
- Courtney made a new **skirt** and **sweater** from the leftover fabric.
- Most of the **counties** and **towns** affected by the earthquake have since been rebuilt.
- The kids saw the **jaws** and **teeth** of a dinosaur at the museum.
- Peter studied the **charts** and **maps** of the area around the Boy Scout camp.
- Ginny sends her **friends** and **relations** boxes of colorful holiday cookies every year.
- It is important to cover your **body** and **face** with sunscreen during the summer.
- David avoids discussing **politics** and **religion** on a first date.
- Richard thought of his **child** and **wife** often while he was away from home.
- Megan wondered whether she had enough **minerals** and **vitamins** in her usual diet.
- Sometimes your **arms** and **hands** become numb if you sit still for a long time.
- Scott prefers **magazines** and **newspapers** over watching shows on television.

- Blake read about **pressure** and **temperature** in his atmospheric science textbook.
- The company is switching away from **gas** and **oil** to use more renewable energy sources.
- Barbara couldn't find a good **place** and **time** to ask for a raise.
- Angela works on **development** and **research** for a large biomedical company.
- The organization provides **goods** and **services** to people in need.
- Jamie showed the **boys** and **girls** how to make a paper airplane.
- The artist painted a **black** and **white** pattern on the canvas.
- The guard instructed the **men** and **women** to form an orderly line.

Appendix B

Additional analyses from Chapter 2

B.1 Experiment 2 region-by-region analyses

Here we present region-by-region analyses of the self-paced reading data from Experiment 2. Our goals in these analyses are to replicate the results of Siyanova-Chanturia et al. (2011) that attested binomial expressions are read faster in their preferred order, and to demonstrate that this finding extends to novel expressions when categorized into preferred/dispreferred orders on the basis of abstract knowledge. Specifically, we analyze reading times by dichotomizing binomials into preferred/dispreferred conditions, rather than using continous abstract knowledge and relative frequency predictors as in Section 2.5.2. For simplicity of presentation, and because we are not concerned here with comparisons across binomial types, we analyze each type (attested/novel) separately.

Residualization on word length and outlier removal are identical to that reported in Section 2.5.2, except that outlier removal was done for each region and each binomial type separately (because each region within each type is analyzed separately in this section).

For each binomial type and region, we fit a linear mixed-effects regression model with residualized reading times (in milliseconds) as the dependent variable. Our independent predictor of interest is a dichotomous preferred/non-preferred variable (treatment coded with "preferred" as the reference level). Details of how preferred order is assessed vary between binomial types and are discussed in more details below. Trial order is also included as a predictor. Following Barr et al. (2013), we use the maximal random effects structure for subjects justified by the experimental design, namely an intercept and a slope for preferred/non-preferred order. We also include a random by-subjects slope for trial order. For items, defined as unordered word pairs, we use an intercept and a slope for a binary alphabetical/non-alphabetical factor (comparable to that used in Section 2.5.2). Results for the predictor of interest are shown in Table B.1.

Novel expressions For novel expressions, we assign each expression a preferred and non-preferred order on the basis of our abstract knowledge model's prediction for ordering preferences. Results are shown in Figure B.1. As seen in Table B.1, we find significant effects of order at the Word1 and Word2 regions, with preferred read faster than non-preferred.

Attested expressions For attested expressions, we consider two ways to sort expressions into preferred and non-preferred order: we can use corpus frequencies, replicating Siyanova-Chanturia et al. (2011), or we can use abstract knowledge model predictions for a more direct comparison with the novel expressions. We will show results sorted both ways.

We begin by showing results with preferred/non-preferred determined by corpus frequencies as reported by Siyanova-Chanturia et al.[1] Results are shown in Figure B.2. We find significant effects of order at the And, Word2, and Spill1 regions, with preferred read faster than non-preferred.[2]

Next we analyze our attested expressions as sorted by abstract knowledge

[1]Siyanova-Chanturia et al.'s reported preferences differ from the Google n-gram preferences for one item, *family and friends*.

[2]Siyanova-Chanturia et al. only report aggregate reading times, not word by word reading times, so we cannot say whether our results directly replicate exactly where in the sentence these effects appear.

Table B.1: Means, standard errors, and t values for the estimated coefficient of the preferred/dispreferred predictor in the region-by-region analyses of the self-paced reading experiment. t values greater than 2 are taken to be significant.

		Prelim	Word1	And	Word2	Spill1	Spill2	Spill3
Novel	Mean (SE)	3.02 (2.14)	11.61 (5.16)	8.21 (6.05)	11.18 (4.79)	-1.74 (4.49)	1.03 (2.41)	2.01 (2.71)
	t value	1.41	2.25	1.36	2.34	-0.39	0.43	0.74
Attested (corpus freq)	Mean (SE)	-2.05 (1.87)	-3.55 (3.34)	7.44 (2.59)	15.26 (3.37)	8.82 (2.40)	2.94 (2.75)	2.90 (1.91)
	t value	-1.10	-1.06	2.87	4.53	3.67	1.07	1.52
Attested (model)	Mean (SE)	-0.12 (1.91)	-3.39 (3.32)	-1.79 (2.78)	5.87 (3.93)	6.12 (2.64)	-2.29 (2.75)	1.84 (2.01)
	t value	-0.06	-1.02	-0.64	1.49	2.32	-0.83	0.92

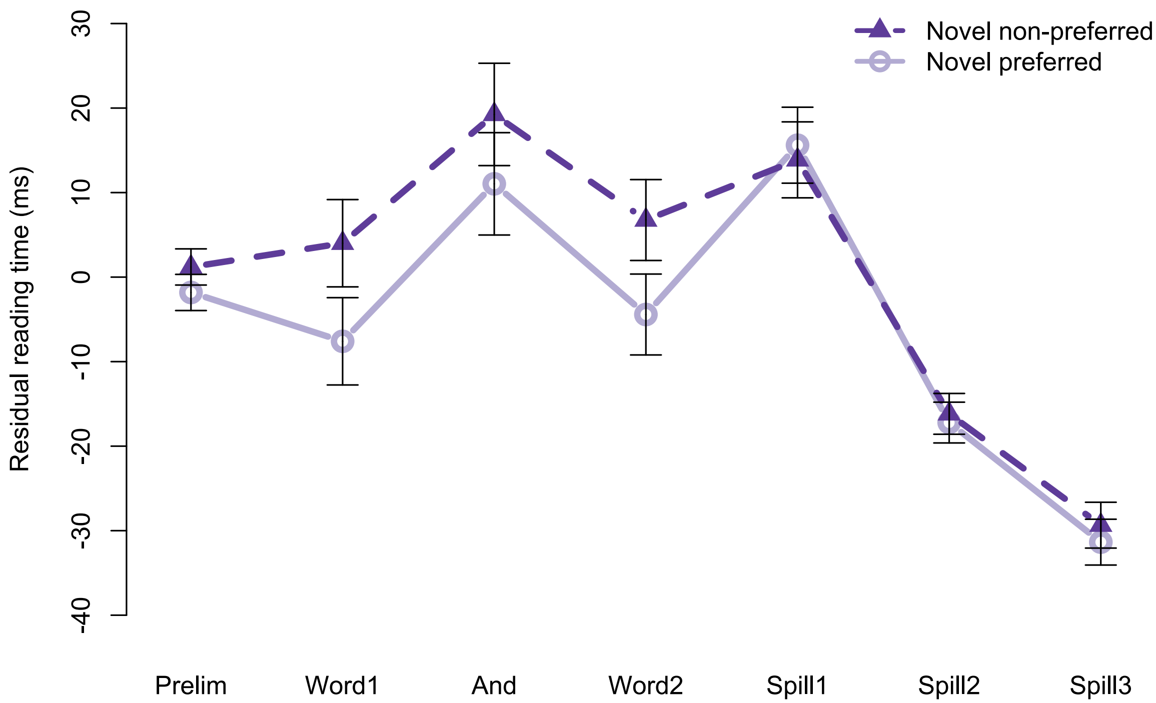


Figure B.1: Self-paced reading times for novel expressions. Error bars show standard errors for the predictor of interest (Table B.1).

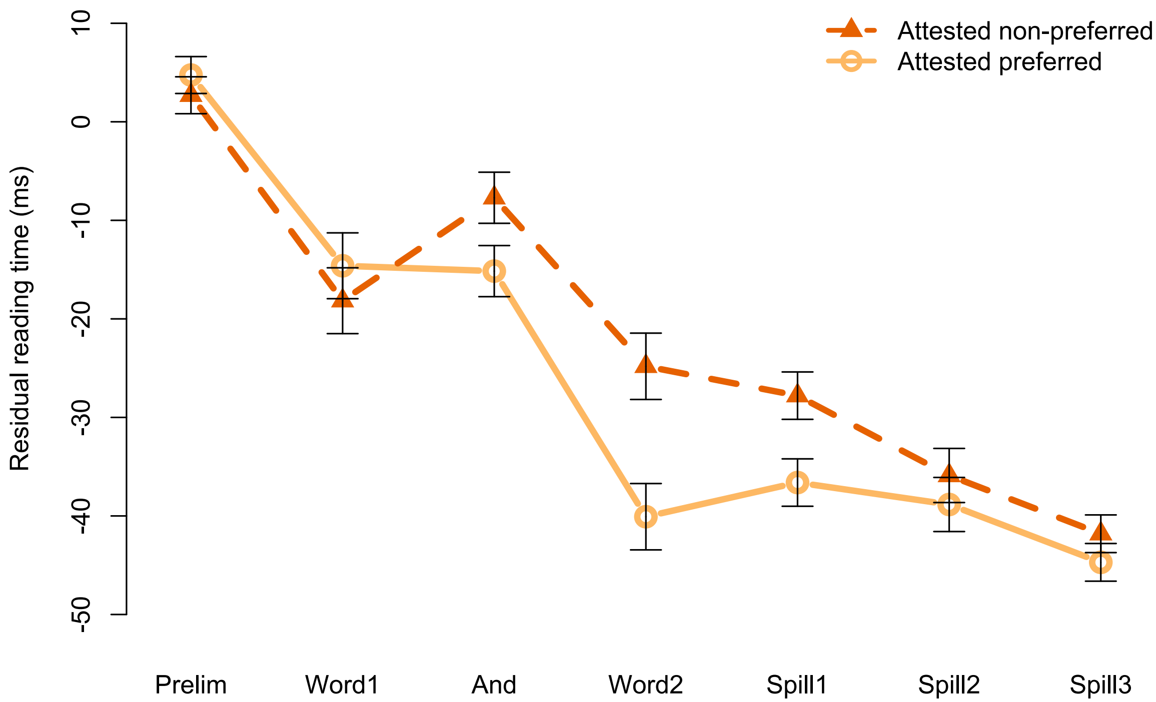


Figure B.2: Self-paced reading times for attested expressions with preferred direction determined by corpus frequency. Error bars show standard errors for the predictor of interest (Table B.1).

model predictions. Results are shown in Figure B.3. We find a significant effect of order at the Spill1 region, with preferred read faster than non-preferred.

Discussion We replicate Siyanova-Chanturia et al.'s (2011) finding that attested binomial expressions are read faster in their preferred order. We also demonstrate for the first time that novel binomials show online effects of abstract constraints on ordering, with faster reading times in our model's predicted preferred direction.

We do not present a region-by-region version of the multivariate analyses presented in Section 2.5.2 because we do not expect the results seen there to hold at each region individually. As noted in Section 2.5.2, the analyses presented there took advantage of the fact that within the six-word region analyzed, participants read the same set of words regardless of order of binomial presentation. Within the word-by-word analyses presented here, however, words differ across conditions: Word1 in the preferred condition becomes Word2 in the dispreferred condition, and vice versa (e.g. "bishops and seamstresses" versus "seamstresses and bishops"). Moreover, recall that effects of lexical frequency are one component of abstract knowledge (Section 2.2), such that binomials in preferred order on average have a more frequent word preceeding a less frequent word, while binomials in dispreferred order on average have a less frequent word proceeding a more frequent word. Thus, on the basis of lexical frequency alone, we would expect to see the preferred order read faster around Word1 (or shortly thereafter, due to spillover), and the dispreferred order read faster around Word2 (or shortly thereafter). In other words, on the basis of lexical frequency alone, we would expect to see a local reversal of the effect of abstract knowledge around Word2 (although we expect this reversal to be smaller in magnitude than the overall benefit of conforming to abstract knowledge across the binomial as a whole). This prediction is born out numerically in the Spill1 region for novel binomials, although it does not approach significance.

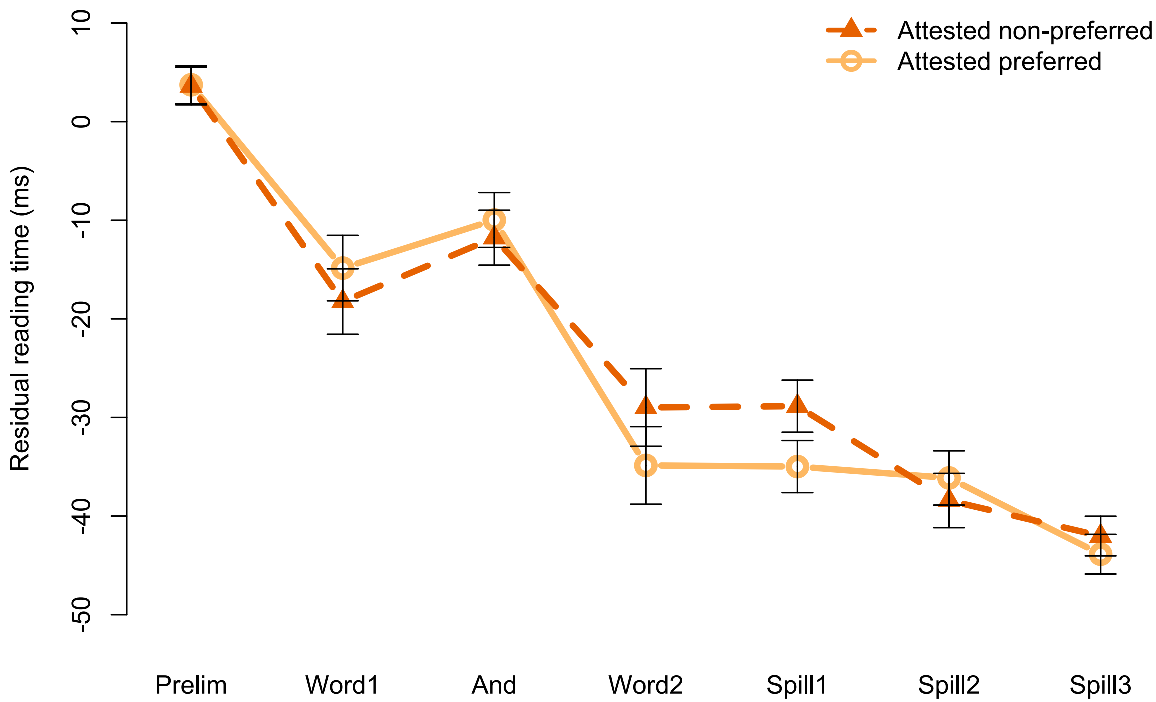


Figure B.3: Self-paced reading times for attested expressions with preferred direction determined by model predictions. Error bars show standard errors for the predictor of interest (Table B.1).

Table B.2: Model fit for results of Experiment 2 using raw reading times. Effects with $t > 2$ are taken to be significant. All VIF < 1.6.

	Estimate	*Std. Error*	*t value*
Intercept	2246.62	41.66	53.93
Type: novel	204.18	28.51	7.16
Abs know (Type: attested)	0.72	24.97	0.03
Abs know (Type: novel)	-56.81	19.88	-2.86
Rel freq	-57.46	19.98	-2.88
Trial order	-198.11	9.46	-20.95

B.2 Experiment 2 results with raw reading times

Here we replicate the analyses presented in Section 2.5.2 with raw rather than word-length-residualized reading times. Model results are given in Table B.2. Crucial effects are very similar to those seen in Table 2.4. In a likelihood ratio test comparing this model to a model with only an additive (non-nested) effect of abstract knowledge, we find a marginally significant difference ($\chi^2(1) = 3.12, p = 0.08$). We attribute the lower significant level here compared to that presented in in Section 2.5.2 to presence of extra noise in the raw compared to the residualized reading time data.

Appendix C

Generative constraint coding guide

This coding guide is largely adapted from Benor & Levy (2006).

Pragmatic An element that has been previously mentioned will appear first.
Examples:

- *English and Americans* in the context: Oh yes, the other day I reread some of EmersonÕs *English Traits*, and there was an anecdote about a group of English and Americans visiting Germany, more than a hundred years ago.

Word order in a neighboring phrase will be preserved
Examples:

- *music and comedy* in the context: I admit that going back to Ralph Waldo Emerson for humor is like going to a modern musical comedy for music and comedy.

When one element is more closely related to a neighboring modifier, it is preferred in the slot closer to the modifier
Examples:

- *buttons and badge* in the context: the buttons and badge of a policeman

Relative Formal Markedness Less marked items should come first. Less marked items:

1. have a broader, more general meaning
2. have greater freedom of distribution
3. have a larger number of subcategorical distinctions

Examples:

- *flowers and roses*: a rose is a type of flower
- *changing and improving*: you can change without improving, but not vice versa
- *first and only*: something can be first without being only, but not vice versa

Less marked items are structurally more simple
Examples:

- *linguistic and paralinguistic*
- *poetry and non poetry*

An item is more marked if it is defined by or discussed in relation to another
Examples:

- *there and elsewhere*

An item is more marked if it a logical subset of another:
Examples:

- *sewing and alterations*: alterations are a subset of sewing projects (if you assume that all alterations involve sewing)

This can also include physical relationships:
Examples:

- *kitchen and pantry*

Perceptual Markedness Less marked elements appear first.

LESS MARKED	MORE MARKED	
Animate	Inanimate	
Singular	Plural	(For count nouns only)
Right	Left	(This one is context dependent.)
Positive	Negative	
Concrete	Abstract	
Front	Back	
Above	Below	
Vertical	Horizontal	
Here	There	

(Other markedness constrasts will turn up too, e.g. land before sea, objective before subjective, solid before liquid/gas.) Although multiple constrasts could in principle apply in conflicting directions for a single expression, this never occurred in our corpus.
Examples:

- *people and soils*
- *individually and cumulatively*
- *individuals and couples*
- *physical and mental*
- *up and down*
- *high and inside* (vertical and horizontal)
- *Harvard and Yale* (if you're at Harvard) vs. *Yale and Harvard* (if you're at Yale)
- *honest and stupid* (positive and negative)

N.B. Percept does not apply in cases of a singular mass noun and a plural count noun, e.g. *integrity and principles*

Cultural Centrality More culturally central or common elements appear first
Examples:

- *north and south*
- *mother and dad*: mother is more central to raising a child
- *day and night*
- *see and hear*: seeing is the more salient form of perception

- *oranges and grapefruit*
- *family and friends*
- *food and drink*: aka. the Condiment Rule—now Centrality rather than Power!

Power More powerful or culturally prioritized elements appear first
Examples:

- *husband and wife*: men before women
- *mother and child*: older first
- *gold and silver*
- *clergymen and parishioners*
- counter-example: *peanuts and emeralds*

Itensity Elements with more intensity appear first
Examples:

- *cruel and unusual*
- *war and peace*

Iconic/Scalar Sequencing When two elements are perceived as existing in a sequence or on a scale they should appear in that same sequence. Sequences may be e.g. chronological or cause-and-effect
Examples:

- *wait and see*
- *kiss and tell*
- *slowed and stopped*
- *there and back*: you must go there before coming back
- *out and about*: you must go out in order to go about
- *unconstitutional and severable*: in a context where a contract was severable because it was unconstitutional
- *eighth and nineth*
- *months and years*

Notes on some subtle distinctions Preconditions count as RelForm if they're in a logical subset relation (e.g. *sewing and alterations*) or Icon if they're temporal/cause-and-effect relations (e.g. *unconstitutional and severable, achieved and maintained*).

Contrasts of intensity count as RelForm if they're in a logical subset relation (e.g. *mad and fuming*, where *fuming* is a subset of *mad*), or Intensity otherwise.

Appendix D

Corpus

Table D.1: The corpus described in Chapter 3.

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
20s	30s	0	0	1	0	0	1	-0.702	0	0	-0	5667	1.000
a.b.	m.a.	0	0	0	-1	0	1	-0.852	0	0	0	1374	1.000
abasement	humiliation	-1	0	0	0	1	1	-2.544	2	0	-0	1085	0.470
ability	age	-0	-1	-0	-0	-0	-1	-0.886	-2	-2	-1	12237	0.095
ability	desire	-0	1	-1	-0	-0	-0	0.141	-1	-1	0	26503	0.464
ability	strength	-0	-1	-0	-0	-0	-0	0.002	-2	-2	-1	20058	0.207
abolition	emancipation	1	0	0	0	0	0	-0.075	0	0	-0	1599	0.871
action	character	-0	-1	-0	-0	-0	-0	0.483	1	1	0	18766	0.325
action	conversation	-0	-0	-0	-0	-0	-0	1.570	2	-0	0	2470	0.367
action	mind	-0	1	-0	-0	-0	-0	-0.088	-1	-1	-1	5184	0.067
action	motion	-1	0	0	0	0	0	1.105	0	0	-0	5237	0.397
actions	feelings	-0	1	-0	-0	-0	-0	0.113	-0	-0	0	31729	0.269
activities	character	0	1	0	0	0	0	-0.139	0	0	-0	2796	0.214
activities	places	-0	-0	-0	-0	-0	-0	0.312	-1	-1	0	3873	0.288
activity	nature	-0	1	-0	-0	-0	-0	-0.617	-1	-1	0	3742	0.132
addresses	names	-0	-0	-1	-0	-0	-0	-1.569	-2	-1	-1	273323	0.007
adventure	romance	-0	-0	-0	-0	-0	-0	-0.013	-0	-0	0	22627	0.376
africa	asia	-0	-0	-1	-1	-0	-0	0.524	-1	-1	0	293128	0.409
age	day	-0	-0	-0	-0	-0	-1	-0.877	-0	-0	-0	98957	0.004
agencies	individuals	-0	-1	-0	-0	-0	-0	-0.837	2	-0	0	20024	0.530
aggression	hostility	0	0	0	0	1	0	0.118	1	1	-0	17481	0.359
agreements	contracts	1	-0	-0	-1	-0	-0	-0.376	-0	-1	-1	22390	0.291
aims	ends	0	0	0	0	0	0	-1.251	0	0	0	7006	0.461
aircraft	ships	-0	-1	-0	-0	-0	-0	-0.422	-1	-0	-0	29331	0.187
alibis	excuses	0	1	0	0	0	0	-2.954	-1	1	1	1103	0.549
aloes	myrrh	-0	-0	-0	-0	-0	-0	-0.591	-1	-1	-1	7735	0.157
ammo	weapons	-0	-0	-1	-0	-0	-0	-3.739	-0	1	1	2212	0.100

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
ammunition	baggage	-0	-0	-1	-0	-0	-0	0.327	-2	-0	0	1494	0.431
ammunition	guns	-0	-0	-1	-0	-0	-0	-1.248	-3	-1	-1	55799	0.040
anger	anxiety	0	0	0	0	1	0	-0.035	1	1	-0	19277	0.422
anger	fear	0	0	0	0	1	0	-1.079	-1	-1	-1	88323	0.414
anger	hatred	-0	-0	-0	-0	-1	-0	0.919	-0	-0	0	29441	0.707
anger	indignation	1	-0	-0	-0	-0	-0	1.662	2	-0	0	13774	0.719
anger	spite	1	0	0	0	1	0	-0.068	-1	-1	-1	1672	0.563
anguish	pain	-1	-1	-0	-0	-0	-0	-2.702	-1	-1	-1	31259	0.195
anguish	rage	0	0	0	0	-1	0	-0.796	-1	-1	-1	3683	0.367
anthropology	psychology	-0	-0	-1	-0	-0	-0	-1.572	-2	-0	0	16325	0.543
anxiety	dismay	0	0	0	0	0	0	2.154	-2	-2	-1	1411	0.617
apprehension	fear	-1	0	0	0	-1	0	-2.651	-3	-1	-1	19096	0.249
apprehension	terror	-0	1	-0	-0	-1	-0	-1.025	-2	-0	0	1482	0.380
archaeologists	geologists	-0	1	-0	-0	-0	-0	0.102	-2	-0	0	1469	0.432
arches	piers	-0	-0	-1	-0	-0	-0	0.525	-1	-1	-1	3647	0.228
argument	debate	-1	-0	-0	-0	1	-0	0.734	-2	-2	-1	7440	0.670
arizona	california	0	0	-1	0	0	0	-1.773	0	0	-0	42843	0.462
arm	back	0	1	0	0	0	0	-2.096	0	0	0	7257	0.581
arms	faces	0	-1	0	-1	0	0	0.993	1	1	1	7402	0.444
arms	hands	-0	1	-1	-0	-0	-0	-0.601	-0	-0	-0	194013	0.453
arms	legs	0	1	0	0	0	0	0.779	0	0	0	667976	0.837
arrangers	composers	-0	-0	-0	-1	-0	-0	-3.617	-0	-0	0	3127	0.179
arsenic	strychnine	0	0	1	0	0	0	1.593	-1	-2	-1	1342	0.652
artillery	mortar	-0	-0	-0	-1	-0	-0	0.906	-1	-1	0	19688	0.631
artist	critic	0	0	1	0	0	1	1.008	0	0	-0	5674	0.809
arts	crafts	1	0	0	0	0	0	2.102	0	0	0	454080	0.995
arts	sciences	0	0	0	-1	0	0	0.354	2	2	1	618850	0.954

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
ash	maple	-0	-0	-0	-0	-0	-0	0.593	1	1	1	2666	0.432
associations	enterprises	-0	-0	-0	-1	-0	-0	0.387	-0	-0	0	2189	0.399
assurance	dignity	0	0	0	0	0	0	-0.345	1	1	-0	1291	0.424
astronomy	geology	0	-1	0	0	0	0	-0.239	0	0	-0	7933	0.583
attention	civility	0	0	0	0	0	1	4.469	1	1	-0	2048	0.392
attitude	experience	0	-1	0	0	0	0	-1.251	0	2	1	3253	0.442
attractions	repulsions	0	1	0	0	0	0	3.462	0	0	-0	12897	0.952
aunt	uncle	-0	-0	-0	-1	-0	-0	-0.342	1	1	1	165376	0.638
aunts	uncles	0	0	1	-1	0	0	-0.028	1	1	1	114866	0.632
authority	force	-0	-0	-0	-0	-0	-0	-0.528	-2	-2	-1	9915	0.330
authority	independence	0	0	0	1	0	0	0.902	1	-1	-0	6920	0.512
azaleas	camellias	-0	-0	-0	-0	-0	-0	0.725	1	-1	-1	3020	0.601
b	c	0	0	0	0	0	1	-0.063	0	0	0	617842	0.945
babylon	nineveh	0	0	1	0	0	0	1.624	0	2	1	12006	0.338
back	behind	0	1	0	0	0	0	1.310	0	0	0	1511	0.807
back	hips	0	1	0	0	0	0	4.440	0	0	0	7534	0.685
back	shoulders	0	-1	0	0	0	0	2.934	1	1	1	63372	0.630
background	intelligence	-0	-0	-1	-0	-0	-0	0.043	1	2	1	1562	0.686
background	interest	0	0	0	0	0	1	-1.502	1	2	1	4132	0.620
bacon	coffee	0	0	1	0	0	0	-1.079	0	0	-0	9230	0.515
bacon	eggs	0	0	1	0	0	0	-0.783	-1	-1	-1	86914	0.644
bacteria	insects	0	-1	0	0	0	0	0.294	-1	-2	-1	1765	0.455
bandits	outlaws	-1	0	0	0	0	0	0.179	0	-1	-1	1364	0.548
bar	grill	0	0	1	0	0	0	2.902	0	0	0	40757	0.963
barges	towboats	-0	-0	-0	-1	-0	-0	3.503	-0	-1	-1	1270	0.390
barley	oats	-0	-0	-0	-0	-0	-0	0.222	-1	-1	-1	27431	0.430
barrels	boxes	-0	-1	-1	-0	-0	-0	-1.101	-0	-0	0	12973	0.371

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
bathrobe	slippers	0	1	0	0	0	0	-1.463	0	1	1	6684	0.896
battle	bloodshed	0	0	0	0	0	1	3.434	0	-1	-1	1555	0.869
beauty	brains	-0	1	-1	-0	-0	-0	1.912	-1	-1	-1	5091	0.620
beauty	color	-0	-1	-0	-0	-0	-0	-0.654	-0	-0	0	8914	0.347
beauty	dignity	-0	-0	-0	-0	-0	-0	1.036	1	1	0	25267	0.521
beauty	inspiration	0	1	0	0	0	0	1.237	2	0	-0	2617	0.822
beauty	poise	0	0	1	0	0	0	3.454	-1	-1	-1	1571	0.610
beauty	truth	-0	-1	-0	-0	-0	-0	-0.812	-1	-1	-1	77484	0.326
beauty	youth	-0	-1	-0	-0	-0	-1	-0.052	-1	-1	-1	65503	0.101
beech	elm	0	0	0	0	0	0	-0.268	0	0	0	1837	0.608
being	existence	0	0	0	0	0	0	1.980	0	0	-0	12923	0.526
belly	chest	-0	-1	-0	-0	-0	-0	-1.298	-1	-1	-1	14929	0.229
berlin	frankfurt	0	0	1	0	0	0	1.974	1	1	1	4154	0.611
betrayal	murder	0	0	0	-1	0	1	-2.089	0	0	-0	1796	0.654
bitterness	disappointment	0	0	0	1	0	0	-0.367	1	-1	-0	5200	0.603
black	white	0	-1	0	0	0	0	-0.212	0	0	0	2268565	0.832
blame	hate	0	0	0	0	-1	0	-0.216	0	0	0	1113	0.661
blood	flesh	-0	-1	-0	-0	-0	-0	1.714	-0	-0	-0	556220	0.026
blood	guts	0	0	0	0	-1	0	4.209	0	0	0	19009	0.961
blossom	leaf	0	0	1	0	0	0	-1.703	-1	-1	-1	4489	0.185
boardinghouses	hotels	-0	-0	-0	-0	-0	-0	-4.366	-3	-1	-1	2816	0.181
boats	ships	-0	1	-1	-0	-0	-0	-0.728	-0	-0	-0	24996	0.379
bodies	faces	-0	-1	-0	-0	-0	-0	0.440	-0	-0	0	26522	0.239
body	breath	-0	1	-0	-0	-0	-0	1.886	-1	-1	-1	5098	0.488
body	face	0	-1	0	-1	0	0	0.124	-1	-1	-1	97417	0.174
body	head	-0	-1	-0	-0	-0	-0	-0.029	-1	-1	-1	172950	0.151
body	mind	-0	1	-0	-0	-0	-0	0.133	-1	-1	-1	820530	0.417

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
bombers	missiles	0	0	0	0	0	0	-0.498	0	0	-0	7505	0.388
bone	flesh	-0	-1	-0	-0	-0	-0	0.476	-0	-0	-0	72784	0.219
bone	muscle	0	0	0	0	0	0	0.083	1	1	1	59672	0.563
bones	skulls	1	0	0	-1	0	0	2.505	0	0	0	19232	0.393
boots	jacket	0	-1	0	0	0	0	-0.002	1	1	1	4308	0.363
boots	shoes	-1	-0	-0	-0	-0	-0	-0.624	-0	-0	-0	81005	0.837
boredom	loneliness	-0	-1	-0	-0	-1	-0	-0.674	1	1	0	6246	0.451
boulders	ledges	0	-1	0	0	0	0	0.891	0	0	-0	1599	0.476
bouquets	wreaths	0	0	1	0	0	0	-0.438	0	0	0	2269	0.376
bourbon	water	0	0	1	0	0	0	-4.845	0	0	-0	5555	0.984
boys	girls	0	0	0	1	0	0	0.056	0	0	0	1835429	0.883
boys	men	-0	-0	-0	-1	-0	-0	-1.892	-0	-0	-0	263583	0.252
brain	intelligence	0	1	0	0	0	0	0.337	2	2	1	5151	0.755
brains	talent	0	0	0	0	0	0	-0.763	1	1	1	2145	0.690
bread	butter	0	0	1	0	0	0	0.479	1	1	1	261886	0.987
bread	cheese	0	0	1	0	0	0	0.676	0	0	0	105642	0.896
breakfast	dinner	0	0	-1	0	0	1	-0.644	0	0	-0	53396	0.909
breakfast	luncheon	0	0	0	0	0	1	1.914	0	0	-0	3299	0.929
brick	glass	0	0	1	0	0	0	-1.526	0	0	0	5652	0.734
bride	groom	0	0	1	-1	0	0	1.215	0	0	0	204540	0.969
brother	mother	0	0	-1	-1	0	0	-0.939	0	0	-0	57486	0.129
brother	sister	0	0	0	1	0	0	0.396	0	0	-0	492450	0.838
brothers	sisters	0	0	0	1	0	0	0.570	0	0	-0	1102312	0.914
brow	cheeks	0	1	0	0	0	0	-0.380	0	0	0	5773	0.449
brush	logs	-0	-1	-1	-0	-0	-0	0.779	-0	-0	-0	4632	0.261
brush	trees	0	-1	-1	0	0	0	-1.397	0	0	0	35404	0.333
burma	india	-0	-0	-1	-0	-0	-0	-2.454	1	1	0	20258	0.258

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
bushes	vines	0	0	1	0	0	0	0.267	-1	-1	-1	7099	0.501
c.	d.	-0	-0	-0	-0	-0	-1	0.204	-0	-0	-0	20006	0.629
cakes	pies	0	0	1	0	0	0	0.817	0	0	0	22636	0.503
calluses	corns	-0	-0	-0	-0	-1	-0	-0.609	-2	-2	-1	5405	0.180
candies	cookies	-0	-0	-0	-0	-0	-0	-1.805	-0	-0	0	2837	0.352
captain	leader	-1	-0	-0	-0	-0	-0	0.186	-0	-0	0	2031	0.587
caresses	endearments	-0	-0	-0	-0	-0	-0	1.598	-0	-0	0	1151	0.705
cattle	horses	-0	-0	-0	-0	-0	-0	-0.448	-0	-0	0	164491	0.495
cause	consequence	0	0	0	0	0	1	1.574	2	2	1	11492	0.942
cause	effect	0	0	0	0	0	1	-0.292	0	0	0	679208	0.989
ceiling	walls	-0	-1	-0	-0	-0	-0	-1.302	-1	-1	-1	90509	0.197
celebration	ritual	-1	-0	-0	-0	-0	-0	-0.598	-1	1	0	1211	0.300
cents	dollars	-0	-0	-1	-1	-0	-0	-0.978	1	1	1	127135	0.003
certainty	order	0	0	0	0	0	0	-3.273	-1	-1	-0	2092	0.323
certainty	security	0	0	0	0	0	0	-1.987	0	0	-0	5706	0.531
chagrin	embarrassment	-1	-0	-0	-0	-0	-0	-1.568	2	2	1	1283	0.506
chair	table	-0	-0	-1	-0	-0	-0	-1.691	1	1	1	34607	0.339
chairs	sofas	0	1	1	0	0	0	2.963	1	1	1	35053	0.561
chairs	table	-0	-1	-1	-0	-0	-0	-2.973	1	1	1	56411	0.106
character	wisdom	-0	-0	-1	-0	-0	-0	1.276	-1	-1	0	3043	0.513
characters	situations	0	1	0	0	0	0	0.092	1	-1	-0	32957	0.818
charm	variety	-0	-0	-0	-0	-0	-1	-1.786	2	2	1	2705	0.438
charts	maps	-0	-0	-1	-0	-0	-0	-0.574	-0	-0	-0	51057	0.240
checks	money	-1	0	0	0	0	0	-2.714	1	1	1	7950	0.795
chest	stomach	0	1	0	0	0	0	0.287	1	1	1	25203	0.711
child	husband	-0	-0	-0	-1	-0	-0	1.057	1	1	1	20834	0.099
child	wife	-0	-0	-0	-1	-0	-0	0.542	-0	-0	-0	166893	0.022

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
childhood	infancy	-0	-0	-0	-0	-0	-1	1.410	1	2	1	103215	0.018
choice	decision	-0	-0	-0	-0	-1	-0	-0.255	1	1	1	11325	0.761
chopping	sawing	-0	1	-0	-0	-0	-0	0.378	-0	-0	0	1425	0.494
christmas	easter	0	0	1	0	0	1	1.317	0	0	-0	47337	0.787
chrome	glass	0	0	1	0	0	0	-3.578	0	0	0	7398	0.551
church	community	-1	0	0	0	0	0	0.328	2	2	1	39253	0.830
church	graveyard	0	1	1	0	0	0	4.674	1	0	0	2319	0.962
city	country	0	0	0	0	0	1	0.148	0	0	-0	106397	0.785
city	county	0	0	0	0	0	1	0.829	0	0	-0	315000	0.821
city	state	0	0	0	0	0	1	-0.657	-1	-1	-1	184793	0.648
clarity	precision	0	0	0	0	0	0	-0.410	-1	-1	-0	28178	0.710
classmates	friends	-0	-0	-1	-0	-0	-0	-3.638	-1	-0	-0	13278	0.325
clay	wax	0	0	1	0	0	0	1.144	0	0	0	2206	0.521
cliches	stereotypes	0	0	0	0	0	0	-1.520	3	0	0	2899	0.595
clippings	pictures	-0	-0	-1	-0	-0	-0	-3.334	-0	-0	0	2541	0.536
closets	drawers	-0	1	-0	-0	-0	-0	-0.961	-1	-1	-1	10283	0.483
clothes	dishes	0	0	1	0	0	0	1.196	1	1	1	4140	0.552
coat	hat	-0	-1	1	-0	-0	-0	-0.084	-0	-0	-0	132073	0.439
cod	salmon	0	0	1	0	0	0	-0.613	1	1	1	1472	0.503
coffee	sandwiches	-0	-0	-1	-0	-0	-0	2.325	1	1	0	19962	0.517
coffee	toast	0	0	-1	0	0	0	1.992	-1	-1	-1	19739	0.458
college	school	-1	-0	-1	-0	-0	-1	-0.886	-1	-1	-1	301287	0.028
college	university	0	0	0	-1	0	0	-0.730	3	1	-0	296202	0.892
colonel	mrs.	0	0	0	1	0	0	-1.318	0	0	-0	18072	0.998
color	imagery	-1	0	0	0	0	0	2.346	1	1	-0	1105	0.650
color	light	-0	-1	-0	-0	-0	-0	-0.899	-1	-1	-1	83699	0.320
color	scent	0	0	1	0	0	0	2.701	-1	-1	-1	4058	0.680

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
color	texture	-0	-0	1	-0	-0	-0	2.293	-0	-0	0	90359	0.761
colors	forms	0	0	-1	0	1	0	-1.499	-1	-1	-1	31064	0.342
comb	wattles	0	1	1	0	0	0	2.991	1	1	1	4962	0.953
comedy	humor	-0	-0	-0	-0	-0	-0	-0.197	-1	-1	0	2098	0.413
comedy	satire	1	1	0	0	0	0	0.981	0	-1	-0	5139	0.682
comfort	companionship	0	0	0	0	0	0	2.073	1	-1	-1	4138	0.773
comfort	ease	-0	-0	-0	-0	-0	-0	0.134	-1	-1	-1	52544	0.288
comings	goings	0	-1	0	0	0	1	-0.473	0	0	-0	141234	0.913
command	control	0	0	0	-1	0	0	-1.239	0	0	0	170958	0.974
commands	statements	0	0	0	1	0	0	-1.028	1	1	1	1897	0.405
commercialism	materialism	0	0	0	0	0	1	-1.963	1	0	-0	1074	0.512
completeness	unity	0	0	0	0	0	0	-2.214	1	1	-0	4700	0.232
complexity	obscurity	-0	-0	-0	-0	-0	1	1.794	-0	-0	0	1319	0.650
comprehension	control	0	0	0	-1	0	0	-3.342	-3	-1	-1	1993	0.772
concord	lexington	-0	-1	-0	-0	-0	-1	-0.212	1	2	1	55328	0.137
concrete	steel	-0	-1	-0	-0	-0	-0	-0.427	-1	-1	-1	83832	0.451
conduct	manners	-1	1	-0	-0	-0	-0	1.405	-0	-0	0	5138	0.396
confusion	congestion	0	0	0	0	0	1	1.876	0	0	-0	1755	0.405
confusion	fear	0	0	0	0	-1	0	-1.222	-1	-1	-1	27042	0.386
confusion	turmoil	0	1	0	0	-1	0	1.874	0	-1	-1	11154	0.348
confusions	contradictions	0	0	0	0	-1	-1	-1.891	2	0	-0	4304	0.556
conscience	religion	0	0	1	0	0	0	-1.399	0	0	-0	15154	0.700
contemporary	friend	-0	-0	-1	-0	-0	-0	-0.975	-3	-1	-1	13050	0.521
contraction	dilatation	0	0	0	0	1	0	1.356	2	0	-0	3691	0.611
contraction	dilation	0	0	0	0	1	0	1.881	0	0	-0	3161	0.462
control	growth	-0	-0	-0	1	-0	-0	0.676	-0	-0	-0	3621	0.180
control	intelligence	-0	-0	-0	1	-0	-0	1.729	2	2	1	2022	0.533

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
corruption	crime	-0	-0	-0	-0	-0	-1	-1.346	-1	-1	-1	16627	0.191
cost	upkeep	0	0	0	0	0	1	5.052	1	0	0	1881	0.939
costumes	masks	0	-1	0	0	0	0	-0.031	-1	-1	-1	12920	0.441
cotton	tobacco	0	0	1	0	0	0	0.547	0	-1	-1	40004	0.710
council	governor	-0	-1	-0	-1	-0	-0	0.444	1	1	0	100783	0.011
counterpoint	harmony	1	0	-1	0	0	0	-2.491	0	2	1	14622	0.125
counties	towns	-0	-0	-1	-0	-0	-0	-0.639	-1	-1	-1	26566	0.434
county	village	0	0	-1	0	0	0	0.692	0	0	-0	2834	0.429
courage	hope	0	0	0	0	0	0	-1.460	-1	-1	-1	37808	0.432
cousins	uncles	-0	-0	-0	-1	-0	-0	1.100	-0	-0	0	21181	0.185
crannies	nooks	-0	-1	-0	-0	-0	-0	-0.555	-1	-1	-1	46060	0.003
cream	peaches	-0	-0	-1	-0	-0	-0	2.236	1	1	1	13614	0.019
crime	poverty	-0	-0	-1	-0	1	-0	0.423	2	2	1	21270	0.373
crime	punishment	0	0	0	0	0	1	0.560	2	2	1	131945	0.989
crime	violence	1	0	0	0	0	0	0.034	2	2	1	63327	0.730
crises	problems	-1	-0	-0	-0	-0	-0	-3.363	-0	-0	0	6694	0.254
crops	livestock	0	-1	0	0	0	0	0.917	1	0	0	41356	0.881
curettage	dilatation	-0	-0	-0	-0	-0	-1	-1.739	1	-1	0	15649	0.000
cuts	wounds	-1	0	0	0	0	0	0.137	0	0	0	5438	0.777
cysts	tumors	-0	-0	-1	-0	-1	-0	-1.416	1	1	1	11633	0.497
dairy	poultry	0	0	0	0	0	0	0.387	0	0	-0	13945	0.690
danger	threat	1	-0	-0	-0	-0	-0	0.494	-1	-1	-1	4180	0.431
dark	light	-0	-1	-0	-0	-0	-0	-0.846	-0	-0	-0	292691	0.245
dates	events	1	0	1	0	0	0	-1.515	0	0	0	15058	0.734
daughter	son	-0	-0	-0	-1	-0	-0	-0.702	-1	-1	-1	190092	0.244
day	night	0	1	0	0	0	0	0.858	0	0	0	1876827	0.698
days	nights	0	1	0	0	0	0	2.673	0	0	0	445957	0.914

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
days	years	0	0	1	0	0	1	-0.850	0	0	0	31764	0.879
death	life	-0	-1	-1	-0	-0	-1	-1.058	-0	-0	-0	1142991	0.062
debate	discussion	-1	-0	-0	-0	-0	-0	-1.205	1	1	1	59247	0.308
december	may	-0	-0	-0	-0	-0	-1	-2.845	-1	-1	-1	9064	0.302
democracy	liberty	1	-0	-0	-0	-0	-0	-0.075	-0	-0	0	29095	0.383
dentist	doctor	-0	-0	-1	-0	-0	-0	-2.737	-0	-0	0	5556	0.148
depth	force	-0	-0	-0	-1	-0	-0	-1.487	-0	-0	-0	4062	0.616
desolation	misery	-0	-0	-0	-0	1	-0	-1.466	-1	1	0	4427	0.372
despair	futility	-0	-0	-0	-0	1	-0	1.868	2	2	1	2723	0.366
despair	rage	-0	-0	-0	-0	-1	-0	0.114	-0	-0	-0	17146	0.173
detection	removal	0	0	0	0	0	1	-0.484	0	0	-0	3333	1.000
detective	police	0	0	0	-1	0	0	-2.046	-1	-1	-1	2346	0.220
determination	faith	-0	-0	-0	-0	-0	-0	-0.830	-3	-1	-1	4173	0.260
development	practice	0	0	0	0	0	0	0.607	-1	-1	-0	7770	0.719
development	research	-0	-0	-0	-0	-0	-1	0.223	-1	-1	0	1380352	0.019
devotion	effort	-0	-1	-0	-0	-0	-0	-1.819	-0	-0	0	1231	0.310
diagnosis	treatment	0	0	0	0	0	1	-1.439	-2	0	-0	448278	0.991
diagrams	sketches	0	0	0	0	0	0	-0.020	-1	1	1	2961	0.391
diamonds	rubies	-0	-0	1	-0	-0	-0	1.870	-0	-0	0	18750	0.609
dirt	sweat	-0	-1	-0	-0	-0	-0	0.156	-0	-0	-0	11621	0.477
discretion	moderation	0	0	0	0	0	0	1.198	2	0	-0	2059	0.469
discussion	thought	-0	1	-0	-0	-0	-1	-1.241	-1	-1	-1	24358	0.076
disdain	indifference	0	-1	0	0	1	0	-1.083	2	2	1	1129	0.327
distress	pain	1	-1	-0	-0	-0	-0	-1.771	-0	-0	-0	24052	0.180
doctors	fathers	-0	-0	-0	-0	-0	-0	-0.085	-0	-0	0	5235	0.127
dogs	men	-0	-1	-0	-0	-0	-0	-2.838	-0	-0	-0	24614	0.349
domination	influence	0	0	0	1	0	0	-2.696	-1	1	-0	2524	0.384

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
doorways	windows	0	0	1	0	0	0	-3.645	0	1	1	14085	0.498
doubts	fears	0	0	0	0	-1	0	-0.538	0	0	0	71893	0.772
doubts	suspicions	0	0	0	0	0	1	0.919	1	1	1	7417	0.833
dreams	plans	-0	-1	-0	-0	-0	-0	-0.798	-0	-0	-0	11614	0.483
drudgery	monotony	0	0	0	0	-1	0	-0.545	0	0	-0	1332	0.495
ears	eyes	-0	-0	-1	-0	-0	-0	-2.090	-0	-0	-0	304481	0.129
ears	hair	-0	-1	1	-0	-0	-0	-1.121	-0	-0	-0	4651	0.433
earth	sky	0	1	0	0	0	0	1.093	0	0	0	131176	0.784
economist	historian	-0	-1	-0	-0	-0	-0	-1.020	-0	-0	0	3086	0.474
editing	publishing	0	0	0	0	0	1	-1.201	0	0	-0	11644	0.875
editor	publisher	0	0	0	0	0	1	0.838	0	0	-0	125503	0.842
education	industrialization	0	0	1	0	0	0	3.687	2	0	-0	1068	0.624
education	skill	-0	-0	-1	-0	-0	-0	1.816	-3	-1	-1	10876	0.727
effort	resources	0	0	0	0	0	0	-0.074	1	1	-0	12297	0.742
eggs	ham	-0	-0	-0	-0	-0	-0	1.311	-0	-0	-0	59981	0.183
egypt	europe	-0	-0	-1	-0	-0	-0	-1.176	-0	-0	0	3595	0.431
elbows	knees	0	1	0	0	0	0	-1.654	-1	0	0	48397	0.499
elegance	purity	0	0	0	0	0	0	-1.080	0	0	-0	3859	0.324
eloquence	force	-0	-0	-0	-1	-0	-0	-3.589	-2	-2	-1	4562	0.278
emotion	meaning	0	0	1	0	0	0	-1.608	0	0	-0	7218	0.569
ends	odds	-0	-0	-0	-0	-0	-0	1.597	-0	-0	-0	188194	0.004
energy	money	-0	-1	-0	-0	-0	-0	-0.307	-1	-1	0	31490	0.527
engineers	scientists	-0	-0	-0	-0	-0	-1	-0.243	-0	2	1	229487	0.251
english	german	0	0	1	0	0	0	0.442	0	0	-0	148633	0.568
enthusiasm	warmth	-0	-0	-0	-0	1	-0	0.494	-3	-1	-1	5100	0.170
environment	heredity	-0	-0	-1	-0	-0	-0	2.961	-0	-0	0	89156	0.079
equipment	personnel	0	-1	0	0	0	0	0.651	1	-1	-1	64626	0.414

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
error	trial	-0	-1	-0	-0	-0	-1	-0.046	-0	-0	0	454209	0.000
errors	omissions	0	1	0	0	0	0	2.620	0	0	-0	58890	0.937
evening	morning	-0	-1	-0	-0	-0	-1	-0.534	-0	-0	0	282270	0.076
evenings	sundays	0	0	1	0	0	0	0.459	0	-1	-1	3744	0.822
everybody	everything	-0	1	-0	-0	-0	-0	-1.442	-1	-1	-1	88771	0.363
example	precept	-0	-0	-0	-1	-0	-0	5.181	-0	-1	-1	37431	0.188
excitement	fatigue	0	1	0	0	0	0	0.402	-1	-1	-1	6150	0.447
expense	time	0	1	0	0	0	0	-3.611	0	0	0	78761	0.093
experiences	reactions	0	0	0	0	0	1	0.237	-2	-2	-0	5219	0.767
eyes	face	-1	-0	-0	-0	-0	-0	-0.012	-0	-0	-0	61214	0.313
eyes	hair	-0	-1	1	-0	-0	-0	0.970	-0	-0	-0	108355	0.350
face	figure	0	1	0	0	0	0	-0.267	1	1	1	57306	0.923
factories	industries	-0	1	-0	-1	-0	-0	-1.027	-0	-0	0	3465	0.661
facts	figures	0	0	1	0	0	0	0.000	1	1	1	183738	0.978
facts	techniques	0	0	0	0	0	0	0.167	0	0	0	1703	0.901
fairness	moderation	-0	-0	-0	-0	-0	-0	0.412	2	-0	0	1517	0.612
faith	life	-0	-0	-1	-0	-0	-0	-1.949	-0	-0	-0	44519	0.696
faith	love	-0	-0	-1	-0	-0	-0	-1.023	-0	-0	-0	130104	0.690
fall	rise	-0	-1	-0	-0	-0	-0	0.335	-0	-0	-0	746325	0.030
fall	winter	0	0	0	0	0	1	0.501	1	1	1	174361	0.987
family	neighborhood	-0	-0	1	-0	-0	-0	2.507	-0	-2	-1	17862	0.836
farms	fields	-0	-0	1	-0	-0	-0	-1.340	-0	-0	-0	11526	0.437
father	mother	-0	-0	-1	1	-0	-0	0.092	-0	-0	0	1590868	0.602
father	son	0	0	0	1	0	0	0.413	-1	-1	-1	486090	0.968
father	uncle	0	0	1	0	0	0	1.974	0	0	-0	52100	0.909
fathers	mothers	0	0	-1	1	0	0	-0.132	0	0	-0	315907	0.492
fear	relief	0	-1	0	0	1	0	0.673	0	0	0	2765	0.646

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
fear	want	-0	-1	-0	-0	1	-0	-1.077	-0	-0	-0	9491	0.642
feelings	thoughts	-0	-0	-0	-0	1	-0	0.057	-1	-1	-1	480072	0.151
feet	hands	-0	-1	-0	-0	-0	-0	-0.044	-0	-0	-0	574548	0.118
feet	heads	-0	-1	-0	-0	-0	-0	1.340	-0	-0	-0	7570	0.212
feet	legs	-0	-1	-0	-0	-0	-0	1.336	-0	-0	-0	194658	0.389
female	male	-0	-0	-0	-1	-0	-0	-0.010	-1	-0	-0	1742846	0.074
fiction	poetry	-0	-0	-0	-0	-0	-0	-0.553	1	1	0	65232	0.501
field	stream	0	1	0	0	0	0	1.558	0	0	0	15417	0.964
fields	homes	0	0	-1	0	0	0	0.609	0	0	0	6083	0.307
figures	words	-0	-0	-1	-0	-0	-0	-1.247	-1	-1	-1	19044	0.117
file	rank	-0	1	-0	-0	-0	-0	0.945	-0	-0	-0	500989	0.000
fingers	mouth	0	-1	0	0	0	0	-0.695	-1	-1	-1	2949	0.381
fingers	thumb	-0	-1	-0	-1	-0	-0	1.162	-1	-1	-1	59440	0.454
fire	light	0	0	0	0	1	0	-0.656	-1	-1	-1	21101	0.562
fire	police	0	-1	0	0	0	0	0.539	-1	-1	-1	126094	0.278
firepower	speed	-0	-0	-0	-0	1	-0	-4.528	-2	-1	-1	1016	0.199
fish	vegetable	0	1	1	0	0	0	1.609	2	2	1	5385	0.667
fits	starts	0	0	0	0	1	0	-0.657	0	0	0	71456	0.995
flame	smoke	0	0	1	0	0	1	-0.541	0	0	0	45767	0.448
flames	smoke	-0	-0	1	-0	-0	1	-1.137	-0	-0	-0	33304	0.429
flames	sparks	-0	-0	-0	-0	1	-0	0.735	-0	-0	-0	4256	0.482
flesh	skin	0	0	0	0	0	0	-0.913	0	0	0	27576	0.354
flight	terror	-0	1	-0	-0	-0	-1	0.691	1	1	1	1152	0.298
flowers	hearts	-0	-0	-1	-0	-0	-0	0.618	-1	-1	-1	9516	0.050
flowers	herbs	0	0	1	0	0	0	1.896	-1	-1	-1	27473	0.418
flowers	plants	-1	0	0	0	0	0	-0.403	-1	-1	-1	78739	0.385
fog	mist	-0	-0	-0	-0	1	-0	0.108	-0	-0	-0	15420	0.571

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
folklore	legend	0	0	0	0	-1	0	-0.950	0	1	1	5914	0.518
food	shelter	0	0	1	0	0	0	2.348	1	1	1	147749	0.877
force	violence	1	0	0	0	0	0	1.309	2	2	1	63181	0.922
foreboding	gloom	-0	-0	-0	-0	1	-0	-1.603	-1	-1	-1	1057	0.136
forefinger	thumb	-0	-0	1	-1	-0	-0	-1.902	-2	-1	-1	193360	0.103
fork	spoon	0	1	1	0	0	0	0.402	0	0	0	20391	0.654
forms	materials	0	0	0	0	0	0	0.347	2	2	1	10721	0.566
fragrance	taste	-0	-0	-1	-0	-0	-0	-2.453	-1	-1	-1	1816	0.456
freedom	security	-0	-0	-0	-0	-0	-0	-0.145	1	1	0	30628	0.584
friends	partners	-0	-0	-1	-0	-0	-0	1.773	1	1	1	10050	0.613
friends	relations	0	0	-1	0	0	0	0.218	1	1	1	83205	0.674
fruit	nuts	0	0	1	0	0	0	1.498	0	0	0	20149	0.795
fruits	vegetables	0	0	0	0	0	0	0.128	2	2	1	667661	0.850
fulfillment	pleasure	1	0	0	0	0	0	-1.863	0	0	-0	2775	0.206
fun	games	1	-1	0	0	0	0	-0.095	0	0	0	54491	0.975
fun	pleasure	-1	0	0	0	0	0	-0.579	1	1	1	6132	0.685
fungicides	insecticides	-0	-1	-1	-0	-0	-0	-1.091	-0	-0	-0	11397	0.218
funnels	masts	-0	-0	-1	-0	-0	-0	-1.339	-1	-1	-1	2592	0.258
fury	pain	-0	-1	-0	-0	-0	-0	-2.365	-1	-1	-1	5310	0.290
future	past	-0	-0	-0	-0	-0	-1	-0.107	-1	-1	-1	283153	0.042
gaiety	spirit	0	0	0	0	1	0	-4.100	-1	-1	-0	1148	0.391
gambling	vice	-1	0	0	0	0	0	-1.591	-2	-2	-1	3997	0.505
games	races	1	0	0	0	0	0	0.696	1	1	1	3815	0.559
gardens	lawn	-0	-0	-1	-0	-0	-0	0.685	-1	-1	-1	2244	0.208
gas	light	0	0	0	0	0	0	-1.150	0	0	0	5578	0.485
gas	oil	-0	-1	-0	-0	-0	-0	-0.310	-0	-0	-0	1040905	0.085
generalists	specialists	1	0	0	0	0	0	-3.537	-1	-1	-0	6474	0.538

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
gentlemen	ladies	-0	-0	-1	1	-0	-0	-0.284	-1	-1	0	501922	0.091
gentleness	sweetness	0	0	0	0	0	0	-0.567	-1	-1	-0	5275	0.466
gin	tonic	0	0	1	0	0	0	-0.019	1	1	1	36799	0.998
glass	rubber	-0	1	-0	-0	-0	-0	1.320	1	1	1	4071	0.608
glass	steel	-0	-1	-0	-0	-0	-0	0.319	-0	-0	-0	41200	0.463
god	man	0	0	0	1	0	0	-0.518	0	0	0	501367	0.824
goods	services	0	1	0	0	0	0	-0.671	2	2	1	1553253	0.993
grace	modesty	0	1	0	0	0	0	2.455	2	2	1	2152	0.587
grains	seeds	0	0	1	0	0	0	-0.545	0	0	0	8920	0.592
grapefruit	oranges	-0	-0	-1	-0	-0	-0	-0.959	1	2	1	7753	0.190
gratitude	joy	-0	-0	-0	-0	-1	-0	-1.273	-2	-0	-0	30564	0.281
grease	sweat	-0	-1	-0	-0	-0	-0	-0.896	-0	-0	-0	2001	0.526
groups	individuals	-0	-1	-0	-0	-0	-0	0.600	4	2	1	441733	0.252
guidance	information	-0	-0	-0	-0	-0	-1	-2.598	2	-0	0	32507	0.176
gums	teeth	0	0	-1	0	0	0	-2.957	0	0	0	36148	0.174
habitat	habits	-0	-1	-0	-0	-0	-0	-1.016	-1	1	1	5478	0.401
habits	movements	0	0	0	0	0	0	-0.597	0	0	-0	2899	0.348
habits	thought	-0	-0	-0	-0	-0	-0	-2.616	-1	-1	-1	2299	0.419
hair	skin	0	1	0	0	0	0	0.182	0	0	0	93412	0.387
hallucinations	visions	-1	-0	-0	-0	-0	-0	-1.197	-2	-0	0	1974	0.274
hands	knees	0	1	0	0	0	0	2.089	0	0	0	339150	0.981
happiness	laughter	-0	-1	-0	-0	-0	-0	0.551	-1	-1	0	3072	0.414
happiness	pain	0	1	0	0	0	0	-1.103	-2	-2	-1	3609	0.600
happiness	warmth	-0	-0	1	-0	-0	-0	0.957	-2	-2	-1	2791	0.182
hardships	trials	-0	-0	-0	-0	-0	-0	-1.572	-0	-0	0	8611	0.338
head	shoulders	0	1	0	0	0	0	2.263	1	1	1	252972	0.955
head	tail	0	1	0	0	0	0	2.410	0	0	0	73606	0.945

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
head	teeth	1	1	0	0	0	0	1.925	0	0	0	1481	0.740
health	spirits	0	1	0	0	0	0	2.006	1	1	1	50385	0.964
hearing	sight	-0	-0	-1	-0	-0	-0	-0.203	-1	-1	-1	79594	0.138
heart	mind	-0	-1	-0	-0	-0	-0	-0.210	-0	-0	-0	433049	0.466
heir	son	-0	-0	-1	1	-0	-0	-2.953	-0	-0	-0	137375	0.003
help	understanding	0	1	0	0	0	0	0.851	3	1	1	10362	0.556
history	identity	0	0	0	0	0	1	1.678	0	0	-0	13980	0.671
honey	milk	-0	-0	-1	-0	-0	-0	-1.142	-1	-1	-1	113327	0.056
horses	wagons	0	1	0	0	0	0	1.912	0	0	-0	28583	0.726
hostility	indifference	0	-1	0	0	1	0	0.164	0	0	-0	4279	0.314
hotels	pensions	1	0	0	0	0	0	0.915	1	1	1	4962	0.907
hours	miles	-0	-1	-0	-0	-0	-0	0.236	-0	-0	0	3820	0.414
houses	shops	-0	-0	1	-0	-0	-0	1.348	-1	-1	-1	33872	0.614
humility	pride	-0	-0	-0	-0	-1	-0	-1.629	-2	-2	-1	12307	0.303
husband	wife	0	0	0	1	0	0	-0.514	-1	-1	-1	1075512	0.975
husbands	wives	0	0	0	1	0	0	-0.653	-1	-1	-1	272233	0.896
ice	snow	0	1	0	0	0	0	0.196	0	0	0	278789	0.409
ice	water	0	1	0	0	0	0	-2.073	1	1	1	54953	0.517
illinois	missouri	-0	-1	-0	-0	-0	-0	0.358	-1	1	1	25009	0.557
imagination	originality	-0	-0	-0	-0	-0	-0	2.042	1	1	0	4763	0.493
importance	validity	-0	-0	-0	-0	-0	-1	1.583	1	1	0	4610	0.394
improvements	land	-0	-0	-1	-0	-0	-1	-2.485	-1	-1	-1	17036	0.039
inclination	time	-0	-0	-0	-0	-0	-0	-5.135	-3	-1	-1	10476	0.025
individualism	socialism	0	1	0	0	0	0	-0.920	-3	0	-0	6126	0.655
influence	power	-0	-0	-0	-1	-0	-0	-1.060	-1	-1	0	314032	0.189
influence	prestige	-0	-0	-0	1	-0	-0	2.461	-2	-2	-1	40191	0.346
information	material	0	-1	0	0	0	0	0.689	-1	1	-0	17241	0.488

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
initials	names	0	0	-1	0	0	0	-3.234	-1	-1	-1	5373	0.274
ink	paper	-0	-1	-0	-0	-0	-0	-2.541	1	1	1	58599	0.472
integrity	unity	-0	-0	-0	-0	-0	-0	-0.687	-0	-0	0	9540	0.254
interest	payments	0	0	-1	0	0	0	2.061	-1	-1	-0	4047	0.225
intonation	phrasing	0	0	-1	0	0	0	-0.025	0	0	-0	1097	0.395
investment	risk	0	0	0	0	0	1	-0.446	-1	-1	-1	3621	0.663
iron	wood	-0	-0	-0	1	-0	-0	0.027	-1	-1	-1	29016	0.299
irony	satire	-1	-0	-0	-0	-0	-0	0.438	-0	-1	0	8231	0.588
jacket	trousers	0	1	0	0	0	0	0.821	0	0	-0	18202	0.789
jams	jellies	0	0	0	0	0	0	0.509	1	1	1	24827	0.803
jaws	teeth	-0	-1	-1	-0	-0	-0	-2.085	-0	-0	-0	26610	0.521
joy	pleasure	-0	-0	-0	-0	-0	-0	-0.245	1	1	1	28376	0.659
july	june	-0	-0	-0	-0	-0	-1	-0.100	-0	-0	-0	154093	0.001
kind	quantity	0	0	1	0	0	0	1.535	2	2	1	18201	0.735
kingston	montreal	-0	-1	-0	-0	-0	-0	-0.733	1	-1	-1	7177	0.365
kitchen	pantry	1	0	0	0	0	0	3.177	0	0	-0	8200	0.787
knowledge	skills	-0	-0	-1	-0	-0	-0	1.139	-1	-1	-1	437430	0.669
labor	money	-0	-0	-0	-0	-0	-0	-0.377	-0	-0	0	17606	0.450
laborers	servants	-0	-0	-0	-0	-0	-0	-0.969	-1	-1	0	7407	0.447
land	sea	0	1	0	0	0	0	0.589	0	0	0	413757	0.689
landings	takeoffs	-0	-0	-0	-0	-0	-1	2.582	-0	-1	-1	17024	0.229
law	order	0	1	0	0	0	1	-0.057	1	1	1	682080	0.985
law	sovereignty	0	0	0	0	0	0	3.319	2	2	1	2097	0.629
leather	paper	-0	1	-1	-0	-0	-0	-2.072	-0	-0	0	3714	0.460
leaves	limbs	-0	-0	-1	-0	-0	-0	1.948	-0	-0	-0	3832	0.306
lecturer	writer	-0	-0	-1	-0	-0	-0	-2.698	-1	-1	0	15994	0.365
left	right	-0	-1	-0	-0	-0	-0	-0.327	-0	-0	-0	1786858	0.433

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
legs	torso	-0	-1	-1	-0	-0	-0	2.899	1	-0	-0	8829	0.484
length	quality	0	1	1	0	0	0	0.010	2	2	1	10492	0.625
length	weight	0	1	0	0	0	0	0.153	0	0	0	30128	0.554
letters	magazines	0	0	1	0	0	0	1.845	1	-1	-1	1182	0.646
liberalism	progress	0	0	0	0	0	0	-2.446	-3	0	-0	1123	0.656
libraries	museums	0	0	1	0	0	0	1.071	-1	0	-0	33304	0.630
life	promise	0	0	1	0	0	0	2.773	1	1	1	1481	0.869
life	religion	0	0	1	0	0	0	2.069	1	1	1	20659	0.568
light	shadow	0	0	0	0	0	1	2.229	1	0	0	128575	0.908
light	sound	0	0	1	0	0	0	0.751	0	0	0	68749	0.590
linen	silk	0	0	1	-1	0	0	-0.800	-1	-1	-1	9625	0.458
liquids	powders	0	-1	1	0	0	0	0.976	0	0	-0	2530	0.433
location	size	0	0	1	0	0	0	-0.890	-1	-1	-1	104236	0.313
love	veneration	-0	-0	1	-0	-0	-0	4.925	3	1	1	11653	0.757
machinery	power	-0	1	1	-0	-0	-0	-2.772	-1	-1	0	11022	0.272
machinery	tractors	1	-0	-1	-0	-0	-0	2.498	-1	-1	0	2213	0.483
magazines	newspapers	0	0	0	0	0	0	-0.600	0	1	1	378784	0.300
magic	superstition	0	1	0	0	0	0	1.669	2	0	-0	11125	0.699
malaria	typhoid	-0	-0	-0	-0	-0	-0	0.260	-1	-1	0	3188	0.642
mamma	papa	-0	-0	1	-1	-0	-0	-1.010	-0	-0	0	30494	0.266
man	nature	0	1	0	0	0	0	1.063	1	1	1	178124	0.723
man	wife	0	0	0	1	0	0	1.533	0	0	0	213993	0.998
man	woman	0	0	0	1	0	0	1.178	1	1	1	731849	0.961
math	science	0	0	0	0	0	0	-2.359	1	1	1	123868	0.765
meaning	origin	-0	-0	-0	-0	-0	-1	0.613	1	1	0	30841	0.155
means	motives	0	1	0	0	0	-1	2.781	1	1	1	2767	0.510
men	objects	0	1	0	0	0	0	1.985	1	1	1	1870	0.743

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
men	officers	-0	-0	-0	-1	-0	-0	1.961	2	2	1	457386	0.091
men	women	0	0	0	1	0	0	0.358	1	1	1	7628255	0.906
methods	principles	-0	1	-1	-0	-0	-0	0.392	1	1	0	107758	0.158
minerals	vitamins	0	0	0	0	0	0	0.779	0	0	-0	179421	0.142
misery	pain	-0	-1	-0	-0	-0	-0	-2.118	-2	-2	-1	24920	0.279
money	numbers	-0	-1	-0	-0	-0	-0	0.741	-0	-0	0	1671	0.367
money	papers	0	0	1	0	0	0	1.119	0	0	-0	4570	0.542
money	time	-0	-0	-0	-0	-0	-0	-1.824	-1	-1	-1	572131	0.103
months	years	0	0	0	0	0	1	-1.537	0	0	0	125753	0.893
moon	sun	-0	-0	-1	-0	-0	-0	-0.935	-0	-0	-0	322714	0.089
morning	night	0	1	0	0	0	1	-0.551	-1	-1	-1	101769	0.395
mother	son	0	0	0	1	0	0	0.321	-1	-1	-1	128987	0.936
mouth	nostrils	0	-1	1	0	0	0	2.970	1	1	1	23816	0.770
movements	positions	-0	-1	-0	-0	-0	-1	0.012	-0	-0	0	13056	0.333
movements	speech	0	-1	1	0	0	0	-0.587	-1	-1	-1	3639	0.611
mr.	mrs.	0	0	0	1	0	0	0.649	0	0	-0	1680824	0.998
muscles	nerves	0	0	1	1	0	0	0.567	-1	-1	-1	55557	0.416
muskets	swords	0	0	0	1	0	0	-1.178	-1	-1	-1	3049	0.401
name	year	0	0	1	0	0	0	-0.332	0	0	0	3444	0.938
names	numbers	-0	-0	-1	-0	-0	-0	-0.202	1	1	1	48724	0.821
nationalism	sovereignty	0	0	1	-1	0	0	-0.049	-2	1	-0	1597	0.500
nature	size	-0	-1	-0	-0	-0	-0	0.540	-1	-1	-1	38101	0.306
needle	thread	0	0	1	0	0	0	0.035	-1	-1	-1	64176	0.934
neighborhood	school	-0	-0	-0	-0	-0	-0	-2.558	-2	-0	-0	8377	0.378
nights	weekends	0	0	1	0	0	0	1.377	1	0	0	17624	0.941
nitrogen	oxygen	0	0	-1	0	0	0	-0.649	0	0	-0	93142	0.419
north	south	0	0	1	0	0	0	-0.031	0	0	0	1532336	0.939

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
north	west	0	0	1	0	0	0	0.140	0	0	0	385427	0.728
nose	throat	0	1	0	0	0	0	0.246	0	0	0	101711	0.921
novelist	poet	-0	-0	-1	-0	-0	-0	-1.917	-1	-1	0	43417	0.343
novels	stories	0	0	1	0	0	0	-1.342	0	0	-0	70893	0.495
novels	tales	-1	0	0	0	1	0	-0.122	-1	-1	-1	16797	0.773
order	truth	0	0	0	0	0	0	1.080	-1	-1	-1	3498	0.387
organizations	states	1	0	0	-1	0	0	-1.897	-4	-1	-1	4310	0.378
pad	pencil	0	0	-1	0	0	0	-0.546	1	1	1	35363	0.800
pail	shovel	0	0	-1	0	0	0	-0.445	1	1	1	2855	0.822
pain	pressure	-0	-0	-0	-0	1	-0	-0.363	1	1	1	10781	0.479
pain	shame	-0	-0	-0	-0	1	-0	1.628	-0	-0	-0	14173	0.665
palm	pine	0	0	0	0	0	0	-0.176	0	0	0	5388	0.597
paper	pen	-0	-0	-1	-0	-0	-0	1.991	-1	-1	-1	86400	0.229
paris	washington	0	1	0	0	0	0	-0.473	1	1	-0	7309	0.530
pattern	sequence	-0	-0	-1	-0	-0	-0	0.475	-0	-0	0	2372	0.544
peace	quiet	0	0	1	1	0	0	0.805	1	1	1	191274	0.951
pepper	salt	-0	-0	-1	-0	-0	-0	-1.263	-1	-1	-1	831039	0.061
personnel	ships	-0	1	-1	-0	-0	-0	0.043	-2	-0	-0	1386	0.167
piano	violin	-0	-0	1	-0	-0	-0	1.104	-1	-1	-1	90587	0.224
pies	puddings	-0	-0	-0	-0	-0	-0	1.401	1	1	1	4565	0.583
pilot	plane	0	1	1	0	0	0	-1.039	-1	-1	-1	3214	0.322
pitch	volume	-0	-0	-0	-1	-0	-0	-1.797	1	1	1	10617	0.665
place	time	-0	1	-0	-0	-0	-0	-1.106	-0	-0	-0	1218090	0.139
plans	projects	0	-1	0	0	0	0	0.450	1	1	1	18234	0.817
poetry	prose	-0	-0	-1	-0	-0	-0	1.229	-2	-2	-1	222228	0.547
politics	religion	-0	-0	-0	-0	-0	-0	-0.148	-1	1	1	162465	0.352
power	resources	0	-1	0	1	0	0	1.408	1	1	-0	45867	0.698

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
power	scope	-0	-0	-0	1	-0	-0	2.524	-1	-1	-1	14744	0.382
power	trade	-0	-0	-0	-0	-0	-1	1.002	-1	-1	-1	3824	0.552
power	variety	-0	-0	-0	1	1	-0	1.572	1	1	0	2907	0.556
powers	processes	-0	-0	-0	1	-0	-0	-0.073	1	1	0	2242	0.769
pressure	temperature	-0	-0	-1	-0	-0	-0	0.238	2	2	0	449437	0.362
prose	verse	0	0	0	0	0	0	-0.658	0	0	0	173062	0.657
radar	radio	-0	-0	-1	-0	-0	-1	-1.970	1	-0	-0	12445	0.318
radio	television	0	0	-1	0	0	0	0.125	1	1	1	651263	0.744
receivers	senders	-0	-1	-0	-0	-0	-1	2.275	-0	-0	0	9936	0.038
research	training	0	0	0	0	0	1	0.670	0	0	-0	108917	0.644
resources	riches	1	-0	-0	-0	-0	-0	2.594	-1	-1	0	1403	0.308
rights	wrongs	0	1	0	0	0	0	3.705	0	0	0	47755	0.970
river	sky	-0	1	-0	-0	-0	-0	1.144	-1	-1	-1	3362	0.652
rock	roll	0	0	0	0	0	0	0.852	0	0	0	248992	0.997
rudder	stick	-0	-0	-0	-0	-0	-0	-2.538	-1	-1	-1	3699	0.100
sailors	soldiers	0	0	1	0	0	0	-1.825	0	0	-0	149549	0.116
school	service	-0	-0	-0	-1	-0	-0	0.376	1	1	1	1380	0.451
schools	synagogues	0	0	1	0	0	0	4.248	2	0	0	3291	0.408
science	tactics	1	0	0	0	0	0	2.531	0	0	-0	8541	0.992
searching	yearning	-0	1	-0	-0	-0	-1	1.619	-0	-0	0	1193	0.277
secrecy	stealth	0	-1	0	0	0	0	1.578	-2	-2	-1	1319	0.385
ships	weapons	0	0	1	0	0	0	0.283	1	1	1	2628	0.775
shirt	waistcoat	0	0	1	0	0	0	2.621	1	1	1	2852	0.483
shoes	socks	0	0	1	0	0	0	1.776	0	0	0	64752	0.762
shooting	yelling	-0	-0	-1	-0	1	-0	1.369	-0	-0	0	3658	0.394
shots	shouts	0	-1	0	0	1	0	0.808	0	0	0	1897	0.542
shower	tub	0	0	1	0	0	0	0.597	-1	-1	-1	17557	0.179

Table D.1: The corpus described in Chapter 3. (Continued)

WordA	WordB	Form	Percept	Culture	Power	Intense	Icon	Freq	Len	Lapse	*BStress	OverallFreq	RelFreq
sidewalks	streets	-0	-0	-1	-0	-0	-0	-2.920	-1	-0	-0	29202	0.200
sights	sounds	0	0	1	0	0	0	-1.933	0	0	0	143294	0.899
silence	space	-0	-0	-0	-0	-0	-0	-1.161	-1	-1	-1	2998	0.311
skirt	sweater	-0	-1	-0	-0	-0	-0	0.833	1	1	1	9969	0.512
snow	water	-0	1	-1	-0	-0	-0	-2.270	1	1	1	7559	0.537
soil	topography	-1	1	1	-0	-0	-0	2.862	2	2	1	4558	0.453
soldiers	subjects	-1	0	0	1	0	0	-0.250	0	0	-0	1315	0.533
son	wife	-0	-0	-0	-0	-0	-0	0.131	-0	-0	-0	118681	0.052
south	west	0	1	0	0	0	0	0.171	0	0	0	455538	0.649
spring	summer	0	0	0	0	0	1	-0.095	1	1	1	381701	0.990
statesmen	warriors	-0	-0	-0	-0	-1	-0	-0.781	1	1	0	5237	0.439
stress	tension	0	0	0	0	0	0	0.739	1	1	1	32225	0.620
stucco	wood	0	-1	-1	0	0	0	-3.813	-1	-1	-1	2813	0.346
sun	weather	0	0	1	0	0	0	0.899	1	1	1	3714	0.889
sunrise	sunset	-0	-0	-1	-0	-0	1	-0.728	-0	-0	-0	66190	0.794
sympathy	understanding	0	1	0	0	0	0	-1.613	1	-1	-0	66140	0.684
teaching	writing	0	0	1	0	0	0	-0.481	0	0	-0	37699	0.615
technology	time	-0	1	-0	1	-0	-0	-2.845	-2	-2	-1	3171	0.179
tensions	uncertainties	0	-1	0	0	1	0	0.779	1	1	-0	1587	0.743
testament	will	-0	-0	-0	-0	-0	-0	-4.261	-2	-2	-1	248768	0.001
thursday	tuesday	-0	-0	-0	-0	-0	-1	-0.033	-0	-0	-0	27948	0.005
tories	whigs	-0	-0	-0	-0	-0	-0	-0.463	-1	-1	-1	24672	0.155
transportation	utilities	0	0	0	0	0	0	1.402	-1	1	-0	8691	0.491
trees	underbrush	0	1	1	0	0	0	4.412	2	0	0	9369	0.876
turns	twists	-0	-0	-0	-0	-1	-0	3.029	-0	-0	-0	126594	0.067
vinegar	water	0	0	-1	0	1	0	-4.254	-1	-1	-0	20603	0.698
wisdom	wit	-0	-0	-0	-0	-1	-0	0.974	-1	-1	-1	43515	0.094

Bibliography

Agresti, A. (2002). *Categorical Data Analysis*. Wiley-Interscience.

Alegre, M. and Gordon, P. C. (1999). Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language*, 40:41–61.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Psychology Press, Hillsdale, NJ.

Arnon, I. and Cohen Priva, U. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech*, 56(3):349–371.

Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.

Baayen, H., Schreuder, R., de Jong, N., and Krott, A. (2002). Dutch Inflection: The Rules that Prove the Exception. In *Storage and Computation in the Language Faculty*, pages 61–92. Springer Netherlands, Dordrecht.

Baayen, R. H. (2011). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 1.4.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1):94–117.

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3):438–481.

Balota, D. A., Yap, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39:445–459.

Barnbrook, G. (2007). Sinclair on collocation. *IJCL*, 12(2):183–199.

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

Benor, S. and Levy, R. (2006). The Chicken or the Egg? A Probabilistic Analysis of English Binomials. *Language*, 82(2):233–278.

Bergen, L., Levy, R., and Gibson, E. (2012). Verb omission errors: Evidence of rational processing of noisy language inputs. *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*, pages 1320–1325.

Bertram, R., Hyönä, J., and Laine, M. (2000). The role of context in morphological processing: Evidence from Finnish. *Language and Cognitive Processes*, 15(4-5):367–388.

Bickerton, D. (1981). *Roots of language*. Karoma, Ann Arbor, MI.

Bock, J. K. and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1):47–67.

Bod, R. (1998). *Beyond Grammar*. An Experience-Based Theory of Language. Center for the Study of Language and Information, Stanford University.

Bod, R. (2008). The Data-Oriented Parsing approach: Theory and application. *Studies in Computational Intelligence (SCI)*, 115:307–348.

Bod, R., Scha, R., and Sima'an, K., editors (2003). *Data-Oriented Parsing*. University of Chicago Press.

Bolinger, D. L. (1962). Binomials and pitch accent. *Lingua*, 11:34–44.

Brants, T. and Franz, A. (2006). *Web 1T 5-gram Version 1*. Linguistic Data Consortium, Philadelphia.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007). Predicting the dative alternation. *Cognitive Foundations of Interpretation*, pages 69–94.

Burani, C., Salmaso, D., and Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language*, 18(4):348–358.

Burns, R. (1785). To a mouse. http://www.poetryfoundation.org/poem/173072.

Bybee, J. (1985). *Morphology*. A Study of the Relation between Meaning and Form. John Benjamins Publishing.

Bybee, J. (1999). Usage-based Phonology. In Darnell, M., Moravcsik, E. A., Noonan, M., Newmeyer, F. J., and Wheatley, K., editors, *Functionalism and Formalism in Linguistics: General papers*, pages 211–242. John Benjamins, Amsterdam.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4):711–733.

Bybee, J. (2009). *Language, Usage and Cognition.* Cambridge University Press, Cambridge.

Chomsky, N. (1957). *Syntactic Structures.* de Gruyter, Berlin, Boston.

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

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* Harper & Row, New York.

Chomsky, N. and Lasnik, H. (1993). Principles and parameters theory. In Jacobs, J., von Stechow, A., and Venneman, T., editors, *Syntax An International handbook of contemporary research.* de Gruyter, Berlin.

Cooper, W. E. and Ross, J. R. (1975). World Order. In Grossman, R. E., San, L. J., and Vance, T. J., editors, *Papers from the Parasession on Functionalism*, pages 63–111. Chicago Linguistics Society, Chicago.

Demberg, V. (2010). *A Broad-Coverage Model of Prediction in Human Sentence Processing.* PhD thesis, The University of Edinburgh.

Dixon, R. M. W. (1982). *Where have All the Adjectives Gone?: And Other Essays in Semantics and Syntax.* Walter de Gruyter.

Elman, J. L. (2003). Generalization from Sparse Input. *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society.*

Enochson, K. and Culbertson, J. (2015). Collecting Psycholinguistic Response Time Data Using Amazon Mechanical Turk. *PLoS ONE*, 10(3):e0116946–17.

Erman, B. and Warren, B. (2000). The idiom principle and the open choice principle. *TEXT*, 20(1):29–62.

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., and Bolhuis, J. J. (2015). Structures, Not Strings: Linguistics as Part of the Cognitive Sciences. *TRENDs in Cognitive Sciences.*

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics*, 27(3):517–556.

Ferdinand, V., Kirby, S., and Smith, K. (2014). Regularization in language evolution. *Proceedings of the 10th International Evolution of Language Conference.*

Ferreira, F. and Clifton, Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3):348–368.

Fine, A. B., Jaeger, T. F., Farmer, T. A., and Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8(10):e77661.

Frank, M. C. and Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, 75:80–96.

Gahl, S. and Yu, A. C. L. (2006). Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review*, 23(3):213–216.

Gaissmaier, W. and Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, 109(3):416–422.

Garnsey, S., Pearlmutter, N., Myers, E., and Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37:58–93.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. CRC Press.

Gibbs, R. W. (1990). Psycholinguistic studies on the conceptual basis of idiomaticity. *Cognitive Linguistics*, 1(4):417–451.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *TRENDs in Cognitive Sciences*, 7(5):219–224.

Goldwater, S., Griffiths, T. L., and Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, 112(1):21–54.

Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, 5(1):173–184.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35:151–166.

Griffiths, T. L. and Kalish, M. L. (2005). A Bayesian view of language evolution by iterated learning. *Proceedings of the 27th annual conference of the cognitive science society*, pages 827–832.

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models. *Journal of Statistical Software*, 33(2):1–22.

Hare, M., Elman, J. L., and Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language Acquisition*, 10(6):601–630.

Harrell, F. E. (2001). *Regression Modeling Strategies*. Springer Series in Statistics. Springer New York, New York, NY.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298(5598):1569–1579.

Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., and Snider, N. (2011). The source ambiguity problem. *Language and Cognitive Processes*.

Holsinger, E. (2013). Representing Idioms: Syntactic and Contextual Effects on Idiom Processing. *Language and Speech*, 56(3):373–394.

Hudson Kam, C. L. and Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, 1(2):151–195.

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

Inhoff, A. W. and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6):431–439.

Jackendoff, R. (2002). What's in the lexicon. In Nooteboom, S., Weerman, F., and Wijnen, F., editors, *Storage and Computation in the Language Faculty*. Springer Netherlands, Dordrecht.

Jackendoff, R. and Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, 97(2):211–225.

Jaeger, T. F. (2006). Phonological Optimization and Syntactic Variation: The Case of Optional *that*. *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society BLS*.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4):434–446.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics*, 50:101–113.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4):485–499.

Johnson, M., Griffiths, T. L., and Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models. *Advances in Neural Information Processing Systems*, 19:641.

Keller, F., Gunasekharan, S., Mayo, N., and Corley, M. (2009). Timing accuracy of web experiments: A case study using the WebExp software package. *Behavior Research Methods*, 41(1):1–12.

Kelly, M., Bock, J. K., and Keil, F. C. (1986). Prototypicality in a Linguistic Context. *Journal of Memory and Language*, 25:59–74.

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245.

Koehler, D. J. and James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, 113(1):123–127.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, volume 1 of *Theoretical Prerequisites*. Stanford University Press, Stanford.

Lee, M.-W. and Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition*, 105(2):446–456.

Levy, R. and Andrew, G. (2006). Tregex and Tsurgeon. *5th International Conference on Language Resources and Evaluation (LREC)*.

Levy, R., Fedorenko, E., Breen, M., and Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, 122(1):12–36.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., and Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163):713–716.

Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., and Petrov, S. (2012). Syntactic Annotations for the Google Books Ngram Corpus. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 169–174.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua*, 8:113–160.

Marcus, M., Santorini, B., Marcinkiewicz, M. A., and Taylor, A. (1999). Treebank-3. Linguistic Data Consortium.

McDonald, J., Bock, K., and Kelly, M. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25:188–230.

Menard, S. (2002). *Applied Logistic Regression Analysis*. SAGE Publications, Incorporated.

Mollin, S. (2012). Revisiting binomial order in English. *English Language and Linguistics*, 16(1):81–103.

Morgan, E. and Levy, R. (2015a). Abstract knowledge versus direct experience in processing of binomial expressions. *Manuscript submitted for publication.*

Morgan, E. and Levy, R. (2015b). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. *Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society*, pages 1649–1654.

Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In DeGraff, M., editor, *Language Creation and Language Change*. MIT Press, Cambridge, MA.

Nooteboom, S., Weerman, F., and Wijnen, F., editors (2002). *Storage and Computation in the Language Faculty*, volume 30 of *Studies in Theoretical Psycholinguistics*. Springer Netherlands, Dordrecht.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language*, 70(3):491–538.

O'Donnell, T., Snedeker, J., Tenenbaum, J. B., and Goodman, N. (2011). Productivity and Reuse in Language. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 1613–1618.

Onishi, K. H., Murphy, G. L., and Bock, K. (2008). Prototypicality in sentence production. *Cognitive Psychology*, 56(2):103–141.

Pierrehumbert, J. (2000). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. and Hopper, P., editors, *Frequency and the Emergence of Linguistic Structure*, pages 137–157. John Benjamins, Amsterdam.

Pine, J. M. and Lieven, E. V. M. (2008). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(02):123.

Pinker, S. (1991). Rules of Language. *Science, New Series*, 253(5019):530–535.

Pinker, S. (2000). *Words and Rules*. The Ingredients of Language. Harper Perennial, New York.

Pinker, S. and Birdsong, D. (1979). Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior*, 18(4):497–508.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Post, M. and Gildea, D. (2013). Bayesian Tree Substitution Grammars as a Usage-based Approach. *Language and Speech*, 56(3):291–308.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria.

Rayner, K. and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201.

Rayner, K., Sereno, S. C., and Raney, G. E. (1996). Eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5):1188–1200.

Rayner, K. and Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4):504–509.

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

Rumelhart, D. E. and McClelland, J. L. (1986). On Learning the Past Tenses of English Verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP research group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA.

Sereno, J. A. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, 25(4):425–437.

Shanks, D. R., Tunney, R. J., and McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15(3):233–250.

Sinclair, J. (1991). *Corpus, Concordance and Collocation*. Oxford University Press, Oxford.

Singh, R., Fedorenko, E., Mahowald, K., and Gibson, E. (2015). Accommodating presuppositions is inappropriate in implausible contexts. Technical report.

Siyanova-Chanturia, A., Conklin, K., and van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3):776–784.

Smith, K. (2009). Iterated learning in populations of Bayesian agents. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 697–702.

Sobkowiak, W. (1993). Unmarked-before-marked as a Freezing Principle. *Language and Speech*, 36(4):393–414.

Steinbeck, J. (1937). *Of Mice and Men*. Covici Friede.

Swinney, D. A. and Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18(5):523–534.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7(4):263–272.

The Carnegie Mellon Speech Group (2014). The CMU Pronouncing Dictionary. Carnegie Mellon University.

Tily, H. (2012). flexspr. *Personal communication.*

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *TRENDs in Cognitive Sciences*, 4(4):156–163.

Truswell, R. (2009). Attributive Adjectives and Nominal Templates. *Linguistic Inquiry*, 40(3):525–533.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2(10):717–726.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., and Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, 93(2):185–238.

van den Bosch, A. and Daelemans, W. (2013). Implicit Schemata and Categories in Memory-based Language Processing. *Language and Speech*, 56(3):309–328.

Vulkan, N. (2000). An Economist's Perspective on Probability Matching. *Journal of Economic Surveys*, 14(1):101–118.

Wiechmann, D., Kerz, E., Snider, N., and Jaeger, T. (2013). Introduction to the Special Issue: Parsimony and Redundancy in Models of Language. *Language and Speech*, 56(3):257–264.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press, Cambridge.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*, volume 93. OUP Oxford.

Wright, S. K., Hay, J., and Bent, T. (2005). Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics*, 43(3):531–561.

Xu, F. and Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2):245–272.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wesley Press, Oxford.