Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora

David Reitter

School of Informatics, University of Edinburgh

Doctor of Philosophy
Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh
2008

Abstract

This thesis addresses the cognitive basis of syntactic adaptation, which biases speakers to repeat their own syntactic constructions and those of their conversational partners. I address two types of syntactic adaptation: short-term priming and long-term adaptation.

I develop two metrics for syntactic adaptation within a speaker and between speakers in dialogue: one for short-term priming effects that decay quickly, and one for long-term adaptation over the course of a dialogue. Both methods estimate adaptation in large datasets consisting of transcribed human-human dialogue annotated with syntactic information. Two such corpora in English are used: Switchboard, a collection of spontaneous phone conversation, and HCRC Map Task, a set of task-oriented dialogues in which participants describe routes on a map to one another. I find both priming and long-term adaptation in both corpora, confirming well-known experimental results (e.g., Bock, 1986b). I extend prior work by showing that syntactic priming effects not only apply to selected syntactic constructions that are alternative realizations of the same semantics, but still hold when a broad variety of syntactic phrase structure rules are considered. Each rule represents a cognitive decision during syntactic processing. I show that the priming effect for a rule is inversely proportional to its frequency.

With this methodology, I test predictions of the Interactive Alignment Model (IAM, Pickering and Garrod, 2004). The IAM claims that linguistic and situation-model agreement between interlocutors in dialogue is the result of a cascade of resource-free, mechanistic priming effects on various linguistic levels. I examine task-oriented dialogue in Map Task, which provides a measure of task success through the deviance of the communicated routes on the maps. I find that long-term syntactic adaptation predicts communicative success, and it does so earlier than lexical adaptation. The result is applied in a machine-learning based model that estimates task success based on the dialogue, capturing 14 percent of the variance in Map Task. Short-term syntactic priming differs qualitatively from long-term adaptation, as it does not predict task success, providing evidence against learning as a single cognitive basis of adaptation effects.

I obtain further evidence for the correlation between semantic activity and syntactic priming through a comparison of the Map Task and Switchboard corpora, showing that short-term priming is stronger in task-oriented dialogue than in spon-

taneous conversation. This difference is evident for priming between and within speakers, which suggests that priming is a mechanistic rather than strategic effect.

I turn to an investigation of the level at which syntactic priming influences language production. I establish that the effect applies to structural syntactic decisions as opposed to all surface sequences of lexical categories. To do so, I identify pairs of part-of-speech categories which consistently cross constituent boundaries defined by the phrase structure analyses of the sentences. I show that such *distituents* are less sensitive to priming than pairs occurring within constituents. Thus, syntactic priming is sensitive to syntactic structure.

The notion of constituent structure differs among syntactic models. Combinatory Categorial Grammar (CCG, Steedman, 2000) formalizes flexible constituent structure, accounting a varying degree of incrementality in syntactic sentence planning. I examine whether priming effects can support the predictions of CCG using the Switchboard corpus, which has been annotated with CCG syntax. I confirm the syntactic priming effect for lexical and non-lexical CCG categories, which encode partially satisfied subcategorization frames. I then show that both incremental and normal-form constituent structures exhibit priming, arguing for language production accounts that support flexible incrementality.

The empirical results are reflected in a cognitive model of syntactic realization in language production. The model assumes that language production is subject to the same principles and constraints as any other form of cognition and follows the ACT-R framework (Anderson et al., 2004). Its syntactic process implements my empirical results on priming and is based on CCG. Syntactic planning can take place incrementally and non-incrementally. The model is able to generate simple sentences that vary syntactically, similar to the materials used in the experimental priming literature.

Syntactic adaptation emerges due to a preferential and sped-up memory retrieval of syntactic categories describing linearization and subcategorization requirements. Long-term adaptation is explained as a form of learning, while short-term priming is the result of a combination of learning and spreading activation from semantic and lexical material. Simulations show that the model produces the adaptation effects and their inverse frequency interaction, as well as cumulativity of long-term adaptation.

Acknowledgements

Thanks

My first thanks go to Frank Keller and Johanna D. Moore. I could not have had more inspiring, reliable, demanding and generous advisors. Over the course of three years, they provided not only feedback, encouragement, contacts and friendship, but most importantly intellectual direction. The School of Informatics and its Institute for Communicating and Collaborative Systems enabled me to pursue the questions I deemed most interesting and the methods I found most promising.

Committee members Dan Jurafsky and Mark Steedman made the viva a stimulating experience. The discussions with Martin J. Pickering, Jon Oberlander, and Fernanda Ferreira were helpful in defining and refining the work. I am also grateful to Julia Hockenmaier and Roger Levy for extensive discussions and collaboration, as well as to Sasha Calhoun and the Paraphrase project for providing essential time alignment data in the Switchboard corpus. The following people provided valuable feedback or pieces of information at some point during the process: R. Harald Baayen, Franklin Chang, Florian Jaeger, Jonas Kuhn, Richard L. Lewis, Geoffrey K. Pullum, Christoph Scheepers, Patrick Sturt, and Benedikt Szmrecsanyi. I would like to thank Bruce Duncan for copy-editing. Errors and omissions remain my own.

The work described in this thesis would not have been possible without software that is often developed selflessly by teams of programmers. I would like to thank the makers of: R with the packages MASS and lme4, ACT-R, Python, Aquamacs Emacs and GNU, BibDesk, Omnigraffle, LaTeX, and Mac OS X.

I acknowledge support through the Edinburgh-Stanford Link. There were many without whom my time at the University of Edinburgh would not have been as enjoyable: Sherice N. Clarke, the lunch crews (Myrosia Dzikovska, Elaine Farrow, Mark McConville, Diane Litman, Theresa Wilson, Josh Schroeder and friends), the members of the Scotch Malt Whiskey Society (Charles Callaway, Christopher Geib, Matthew Stone) and my friends at Edinburgh University Gliding Club.

Collaborations and Publications

The experiments presented in Chapter 3 were published in Reitter et al. (2006b,d); Reitter and Moore (2007a). Most of the experiments presented in Chapter 4 have in published in Reitter et al. (2006a); Reitter and Keller (2007). The thesis has also benefited from the comments of anonymous reviewers and the audiences of ACL-07, EMNLP-06, CogSci-06/07, HLT-NAACL-06, CUNY-07/08 and AMLaP-06, which improved form and content. I am indebted to Julia Hockenmaier for providing the normal-form and incremental CCG variants of the Switchboard corpus.

The model parameters and effect reliability estimates given in the thesis will occasionally differ in magnitude but not qualitatively from previously published results, as the methodology has since been revised.

Warning

This thesis makes extensive use of statistical and cognitive modeling. Models allow us to explicitly state our assumptions and to test them empirically. As with Newton's description of an apple falling from the tree, even influential models will be superseded by better ones. It also helps to keep Shakespeare in mind: *The fool doth think he is wise, but the wise man knows himself to be a fool.* Models not only clarify what we do know, they also bring to light what we do not know. Yet.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

David Reitter

Table of Contents

Chapter 1

Introduction

1.1 Priming influences the form of our sentences

"Form Follows Function"—this principle of modern architecture could apply to natural languages as well. *Function* results from the meaning to be conveyed: when we speak, we convert the conceptual representation of a message into a sequence of sounds. *Form* is not just a consequence of the message in linguistic expression. It is influenced by the cognitive system that allows us to produce language: basic principles influence the form of what we say. If form followed function in natural language, humans would compose every utterance in a deterministic manner, reproducing the same sentence given a meaning. Instead, we learn and contextualize our linguistic output, adapting to the close and distant past. This thesis is about the variation of form in language production, independently of the immediate meaning conveyed. We examine how and why the structure of sentences depends on the language previously produced and comprehended.

The task of *language production* is often analyzed in terms of a processing chain which includes conceptualization, formulation, and articulation (Levelt, 1989). The conceptualization module selects concepts to express, and the formulation module decides how to express them. Formulation involves determining the lexical, syntactic, and semantic representation of the utterance. Syntax determines the systematic relationship between meaning and form of an utterance, without which language could not be produced. Variation in syntax is what we are concerned with here. If we made conscious decisions about the structure of our sentences, typical deliberations for a speaker would include the questions "Should a clause be formulated as passive, or should I formulate it as active? Should I *give the man a book* (*double object*), or should I *give a book to the man* (*prepositional object*)? Are the children *dropped off* at the swimming pool, or do we *drop them off*?"

Experimental results (e.g., Bock 1986b) show that participants that have a choice between producing the double object (DO) and the prepositional object (PO) construction (e.g., in a picture naming task) are more likely to choose the construction that they (or their interlocutor) have produced previously, and similarly for the use of passives. The general conclusion is that syntactic choices are sensitive to *syntactic priming*: any decision for a particular structure renders following decisions for the same or a related structure more likely.[1]

For how long this effect lasts is subject to debate. In some studies, the effect disappeared after just a clause or a sentence (Levelt and Kelter, 1982; Branigan et al., 1999)—we call this adaptation effect *short-term priming*. Others find that priming persists (Bock and Griffin, 2000; Branigan et al., 2000b)—we call this adaptation effect *long-term adaptation*. Such a duality begs the question about the cognitive substrate of syntactic priming. Is there really only one effect, or are we, in fact, seeing two?

To determine the cognitive basis of syntactic priming, we need to isolate the precise point at which priming affects the language production process. Repetition effects like priming are interesting because repetition above chance levels indicates processing units: the very patterns that are used to produce and comprehend natural language. If the repetition of linguistic material is influenced by its context, then those structures supply evidence for the units of linguistic processing. Our basic hypothesis is that syntactic priming is due to low-level, general effects that affect not just language, but any aspect of cognition. To reduce priming effects to their cognitive bases, we first need to develop an idea of the elementary steps in the decision-making process governing syntactic structure. We demonstrate empirically that priming applies to those steps. In a cognitive model, we then show how it emerges from the general "Rules of the Mind" (cf. Anderson, 1993).

Most work on syntactic priming has been carried out with a psychological perspective in mind. A common view of the cognitive architecture treats syntactic de-

[1]We refer to syntactic priming simply as *priming* when the context does not leave room for ambiguity. *Structural priming* is often used in the literature to indicate effects that include priming of syntactic structure; in this thesis we focus on syntactic structures exclusively and use the term *structural priming* synonymously. Nevertheless, priming can apply to hierarchical *structure* or just sequences. It is considered syntactic in both cases.

cisions as high-level choices that occur whenever there is a chance to decide about form, given meaning. As explained above, we would expect effects such as priming to apply to the general syntactic process, that is, to a broad variety of syntactic structures. It is for this reason that we break up syntactic constructions using the computational tools that linguistic theories provide: grammars with phrase-structure rules and, later, combinatorial restrictions associated with lexical material. Such linguistic grammars encode the allowable set of linguistic results, i.e., they specify the acceptable *form* used to express a given meaning. Grammars afford us with a means to model exactly where priming occurs. Using examples such as the ones with *give* or *drop off*, we argue that priming can take place on the level of syntactic rules. Only if priming holds for syntactic choices in general can we assume it to be a general effect with an underlying cognitive basis, rather than a learned convention.

Given that priming concerns linguistic decision-making *in context*, it seems surprising that the classical studies were carried out in a lab setting with an experimental design that controlled the context tightly. Commonly, subjects are first asked to produce a prime sentence, during which they are forced to choose a particular syntactic form. In a second step, subjects are asked to describe a picture or otherwise formulate a sentence that, semantically, lends itself to the kind of syntactic variability (e.g., DO / PO) that is being examined. Thus, the experiments not only focus on very specific syntactic choices, they also use linguistic data that are elicited using artificial constraints and a high rate of repetition—after a number of trials, subjects may not be as naïve as they are taken to be. This raises the question whether priming can also occur in naturally occurring discourse. Early observational studies found this to be true for conversational actions (Schenkein, 1980)—linguistic repetition is a common feature of conversation (Tannen, 1989). A series of studies has found syntactic adaptation effects in corpus data (Estival, 1985; Gries, 2005; Szmrecsanyi, 2005; Dubey et al., 2005; Jaeger, 2006), each of them looking at selected syntactic constructions. In this thesis, we draw data from corpora in order to demonstrate that the interactions of priming effects affect all syntactic choices. We focus on why that is the case.

1.2 Questions, hypotheses and contributions

The underlying question we shall ask is simple. *What is the cognitive basis of syntactic priming?* The quest for answers leads us to two different levels of discovery. First, we want to examine whether syntactic priming is simply a form of long-term adaptation (*learning*), or if there is a dedicated contextualization mechanism that pre-activates representations as soon as they are accessed. Priming effects can decay rapidly, but they can also last over minutes or even days (see Section 1.3.3). Learning could explain this, but then we would also expect short-term priming to be an indicator of future learning. In other words, any short-term priming and long-term learning effects should co-vary. We probe this in the context of an analysis of the function of priming in dialogue. Priming has been hypothesized to support speakers in their mutual alignment of language and semantic understanding of the dialogue context (Pickering and Garrod, 2004). We examine whether priming differs among different types of dialogue, and whether short-term, long-term, or both types of adaptation can predict the communicative success of interlocutors. Determining the role of syntactic priming in dialogue may help us understand why priming has emerged as a useful, if not strategic effect, but it also allows us to draw qualitative distinctions between short-term priming and long-term adaptation effects.

Second, we examine the basis of priming in a syntactic context. *Is syntactic priming structural in nature?* This line of inquiry concerns the units of cognitive processing that syntactic priming applies to. We not only show that priming can be modeled using common theories of syntactic processing; we use syntactic priming effects to argue in favor of more theoretical, syntactic assumptions, such as core statements of categorial grammar and the *flexible incrementality hypothesis*. The research program that we would like to identify is based on the insight that *repetition of linguistic material is indicative of structure*: repeatable structures are evidence for the units of linguistic cognition.

Any explanation of how humans learn, comprehend or produce language must provide answers to the following questions. What are the data structures or *processing units* used to store and process linguistic information, be they innate or learned? In which order is information accessed and how is it combined to form sentences? What facilitates or inhibits access to information?

Empirically, we want to observe conversation that is unconstrained by experimental setup. This corroborates existing, classical experiments on syntactic priming, which were done in an experimental setting, and corpus-based experiments done on written text. We examine language in the context of dialogue. Wherever possible, we want to determine priming effects as they apply to general syntactic choices rather than selected syntactic alternations.

Using this methodology, we examine syntactic priming both in the context of its possible functions in dialogue, but also with respect to its locus in the language production process. Here, we progress from the examination of the *outcome* of the language production process to the process itself. The observed priming effects are exploited to pinpoint structural properties of syntax.

The empirical portion of this thesis yields results relevant to language production models, in particular for syntactic realization. The model conceptualizes these and other results in an implemented, testable and extensible form. In particular, the contributions of this thesis are:

1. We introduce two methods based on linear regression to measure syntactic priming and long-term adaptation in corpora. With these models and two English-language dialogue corpora, we generalize the known syntactic adaptation effects to phrase-structure rules (Experiments 1 and 2). Further instantiations of this methodology yield a number of results.

   (a) Short-term priming is greater in task-oriented dialogue than in spontaneous conversation (Experiments 3 and 5). Semantic activity may increase syntactic priming. The effect is due to a basic cognitive property and not merely a learned strategy, as priming is increased both within and between speakers.

   (b) Short-term priming is greater for less frequent rules (inverse frequency interaction; Experiment 5).

   (c) Short-term priming and long-term adaptation differ qualitatively. Long-term adaptation, but not short-term priming, correlate with task success. Hence, there may be several sources of priming (Experiments 6 and 7).

   (d) We confirm a prediction of the Interactive Alignment Model (Pickering and Garrod, 2004), stating that task success is correlated with syntactic priming in task-oriented dialogue (Experiment 7).

(e) Syntactic priming is *structural*: it is sensitive to constituent structure (Experiments 9 and 10).

(f) We lend support to *flexible incrementality* in language production in an experiment that uses syntactic priming and a corpus annotated with Combinatory Categorial Grammar (Steedman, 2000) (Experiment 12). We show that the statistical priming model supports both incremental and non-incremental (planning based) language production.

(g) Short-term priming can be modeled as a lexical access effect that applies to combinatorial, syntactic categories as defined by lexicalized, categorial grammar formalisms. We confirm a prediction of Categorial Grammar by showing that priming can be statistically modeled as an effect that applies to types encoding open subcategorization frames (Experiment 13).

2. We define two tasks that serve to evaluate methods to predict task success. We show that a combination of repetition features can predict task success in task-oriented dialogue in a machine-learning approach (Experiment 8).

3. We present a model of language production situated in a general cognitive architecture (Chapter 5). The model explains a number of syntactic adaptation effects. We show that short-term priming and long-term adaptation and their interactions emerge from two basic learning properties. The first one is base-level learning, which leads to long-term adaptation and short-term priming. The second one is associative learning, contributing to short-term priming and lexical boost effects. The model accounts for the results (1a)–(1g).

1.3 Background: Priming and language processing

When we speak, we repeat ourselves for many reasons. There is lexical repetition: the repetition of words and of their meanings, for instance, because we focus on one or a small number of topics at a time. Some of this repetition is to be expected, but the actual repetition found in experiments and in naturally occurring discourse is greater than that (see Experiments 2 and 1). This indicates *priming* effects. In the following, we shall shed light on the reasons for this increased repetition.

Priming is a common phenomenon affecting all levels of language production and comprehension. A prototypical example of the effect in comprehension is that a word (the target) is recognized more quickly and more accurately if it is semantically similar to a preceding word (the prime). An example of priming in language production would be that subjects prefer a specific synonym over another one to express a given meaning for a few seconds or even half an hour after the synonym was used initially. The corpus data that this thesis is based on reflect production priming; either from comprehension to production (CP), or from production to production (PP). Syntactic priming is a strong clue that syntactic language production does not occur in a sentence-by-sentence fashion, with each sentence being syntactically independent of the previous one. Existing models of discourse coherence do not account fully for the influence that context exerts on linguistic choice.

Repetition occurs on higher levels, too: there is repetition of whole phrases, which may happen for rhetorical reasons, or due to disfluencies. Finally, there is also repetition in structural choices, as we have introduced earlier. This thesis is about the increase in such repetition above chance level: the effect is referred to as *syntactic priming*. We focus on dialogue. Here speakers are not only sensitive to priming from their own speech. They also accept priming from their interlocutors. Following Pickering and Garrod (2004), we say: they *align* their linguistic representation.

Over the past decade, a number of priming phenomena have been explored experimentally. In the following sections, we describe empirical work on syntactic priming.

1.3.1 Structural priming

As said, we focus on priming that applies to *syntactic* decisions. Such syntactic priming effects have been demonstrated for syntactic constructions in language production and comprehension.

Levelt and Kelter's (1982) study can be seen as one of the first investigations into priming in dialogue. They described how replies are syntactically related to questions: Shopkeepers tended to reply to the question *At what time does your shop close?* (In the Dutch original: *Om hoe laat gaat uw winkel dicht?*) with a sentence that repeated the preposition, e.g., *At five o'clock.* (*Om viif uur.*). If the question did not contain a preposition (*What time does your shop close? / Hoe laat gaat uw winkel* *dicht?*), their replies tended not contain the preposition (*Five o'clock. / Viif uur*).

A much-cited experiment by Bock (1986b) showed priming effects that were clearly structural in nature. In her experiments, subjects were asked to repeat prime sentences, and then to describe semantically unrelated pictures, which served as targets. Primes consisted of sentences with ditransitive verbs, whose dative argument could either be realized as a prepositional object (PO) or in a double object (DO) construction, for instance, *A rock climber sold some cocaine to an undercover agent*, vs. *A rock climber sold an undercover agent some cocaine*. In the targets, subjects were more likely to use a DO construction after a DO prime, and a PO construction after a PO prime.

In general, experimental studies on structural priming have used a small number of selected well-known alternations of English, which are assumed to be synonymous:

- double (DO) vs. prepositional objects (PO): *the man gives the woman the flower* (double) vs. *the man gives the flower to the woman* (prepositional) (Bock, 1986b; Branigan et al., 2000a)
- participle placement: *the man switches off the light* (post-verbal) vs. *the man switches the light off* (sentence-final) (Konopka and Bock, 2005)
- active vs. passive voice: *the prince told an anecdote* (active) vs. *an anecdote was told by the prince* (passive) (Weiner and Labov, 1983; Bock, 1986b)
- the structure of noun phrases with modifiers: *the red sheep* (adjectival) vs. *the sheep that's red* (relative) (Cleland and Pickering, 2003)
- the omission of optional *that* complementizers in English (V. Ferreira, 2003; Jaeger, 2006)
- high vs. low relative clause attachment in German: *Gabi bestaunte das Titelbild*$_{masc}$ *der Illustrierten*$_{fem}$, *das*$_{masc}$ / *die*$_{fem}$ ... (*Gabi admired the cover*$_{masc}$ *of the magazine*$_{fem}$, *which*$_{masc}$ / *which*$_{fem}$ ...) (Scheepers, 2003)

A common experimental design elicits a *prime* by constraining the subjects in some way. Branigan et al. (2000a), for instance, use a booklet that must be completed by participants. First, a scene to be described verbally is depicted (e.g., a man giving a flower to a woman), along with a fill-in-the-blank sentence, such as *the man gives the flower* ___. Subjects can only complete this sentence with a prepositional object. After a filler sentence that does not contain a choice between DO and PO complements, subjects are asked to describe a different scene. This time (target), they are unconstrained in their choice of construction.

Priming has been shown to increase the probability of one of these syntactic forms appearing by 12% on average—that is, with the above forms and the limited set of lexical contexts that were used in the experiments. Such alternations have also been used in corpus studies (see below).

Syntactic priming effects have mostly been demonstrated in carefully controlled psycholinguistic experiments, thus raising the question of whether priming also occurs in natural, spontaneous conversation. Recent work addressed this question. Estival (1985) found priming effects of actives and passives in a corpus. Gries (2005) uses an English-language corpus to show not only syntactic priming effects, but also that verbs differ in their sensitivity to priming. Szmrecsanyi (2005) presents a study demonstrating the long-term persistence of various alternations in a dialogue corpus. Dubey et al. (2005) argue that syntactic parallelism in coordinate constructions is best explained by priming effects. Jaeger's (2006) study finds significant priming influence on the use of an optional *that* complementizer.

Rather than selecting particular syntactic constructions, we examine the repeated use of phrase structure rules that license a particular syntactic form. Constructions such as *passive voice* or a certain particle placement translate to particular sets of syntactic rules. Indeed, corpus-based studies found an increased repetition of selected syntactic rules (e.g., Dubey et al., 2005; Jaeger, 2006). But just which rules tend to be repeated? Are there patterns? We hypothesize syntactic priming to be a result of more general cognitive phenomena affecting the syntactic process (cf., Bock and Loebell, 1990; Hartsuiker et al., 1999; Pickering and Branigan, 1998; Pickering et al., 2002; Cleland and Pickering, 2003, and others). Syntactic priming does not primarily arise from surface-level or directly from semantic effects, but it applies to representations of syntactic structure.

Priming indeed affects *structural* properties rather than just word sequences (Experiment 3 in Bock and Loebell, 1990). However, other experiments in the same study also show that surface structure is also sensitive to priming, i.e., sentences that differ in the thematic roles of their verb complements may still show syntactic priming. In Bock and Loebell's (1990) study, prepositional phrases with a locative *by* led to the increased use of any prepositional datives. That is, the sentence *The wealthy woman drove the Mercedes to the church* (locative) boosted the production of prepositional datives in *The wealthy woman gave the Mercedes to the church* (goal dative). Similarly, *The 747 was landing by the control tower* primed *The 747 was alerted by the control tower*. This suggests that the syntactic compositions of prime and target share some common structures. The structural view of priming, using rules or combinatorial categories, accounts for these data. We present studies testing the predictions of such a structural analysis in Chapter 4.

In this thesis we refer to priming of syntactic processes as *syntactic priming*. We show, using corpora, that this priming is sensitive to *structure*, thus, the expression *structural priming* refers to the same effect. The hypothesis of a non-structural, syntactic priming effect is examined in Chapter 4. (Note that in the context of psycholinguistic experiments, authors sometimes use the term *structural* to refer to further, non-syntactic, levels of linguistic decision-making. We do not consider non-syntactic effects and do not adopt this terminology.)

1.3.2 Syntactic priming as indicator of structure

Syntactic priming can help us determine the components of a language processing architecture. The basic assumption is that priming applies to processing units and can thus indicate their boundaries. Interactions of the priming effect must follow interactions of such processing units, for instance because the retrieval of one unit facilitates or necessitates the retrieval of another.

In particular, syntactic priming is a valuable tool to examine structure. Syntactic structure eludes direct examination (we can only observe the product), but the crucial assumption is that if, for example, a passive voice construction as a prime facilitates the comprehension or production of another passive voice construction as a target, then prime and target must have structural commonalities. If a passive voice construction also facilitates another construction, such as one involving a locative *by* or a prepositional phrase, then these materials will share some syntactic processing units. Much work in syntax has assumed that syntactic decisions operate on *structured* representations, dividing sentences into a hierarchy of constituents. In contrast, connectionist models (e.g., Elman, 1990) and computational natural language processing models (e.g., Brown et al., 1992) suggest that such hierarchical structure, if any, emerges from low-level word-to-word transitions. If that was true, syntactic priming would apply to basic word-to-word transitions. That is, it would apply to word *sequences* rather than *hierarchical syntactic decisions*. This thesis will formalize the two models of priming and evaluate which one is supported by the language data.

Priming can also point to segmentation in another sense. The amalgamation of language-specific syntactic processing units and information traditionally seen as *lexical* has resulted in lexicalist theories, which posit that combinatorial knowledge (syntax) is stored alongside knowledge about words. This implies, for instance, that the lexical properties of the finite verb (the clausal head) will determine the shape of the clause. Lexicalization can also relate to the encoding of language-specific parameters in what is usually called *Principles and Parameters* approaches. There, language-specific information is stored separately from an innate syntactic model. For instance, the fact that subject noun phrases can be dropped in Italian would be stored as a generalization and applicable to all subjects. In an extreme lexicalized formalism, such information would be stored along with all verbs, implying that verbs (or full verb forms) differ in their preference to drop their subject. However, even lexicalist theories generally assume a hierarchy of lexical types, so that generalized information is not replicated many times. Seen from a cognitive perspective, this seemingly academic discussion concretely condenses to the question of memory retrieval: is syntax retrieved along with lexical knowledge? Syntactic priming can help to differentiate accounts and specify a processing model.

Melinger and Dobel (2005) demonstrate that even a single verb as a prime can activate syntactic information, serving as syntactic prime for a subsequently elicited target phrase. Lexical boost effects serve to show that syntax and lexicon are at least closely linked (see Section 1.3.6, and also Experiment 14), and they are modeled as such in the production model presented in this thesis (Chapter 5).

We rely on a lexicalized account of syntax. This does not mean that our empirical work is concerned with lexical repetition. On the contrary, all Experiments other than 8 and 14 are concerned with syntactic repetition. Among the motivating reasons for this is that we can assume that speakers are unaware of their syntactic choices, while they may reflect on their choice of words. A clustering of topics, which is natural in any coherent discourse, also causes strong local lexical repetition effects, which would be difficult to distinguish with the corpus-based methods proposed here.

The focus of this thesis is *language production*. We show priming as it affects the production process, and we suggest a production process in the form of a cognitive model. Priming itself can be caused by produced language, but we also show how comprehension causes priming in production. This effect points to shared structures in production and comprehension, and the model we develop suggests that lexical-syntactic information stored in memory is shared. Thus, the empirical priming results and the language production model have consequences for language production in general.

1.3.3 Long-term vs. short-term priming

In Levelt and Kelter's (1982) priming study, the repetition bias was remarkably short-lived: the effect disappeared after one clause. In addition, in a later study involving written sentence production, structural priming ceased to be detectable when just one sentence intervened between prime and target (Branigan et al., 1999).

Other studies contrast strongly with this. Hartsuiker and Kolk (1998) found no decay of priming when a one-second temporal lag was inserted between prime and target. Bock and Griffin (2000) demonstrate a form of structural priming that persists with two and even ten intervening sentences. These results were corroborated by Branigan et al. (2000b), who found that priming in spoken (as opposed to written) production persists, whether there is a temporal lag or intervening linguistic material that delays the elicitation of the target. At this point it is not entirely clear what causes priming to be transient in some experiments and long-lived in others (cf., V. Ferreira, 2006). Hartsuiker et al. (2008) finds that decay in verb phrase structure priming is related to the lexical repetition of the verb itself: if the verb is repeated, as in Bock and Griffin's (2000) study, syntactic priming is very short-lived. If it is not repeated, as in Branigan et al. (1999), priming will last longer. (We call the long lasting repetition bias *long-term adaptation*.)

In this thesis, we treat short-term and long-term repetition biases as separate effects initially, with different metrics to measure them. Some of our results present evidence as to why short- and long-term adaptation differ in their nature (compare Experiments 6 and 7) rather than the result of a single learning process. Ultimately, we present a combination of two different cognitive bases (semantic/lexical and syntactic) that account for syntactic adaptation and the lexical boost, as well as the qualitative differences we find between short- and long-term adaptation.

1.3.4 Comprehension vs. production priming

This thesis focuses on syntactic priming (henceforth: priming) that affects the language production process, commonly called *production priming*.

Experimentally, production priming has been elicited early and often. Priming in comprehension (priming that affects the comprehension process) has also been found. Potter and Lombardi (1998) showed that the mere comprehension of a related construction can aid production.

We assume that production and comprehension processes share linguistic information; syntactic priming from comprehension to production lends credence to this assumption. In a dialogue context, comprehension-production priming has been demonstrated by Branigan et al. (2000a) and Cleland and Pickering (2003). Bock et al. (2007) repeated an earlier study (Bock and Griffin, 2000), but presented the same primes auditorily. They obtained the same structural persistence, i.e. over the same prime-target distances and at similar magnitude. Notably, comprehension-production priming is synonymous with priming between speakers in the context of our dialogue studies.

1.3.5 Structural properties of priming: the case of linearization

A major difference between the alternations approach and the one using phrase structure rules is that the latter always combines immediate dominance (in a syntactic description that assumes phrase structure trees) and linear precedence (linearization). It assumes that priming affects these potentially different structural features simultaneously.

This is compatible with Pickering et al.'s (2002) study, which shows that there is no explicit linearization phase. These experiments addressed the question of whether NP-shifted constructions prime their non-shifted counterparts. For instance, *The racing driver showed to the helpful mechanic the problem with the car* as a prime (shifted) did not increase the subject's tendency to produce *The racing driver showed the extremely dirty and badly torn overall to the mechanic* (non-shifted). While the two variants differ in their surface form, they are similar on a deeper syntactic level: their hierarchical structures are the same. A multi-stage account, which predicts the dominance relations (i.e. the structural hierarchy) to be constructed separately, would have predicted priming effects between the two sentences. However, the experiments could not confirm such a priming effect for language production in writing. Pickering et al. (2002) conclude that "constituent structure is formulated in one stage". Salamoura and Williams (2007) present similar results in Greek, where shifted NPs are common.

Earlier experiments by Hartsuiker et al. (1999) have pointed to the existence of such a separate linearization stage. There, a very similar alternation with equivalent functional structure, but different linear ordering, is tested for priming: *Op de tafel ligt een bal* ("On the table is a ball.") vs. *Een bal ligt op de tafel* ("A ball is on the table."). Subjects show that the variants prime one another, which would be consistent with an approach where dominance (hierarchy) is determined separately, before linearization takes place (see also Hartsuiker and Westenberg, 2000). Their results are re-examined by Pickering et al. (2002), who assume a *functional* level of representation. Constituent structure is constructed from this representation. Priming takes place during the process.

The assumption of a combined structural construction stage does not imply that linearization cannot be primed. The experiments discussed in Hartsuiker et al. (1999) actually point to cross-modal priming of linearization decisions. They found a strong effect of the position of pictograms on word order in a description task. When a drawing representing a noun was shown in the upper left part of the picture shown to the participants, it was likely to be used first (i.e., as subject) in the resulting sentence. Also, priming seems to have a stronger effect on target structures within the first phrase of an utterance, but not in later phrases (Smith and Wheeldon, 2001). However, such results may have more to do with the order of realization and differences in preactivation needed at each stage.

Incremental production would be sensitive to linearization priming. V. Ferreira (1996) explicitly supports incrementality in a serial form, as opposed to an approach where several competing linearization variants are maintained in parallel.

The model suggested in this thesis combines decisions about syntactic dominance and linear order into either phrase-structure rules, categorial types, or part-of-speech bigrams, and finally (Chapter 5), we propose an algorithm using categorial types that proceeds incrementally and deterministically, deciding about hierarchy and linearization in one step.

1.3.6 Boost effects

Structural priming is affected by repetition on other levels (see Figure 1.1 for an overview). Pickering and Branigan (1998) demonstrate that structural priming effects are stronger when there has been successful priming on the lexical level, that is, when the verb is repeated. This effect is called the *lexical boost*.

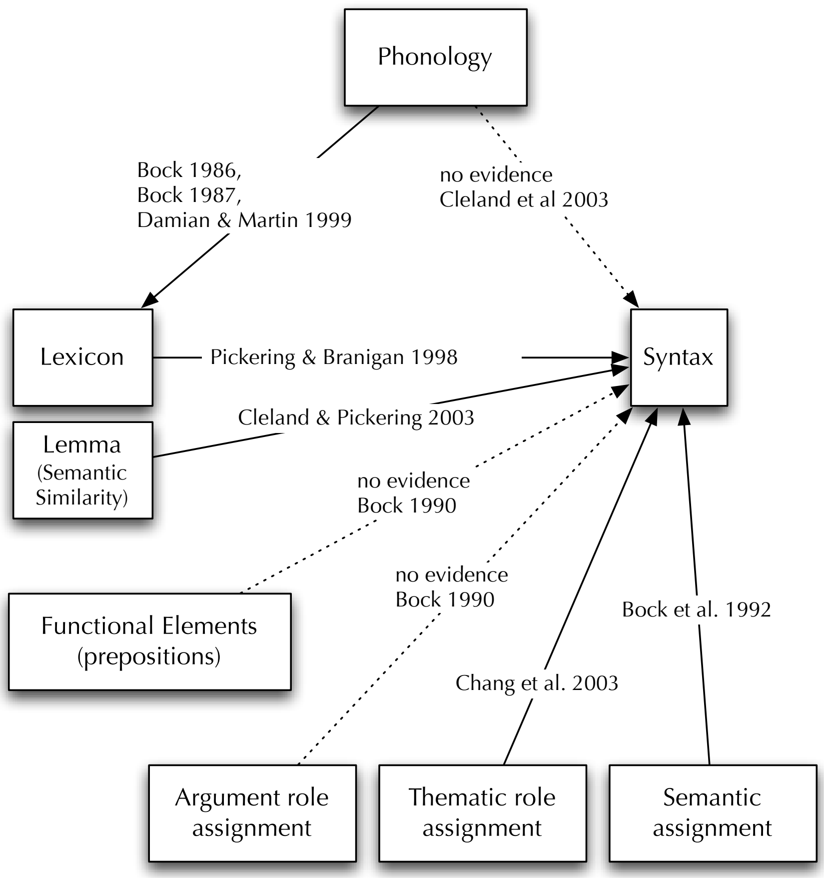


Figure 1.1: Evidence for boost effects on one level by repetition on another

The question arises whether boosting is an additive interaction of priming and lexical-syntactic preferences. Different verbs attract different syntactic configurations, and the higher likelihood of syntactic priming in repeated-verb situations may simply be an additive effect of such lexical-syntactic attraction. Evidence from different pairs of linguistic levels suggests otherwise. Cleland and Pickering (2003) found significantly enhanced structural priming effects for the alternation of a relative phrase (*the goat that's red*) vs. adjectival phrases (*the red goat*) whenever prime and target nouns were semantically similar such as *goat* and *sheep* (priming effects: 47% for same nouns, 31% for similar nouns and 8% for unrelated nouns.)

The same study did not find evidence for an influence of phonological relatedness on structural priming effects. Earlier experiments by Bock (1986a) revealed small boosting effects of phonological relatedness on lexical priming. Pickering and Branigan (1998) could not demonstrate an influence of morphosyntactic features (tense and number) in written language production. However, their conclusion that morphosyntactic features are represented separately does not necessarily follow, given that their correlations merely failed to reach significance. An alternative hypothesis is that the effect is too small given the sample size.

Such boosting appears to be limited in its direction. For instance, no boosting could be found from a discourse-semantic level to the syntactic form. In experiments carried out by Bock and Loebell (1990), subjects showed a tendency to repeat syntactic structures, whether there were changes in the argument structure or not (*The wealthy widow gave her Mercedes to the church* vs. *The wealthy widow drove her Mercedes to the church.*).

The language production model discussed in Chapter 5 provides an explanation for lexical boost effects rooted in well-understood, general properties of the cognitive apparatus.

Lexical boost is relevant to the Interactive Alignment Model (Pickering and Garrod, 2004), a model of dialogue in which priming effects receive a boost from priming-induced repetition on other levels. Thus, priming at the lexical or syntactic levels can support alignment at the levels of semantics and, crucially, the interlocutors' common interpretation of the situation (*situation model*). This idea provides the central motivation for the work presented in Chapter 3.

1.3.7 Practical applications of syntactic priming

In addition to shedding light on cognitive questions, understanding priming in dialogue will be very useful in practical applications.

Alignment can be reproduced by a Natural Language Generation System. For instance, the system can introduce a lexical and to some extent structural bias in a language model which is used to determine the output (Brockmann et al., 2005). That said, psycholinguistically motivated approaches to alignment in natural language generation are rare.

Implementing linguistic alignment in a spoken dialogue system requires a flexible natural language generation system which maintains information about the user's speech input. User input needs to be parsed (analyzed syntactically), and that parsing and generation take place on the same syntactic platform.

We evaluate the link between priming and task success in Sections 3.7ff., and subsequently build a machine-learning based model to estimate success levels (Section 3.10). As we argue, both lexical and structural priming can be used to detect alignment levels in dialogue systems and, ultimately, predict the success of interlocutors at completing a given task that requires alignment. We will see that lexical and syntactic repetition and other length features account for about 15% of the overall variance of success in human-human dialogue (Experiment 8).

1.4 Overview

This thesis is structured as follows. Having given an overview of the issues at hand in Chapter 1, we proceed to describe the methodology used to measure priming levels in corpora of transcribed and syntactically annotated speech in Chapter 2.

Chapter 3 is concerned with the function of priming in dialogue. We test claims that priming aids interlocutors in establishing a common model of their object of discourse.

In Chapter 4, we then turn to the processing levels on which syntactic priming applies. We present experiments that use syntactic priming effects to determine the units of syntactic processing and investigate predictions arising from a flexible treatment of constituent structure. This informs an investigation of the cognitive bases of syntactic priming effects. We show that basic learning and spreading-activation effects can account for long-term and short-term repetition effects. In Chapter 5, we present a cognitive model of syntactic realization in speech production to evaluate this claim. We conclude in Chapter 6 with an overview of the contributions of this thesis.

Chapter 2

Measuring Priming and Adaptation

In this chapter, we describe the methodology to examine two spoken-language corpora with respect to structural repetition. The Switchboard (Marcus et al., 1994) and HCRC Map Task (Anderson et al., 1991) corpora both contain transcriptions of spoken dialogue and phrase structure-based syntactic tree annotation.

2.1 Measuring priming

2.1.1 Corpus studies as opposed to experiments

Experimental studies have uncovered structural priming using selected syntactic constructions. But do experiments in psycholinguistics create a natural, fully spontaneous situation? Not necessarily: it has been shown that findings regarding verb-argument preferences in experimental conditions do not correlate well with corpus studies (Roland and Jurafsky, 2002). Gries (2005) argues while experimenters can control potentially influential factors much better in designed experiments as opposed to corpus analyses, variationist work and the history of confirmed and disconfirmed experiments in structural priming research points to a variety of factors in linguistic choice, which are hard if not impossible to control experimentally. New corpus-based studies (Gries, 2005; Szmrecsanyi, 2006, 2005; Dubey et al., 2005) address such criticism, showing structural priming effects that differ in strength for different lexical items.

Such studies pick out a small set of syntactic rules or constructions such as active vs. passive voice or double object vs. prepositional object use for arguments to verbs e.g., *give*: *give your friend the book* vs. *give the book to your friend*. From a computational perspective, this approach leaves one question open: does structural priming and alignment affect only certain syntactic structures?

While there is a quite substantial effect for the alternations tested in experiments, how large is the effect when considering all syntactic configurations? We would expect syntactic priming to occur with a range of syntactic configurations. This thesis is concerned with confirming and quantifying syntactic priming effects for the general case.

2.1.2 Priming effects only for controlled semantics?

All studies in structural priming that use dedicated experiments with human subjects rely on the control of semantics: for instance, in Branigan et al.'s (2000a) confederate scripting experiments, the naïve subject had to identify cards out of a deck by describing the pictures they showed. Prime and target consisted of a linguistic construction that was chosen from one of several *semantically equivalent* alternations.

Prior designs in structural priming experiments examine the use of alternate syntactic choices for the same semantics. One issue with this is the incomplete notion of semantics. In terms of "truth conditions", active and passive sentences are indeed equivalent. But when semantics include information structure, or the connotations that each passivized verb may carry, equivalence is a weak concept.

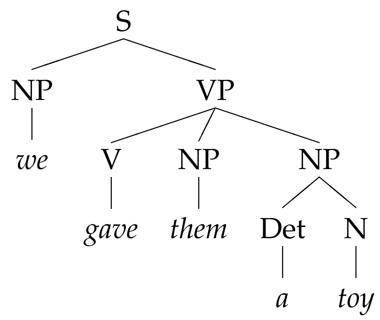
Turning to earlier priming studies, we find that the classical notion of *priming* does not imply a primed preference for one choice of alternative behavior over another one. For example, priming occurs when lexical access is sped up after a semantically related picture had been shown to the subject (Swinney, 1979).

So, semantic equivalence is not needed in order to define or measure "priming". It is a useful tool to determine syntactic choice points, where a speaker is known to decide about which construction they will use. When counting just occurrences of actives or passives over a range of verbs and semantics, we lose this distinction. Still, we know that some variation in the data is still due to the choices a speaker makes. Here, we need to contrast the effect from random variation. Taking the example of passive constructions again, we can compare the use of passives under two conditions, one of them a control condition. And this is exactly what each of the two methods proposed does: *priming is not repetition*. Priming is the differential between the probability of normal (chance) repetition and repetition probability in situations where a prime is present.

In corpus studies like the ones presented in the present thesis, the concept of *covariance* replaces the strict control of semantics in the experiment. Looking at large amounts of data, we see a high degree of variance in the subject's choices of syntactic forms, which is natural, as the underlying semantics largely dictate how to construct the sentences. Examining a large number of data points allows us to treat semantic variation as noise. As is usual, the statistical analysis then is to show that the variance associated with covariates (or: predictors) exceeds the variance we see due to the semantic and other goings-on in the dialogue.

2.1.3 Corpus processing

The trees were converted into phrase structure rules in order to list the rules that *license* the trees. For example, the (hypothetical) tree



would have been converted to three phrase structure rules:

(R1) S → NP VP,

(R2) VP → V NP NP and

(R3) NP → Det N.

Table 2.1 gives actual rule instances extracted from one of the corpora used. This conversion is unique.[1]

Given the phrase structure rules for each utterance, we can now identify the repeated use of rules. A certain amount of repetition will obviously be coincidental. But structural priming would predict that a rule (*target*) occurs more often closely after a potential *prime* of the same rule (stimulus) than further away. Therefore,

[1] Obviously, when dealing with speech, we encounter constructions that cannot be analyzed with a traditional phrase structure rules. The annotation of both corpora commonly assigns ad-hoc rules with flat derivations in such cases. This leads to a large set of extracted rules. Such rules are unlikely to be repeated. For the analysis of repetition, they represent no theoretical obstacle.

onset time (s)	speaker	syntactic rule	yield
185.105	f	VP → VBG PP	*keeping on the edge of the page*
185.363	f	PP → IN NP	*on the edge of the page*
185.490	f	NP → AT NN	*the edge*
185.490	f	NP → NP PP	*the edge of the page*
185.692	f	PP → IN NP	*of the page*
185.729	f	NP → AT NN	*the page*

Table 2.1: Syntactic rules and additional information extracted from the Map Task corpus. The speaker here is the direction follower (f), as opposed to the direction giver.

we can correlate the probability of repetition with the distance between prime and target.

As syntactic structure, we count each syntactic rule which licenses part of the syntactic analysis for a tree. For example, if a sentence-level conjunction leads to the rule $S \rightarrow S\ conj\ S$, and such a conjunction occurs in utterances 3 and 11, we would observe a repetition at distance $d = 8$. This way, every syntactic rule is counted as a potential prime and (almost always) as a target for priming. Because interlocutors tend to stick to a topic during a conversation for some time, we exclude cases of syntactic repetition that are solely due to word-by-word repetition of the rules' yields. Experiment 14 (p. 151 in Chapter 5) examines the relationship of lexical repetition and priming explicitly.

In Chapter 3, we motivate a modification to this methodology, expressing d in terms of time (seconds). This distance is measured from the onset of the prime rule's yield to the onset of the target rule's yield.

2.1.4 Generalized Linear Mixed Effects Regression

There are several ways to identify an effect of distance on repetition probability. One can normalize the number of observed repetitions by the number of expected repetitions for each syntactic rule by taking its prior probability of occurrence into account. The disadvantage of this is that for rare rules, we see a grossly higher error than for rules with higher frequency. Such a dataset would be difficult to model. Alternatively, one can examine the distribution of repetition counts over prime-target distances and use a sampling technique to balance the number of trials across distances. Thirdly, we can contrast cases of structural repetition and cases where no repetition occurs between two speech units that occurred a chosen distance apart. We adopt the latter technique.

In all cases, a rule instance *target* is counted as a repetition at distance d iff there is an utterance *prime* which contains the same rule, and *prime* and *target* are exactly d units apart. In the studies presented in this thesis, we use *Generalized Linear Mixed Effects Regression Models* (GLMM). GLMMs with a binary response variable can be considered a form of *logistic regression*.[2]

Regression allows us to fit a *model* to our data. A *linear model* is simply a choice of coefficients β_i, one for each explanatory variable i (and one for each of their interactions). β_i expresses the contribution of i to the probability of the outcome event, that is, in our case, successful priming. Our data is represented by extracted features—in our context, we call them factors (discrete) and predictors (continuous explanatory variables).

For example, the β_i estimates allow us to predict the decline of repetition probability with increasing distance between prime and target, or other variables such as corpus choice. If we see priming as a form of pre-activation of syntactic nodes, it indicates the decay rate of pre-activation. The scale for this coefficient is the logarithmic distance in number of utterances.[3]

To sum up, *Linear Regression Models* (LMs) can model the decay of the priming effect by estimating the relationship between d and the probability of rule repetition. The model is designed to predict whether repetition will occur, or, more precisely, whether there is a prime for a given target (priming). Under a no-priming null hypothesis, we would assume that the priming probability is independent of d. If there is priming, however, increasing d will negatively influence the priming probability (decay). So, we expect a model parameter (also termed *covariate*) DIST for d that is reliably negative, and lower, if there is more priming.

With this method, we draw multiple samples from the same utterance—for different d, but also for different syntactic rules occurring in those utterances. Because these samples are inter-dependent, we use a grouping variable indicating the

[2]The data are assumed to be binomially distributed. We do not generally give classical R^2 figures, as this metric is not appropriate to such GLMMs.

[3]In our analysis, we focus on the coefficients rather than on the intercept β_0 because long-term adaptation effects and the granularity of syntactic annotations show up in β_0. Both lie out of the scope of this study.

source utterance. Because the dataset is sparse with respect to PRIME, balanced sampling is needed to ensure an equal number of data points of priming and non-priming cases (PRIME) is included.

When a trained model is used to predict the actual outcome, the estimated parameters act as coefficients in a function like the following:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_{12} x_1 x_2 + \ldots$$

The model contains a β for each explanatory variable, an overall bias (intercept) β_0 and a coefficient for each interaction between explanatory variables. The model yields the best fit for y against the actual data x_i (for each data point). For all experiments discussed in the following, y will be a ratio of syntactic repetitions vs. trials, that is, vs. opportunities for a repetition to occur. Note that for binary explanatory variables, x_i is 0 or 1, and the estimate can be seen as a probability $[0,1]$.

Table 3.1 (p. 56) summarizes a GLMM along with further figures that allow us to estimate whether the coefficients obtained are reliable (statistically significant). Coefficients like β_{12} for example estimate an *interaction* between two explanatory variables. They give a coefficient for the influence of e.g., x_2 on the coefficient of x_1, as rewriting the above equation demonstrates:

$$y = \beta_0 + (\beta_1 + \beta_{12} x_2) x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots$$

It is important to keep in mind what the estimated coefficients in a model mean. Let's focus on just the β_1 coefficient. Its quantity measures the (linear) influence that the predictor x_1 has on the response y. That means that for each increase of x_1 by one unit, we expect to find y elevated by β_1 units. To interpret the case $\beta_1 < 0$, we would expect to find y decreased by $-\beta_1$ units. Coefficients as given in this thesis are only meaningful when taking the measure of the predictors x_n and the response y into account. To assess the relative contribution of each predictor, *normalized* β *coefficients* can be derived.

In a practical application, we could, for instance, use the linear model to estimate the relationship between the number of disfluencies (response y) in each sentence and the frequency of the head verb (x_1). (In practice, we would probably want to log-transform and normalize this frequency.)

Interactions with another variable (e.g., x_2), can have a sometimes surprising influence on the estimate for β_1. Again, the measure of x_2 will have an influence, so that it is often more transparent to compare the sum of the coefficients for x_1, i.e., $\beta_1 + \beta_{12}x_2$ for various values of x_2. Often, x_2 will be a binary factor, which is assigned the nominal levels $0,1$. This would be the case in an experiment with two conditions, in which we would like to measure the varying effect of x_1. In this case, we would compare $\beta_1 + \beta_2 * 0$ to $\beta_1 + \beta_2 * 1$. For example, we might compare native and non-native speakers in the above experiment. If we are explicitly interested in the effect of language proficiency (or L1/L2 acquisition) *on the frequency effect*, then we would model the acquisition type as x_2 and include the interaction in the model.

In more complex situations, we may add further interactions. Often, such interactions are included initially, but eliminated from the model if they do not show a significant effect (see Crawley 2005, pp. 105–). The sufficient (minimal adequate) model is reported in such cases.

Where many interactions are included in the final model (as in the experiments reported in Chapters 3 and 4), we *contrast* the sizes of the effect of interest for each combination of factors, independently of the effects of other continuous predictors (controlling for them). Usually, the effect size we are interested in is the one for ln(DIST), under combinations of conditions that are determined by the particular experiment.

In the experiments reported in this thesis, we usually estimate a response variable y that indicates *repetition*, that is, given a syntactic construction (e.g., syntactic rule or part-of-speech bigram), *did this construction occur before?* A certain amount of repetition is to be expected, but this chance repetition is independent of the distance between the two repeated constructions. In the short-term priming experiments, this distance is coded as the first effect variable, i.e., x_1. The estimated parameter β_1 indicates the development of repetition probability with increasing distance, and $\beta_1 < 0$ indicates a priming effect. We are usually interested in the interaction of this effect with various other factors and predictors, depending on the particular experiment.

For instance, if the model in Table 3.1 was a simple Linear Model with only the main effects ln(DIST), ln(FREQ) and $\text{SOURCE}_{MapTask}$ and two two-way interactions, it would specify a function predicting the probability of repetition for a prime-target pair i, $y = p(repetition_i)$.

$$p(\text{repetition}_i) = \beta_0 + (\beta_1 + \beta_{13}x_{i3})x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon$$

x_{i1} gives the value of $\ln(d)$ for the prime-target pair i (covariate: $\ln(\text{DIST})$). x_{i2} gives the frequency $\ln(f)$, and x_{i3} the level coding for the factor PRIMETYPE, which is coded as 0 for repetition within a speaker, and 1 for repetition between speakers (CP priming). Note that the model specifies further interactions and extends the model in other ways which we shall introduce. The β_i represent the model parameters.

Our models are generally trained with a binary response variable. This has an important consequence for the methodology: binary values and probabilities are generally not normally distributed. This is an issue for linear models, which are constrained to data with normally distributed responses. The variance of binary variables is regularly smaller for high and low y. A *logit* transformation can be applied to address this problem if y is a probability ($0 < y < 1$):

$$logit(y) = \ln(\frac{y}{1-y})$$

The result is a logistic regression model, an instance of Generalized Linear Models (GLMs). Consequently, the models do not predict probabilities, but logits. Here, we show the fixed-effects portion of the model given in Table 3.1:

$$\begin{aligned} logit(p(\text{repetition}_i)) &= \\ \ln(\frac{p(\text{repetition}_i)}{1-p(\text{repetition}_i)}) &= 0.584 + (-0.134 + 0.042xi3)x_{i1} + 0.831x_{i2} - 0.299x_{i3} + \varepsilon \end{aligned}$$

In this thesis, we usually test hypotheses using interactions with the main (decay/priming) variable ($\ln(\text{DIST})$ or β_1). The actual magnitude of the effects is of secondary concern in most experiments, but can be easily derived.

A further extension is to add further effects. We not only fit parameters for covariates describing *fixed effects* as discussed so far, but also add further *random* effects. In the analysis of experimental, repeated-measures data, these effects describe subject- or item-specific variation and are held constant for these groups (i.e. for each item or each subject). In the corpus-specific methodology, they group data from each target utterance.[4] These effects can be seen as utterance-specific error term ε in the above model function. Their magnitude for each utterance (or subject, or item) is usually not of interest. With the logit transform and the addition of random effects, we have arrived at *Generalized Mixed Effects Models*.

[4]Also note that the β parameters represent per-subject averages rather than population averages.

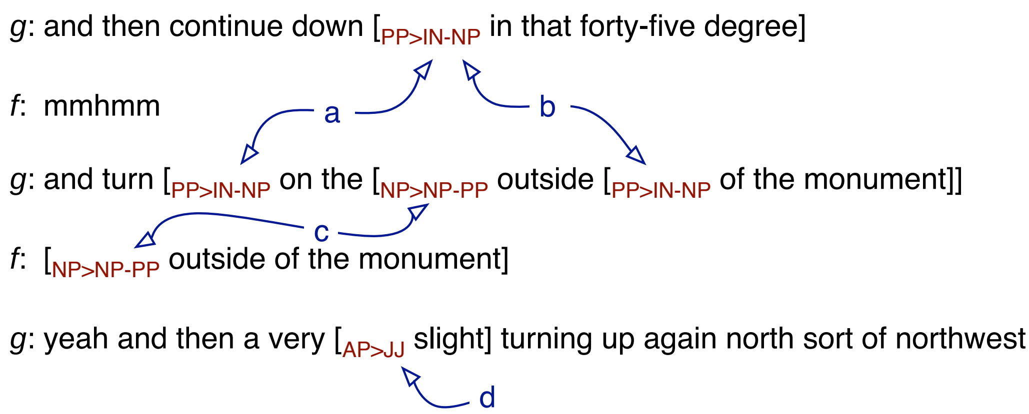


Figure 2.1: Two instances of syntactic repetitions (a,b), a lexical-syntactic one (c) and a preterminal rule (d) from Map Task.

The models for all the experiments reported in this thesis were fitted using Penalized Quasi-Likelihood (Venables and Ripley, 2002), with the exception of Experiments 3, 4, 8 and the tests in Section 2.5.4, for which we followed Baayen et al. (2008) and used an implementation of the same fitting method or the more precise Laplace method by Bates (2008).

2.1.5 Syntactic repetitions

Every pair of two equal syntactic rules up to a maximal distance is a potential case of priming-enhanced production. Consider the example shown in Figure 2.1, where a small subset of the rules that license constituents are marked. Two syntactic repetitions shown here are data points for our analysis. Repetitions *a* and *b* are both at distance 2, because the occurrences (prime and target) are two utterances apart. Repetition *c* would be included at distance 1, if the lexical content of prime and target differed. In *c*, however, we see a syntactic repetition that is due to lexical repetition. Repetitions of unary rules such as the one marked as *d* are not included. The third sentence lends the opportunity to include another repetition (of the prepositional phrase rule PP→ IN NP), but unlike Dubey et al. (2005), this study is not concerned with within-utterance repetitions.

The following analysis shows the distribution of repetition probability over distance from the repetition (target) to the prime. In our data, each repetition occurrence of a syntactic rule *R* at distance *d* counts as *priming*. Each case where *R* occurs, but isn't primed d units beforehand in the dialog, is counted as *non-priming*.

Our goal is to model $\hat{p}(prime|target,n)$, that is, the sampling probability that a *prime* is present in the n-th utterance before *target* occurs. Without structural priming in the general case, we would assume that

$$\hat{p}(prime|target,n) = \hat{p}(prime|target).$$

In order to eliminate cases of lexical repetition of a phrase, e.g., names or lexicalized noun phrases, which we consider topic-dependent or cases of lexical priming, we only collect syntactic repetitions with at least one differing word.

For instance (Figure 2.1), we would have two cases of priming for the rule PP→ IN-NP, namely at distance 2 (a,b), and two of non-priming at distance 1 (two occurrences of that rule and their non-occurrence in the previous utterance).

The distance between stimulus and target (DIST) is initially counted in utterances (Experiments 1–3), but later in seconds, which also includes within-utterance priming. Additive priming by a stimulus that is repeated several times is not captured by the model. We looked for repetitions within windows of 25 utterances or 15 seconds. So, each rule occurrence in the dialog can lead to up to 25 or 15 data points for the various distances. Memory effects generally decay non-linearly, and an exploratory analysis of the repetition probabilities as they develop with increasing d confirmed this non-linear decay. We therefore include a transformed distance, in our models $\ln(\text{DIST})$. Early, informal experiments showed improved fits of the transformed models.

From our analysis, we drop all hapax rules (frequency $f = 1$) as well as outliers, that is 15 highly frequent rules ($f > 2,000$, out of 759) in the case of Map Task, and accordingly 9 ($f > 12,000$, out of $4,695$) in the larger Switchboard corpus.

We include a random intercept in our model grouped by target utterance. This declares the several measurements (up to 25 for utterances or 15 for time) as *repeated measurements*, since they depend on the same target rule occurrence and are partially inter-dependent.

Again: without priming, one would expect that there are equally many cases of syntactic repetition, no matter the distance between first (*prime*) and second (*target*) occurrence. The analysis attempts to reject this null hypothesis and show an of the distance effect with the type of corpus used. We expect to see the structural priming effect found experimentally translate to more cases for shorter repetition distances, since priming effects usually decay rapidly (Branigan et al., 1999). (cf. Figure 2.2, which illustrates the decay.)

We distinguish *comprehension-production* (CP) priming, where the speaker first comprehends the prime (uttered by his/her interlocutor) and then produces the target, and *production-production* (PP) priming, where both the prime and the target are produced by the same speaker. This distinction is encoded in the factor PRIMETYPE.

A predictor ln(DIST) is included to express the logarithm of the normalized frequency of the repeated syntactic rule in the corpus. Frequency is an important covariate in many psycholinguistic models. It has long been suspected to interact with priming (e.g., Scheepers, 2003).

In summary, our modeling effort tries to establish a priming effect. To do so, we can make use of the fact that the priming effect decays over time. How strong that decay is gives us an indication of how much repetition probability we see shortly after the stimulus (prime) compared to the probability of chance repetition—without ever explicitly calculating such a prior.

Thus we define the strength of priming as the decay rate of repetition probability, from shortly after the prime to 15 seconds afterward (predictor: DIST). Thus, we take several samples at varying distances (d), looking at cases of structural repetition, and cases where structure has not been repeated.

Related methods have been used to show the effect of distance for repetition magnitudes. Gries (2005) shows a significant correlation of distance with repetition, but also demonstrates that at distances greater than one parsing unit (which usually coincides with an utterance), distance has no measurable effect. This is compatible with our findings, where we see a strong decay for $4-5$ seconds. Unlike Gries, we use DIST as the measure of priming for short-term priming and examine its interactions.

2.1.6 Analogies to experimental designs

The predominant statistical method to analyse repeated-measures experiment data is *Analysis of Variance* (ANOVA). The well-accepted and standardized method is applied across a range of problems, and this is the case despite it being a *parametric* method assuming normally distributed data when the data at hand are decidedly non-normal.

More specifically, the response (outcome) variable used in the model is assumed to be normally distributed. Where this assumption is violated, the quality of an ANOVA analysis degrades gracefully: this is, for instance, the case for the analysis of reaction times. However, where categorical (e.g., binary) responses or counts are used, ANOVAs either lack power or are overly optimistic. Arcsin transforms, which are often used, do not yield satisfactory performance, especially at the margins of the probability space (i.e., high or low probabilities).

Generalized linear models offer an alternative that generalizes the ANOVA approach (an ANOVA is just an instance of linear models). While one of their advantages is the more complex structure of dependent variables, the core argument pertains to the transformation applied. To analyze binary response variables, a logit-link transformation is applied, which transforms the probability obtained by analyzing the data points for each factor combination into log-odds space. (See Agresti (2002); Baayen et al. (2008) for an overview.)

ANOVAs for repeated measures are usually reported using two analysis variants. F_1 gives an F measure analyzing the data *by subject*, treating each participant of a study as a single data point. Such an analysis takes into account that measurements repeated for each participant are not independent by aggregating them over subjects. The result is a model that allows a generalization beyond the particular sample of subjects to the population that they were sampled from. Similarly, F_2 ANOVAs aggregate each (repeated) item, generalizing to other items.

An analogous approach in GLMMs is the use of random effects, which can be grouped by subjects or items. Notably, nested groupings such as "items within subjects" or several random effects (once grouped by subjects, one by items) can be used to fit models that generalize across subjects and items at the same time. More generally, specifying grouping variables allows us to fit GLMMs to repeated measures data.

In the corpus models of short-term priming presented in this thesis, we group data points by *target utterances*, implying that a group of data points stemming from the same target utterance is the result of a repeated measure, i.e. the data points have not been randomly sampled. Note that this does not imply a *by item* analysis, which would be inadequate for observational data, where none of the utterances was intentionally repeated, as is done in controlled experiments. Similarly, corpora usually offer only limited repetition of subjects. Dialogue corpora involve dyads of subjects. Future work may address issues related to this, for instance using nested or multiple error terms. In this thesis, we concentrate on *by utterance* analyses, which appear to be a conservative choice.

Despite the simple measure of grouped random effects, it remains a caveat of corpus-based methods that dependencies between samples from a corpus may lead to underestimated errors, i.e. inflated significance estimates for the level of generalization intended. Samples from a corpus are never random (Kilgarriff, 2005). The conservative interpretation of this is that the results generalize across the particular utterances in the corpora, but not necessarily beyond the subjects and the corpora chosen, i.e. they can be seen as case studies. The less conservative approach, however, is to point out that corpus studies usually achieve relatively high confidence, i.e. low significance levels, and that the error from non-independent sampling is negligible.

2.1.7 Sampling techniques

Taken together, positive and negative samples amount to a very large dataset. Regression analysis has, in our experiments, proved to be computationally intractable with such a dataset. One method to address this issue is to conduct the experiments on random samples of the corpora (as done in Reitter et al., 2006b).

As an alternative, contingency tables could be used instead of the dataset that contains a binary response variable. Contingency tables keep the size of the dataset manageable, while inheriting many of the advantages of binary logistic regression. Working with counts of repetition is more feasible in our case than would be in binary logistic regression models (Szmrecsanyi, 2005; Gries, 2005), which take all instances of positive and negative cases (repetition vs. non-repetition) into account, and which yield a manageable amount of data only when cases of selected syntactic alternations are extracted.

The third alternative is *balanced sampling*. This method is suitable especially in situations where logistic (or multinomial) regression is performed on datasets with rare events, such as the repetition data: while there are many instances of non-repetition, there are only a few (less than 2 percent) cases of repetition. Thus, positive examples are sparse. In this case, we perform regression on a sample that contains all of the sparse events (i.e. cases of syntactic repetition), and a random sample of the more frequent events as control. This results in an intercept parameter near 0.5 for most models, as this is the baseline probability of a repetition occurring in the sample. Since the intercept is of no further relevance, balanced sampling can be performed. This way, model fitting becomes tractable using all examples of repetition. A side-effect is that the penalized quasi-likelihood fitting algorithm is now terminates reliably (cf., Venables and Ripley, 2002).

To demonstrate that both simple sampling and balanced sampling yield comparable results, we perform the first set of experiments (comparing the two corpora) using the simple sampling method, and further experiments (correlating task success with adaptation) with the balanced sampling method.

In two first experiments, we show that the method replicates syntactic priming effects using two corpora.

2.2 Experiment 1: Repetition in spontaneous conversation

2.2.1 Method

The method used to measure priming effects has been described in Section 2.1.

The dataset used in this experiment is *Switchboard* (Marcus et al., 1994), a corpus of spontaneous spoken telephone dialogue among randomly paired, North American speakers who were given a general topic, but otherwise remained unrestricted.

The conversations were transcribed, and $80,000$ utterances were annotated with phrase structure trees by Marcus, Kim, Marcinkiewicz, MacIntyre, Bies, Ferguson, Katz, and Schasberger (1994). This portion, included in the Penn Treebank, has been time-aligned (per word) in the Paraphrase project (Carletta et al., 2004).

$1293,000$ repetitions could be found in $472,000$ extracted phrase structure rules, of which $4,700$ rules are distinct. These data were balanced by re-sampling, yielding all examples of repetition and a sample of non-repetition cases.

The fitted model contained the $\ln(\textsc{Dist})$ covariate to estimate priming levels (negative effects indicate stronger priming), $\ln(\textsc{Freq})$ for the effects of frequency, and a factor PRIMETYPE (CP for comprehension-production priming between speakers, PP for production-production priming within a speaker).

Covariate	β	*SE*	$p(>\|z\|)$
Intercept	0.024	0.011	< 0.05 *
ln(DIST)	−0.111	0.005	< 0.0001 ***
ln(FREQ)	0.793	0.01	< 0.0001 ***
PRIMETYPE$_{CP}$	−0.109	0.013	< 0.0001 ***
ln(DIST):ln(FREQ)	0.042	0.005	< 0.0001 ***
ln(DIST):PRIMETYPE$_{CP}$	−0.037	0.006	< 0.0001 ***
ln(FREQ):PRIMETYPE$_{CP}$	−0.057	0.012	< 0.0001 ***
ln(DIST):ln(FREQ):PRIMETYPE$_{CP}$	0.043	0.006	< 0.0001 ***

Table 2.2: The model of rule repetition in Switchboard. Prime-target distance in utterances. As is standard, *SE* indicates standard error.

2.2.2 Results

The model shows a reliable effect of ln(DIST) ($\beta = -0.111, p < 0.0001$): repetition of a rule becomes less likely as the distance from the first occurrence increases. PRIMETYPE interacts with the decay coefficient for ln(DIST) ($\beta = -0.037$, $p < 0.0001$). The resulting contrast of this interaction is that the parameter for ln(DIST) in our model is -0.111 as above in PP priming, but -0.148 in CP priming, i.e. the decay is stronger for CP priming.

Further parameter estimates can be found in the full specification of the model (Table 2.2).

2.2.3 Discussion

Syntactic rules (targets) are used more frequently when they occur shortly after the same rule (prime). The closer prime and target occur to one another, the stronger the preference is to repeat. Priming is present within a speaker (PP) and it decays rapidly as well as between speakers (CP).

A log-linear model (for distance) yielded a better fit than a linear-linear one, which is in line with psychological models of attention: activation (salience) of objects decays logarithmically. We revisit this in Chapter 5.

2.3 Experiment 2: Repetition in task-oriented dialogue

The effects shown in Experiment 1 could be attributed not just to a low-level syntactic priming effect. Due to the uncontrolled nature of the corpus, semantic effects such as topic chains could have played a role, even though lexically repeated material was excluded.

Thus, in this experiment, we try to replicate the previous priming effect on different data.

2.3.1 Method

The method to detect a priming effect is as in Experiment 1.

To determine whether syntactic repetition effects can occur in dialogue where topics aren't radically shifted, and where the overall semantics are controlled using a set task, we analyzed the *HCRC Map Task* corpus. Map Task comprises more than 110 dialogs with a total of $20,400$ utterances, using 759 different phrase structure rules. Using exactly the same methodology as for Switchboard, we find $402,000$ syntactic repetitions in Map Task between the $157,000$ rules extracted from its syntactic analyses.

Like Switchboard, the Map Task is a corpus of spoken, two-person dialogue in English. Unlike Switchboard, Map Task contains *task-oriented dialogue*: interlocutors work together to achieve a task as quickly and efficiently as possible. Subjects were asked to work together to find a route on a map. The interlocutors are in the same room, but have separate maps and are unable to see each other's maps. One of them, the Instruction Giver, is to describe a route, while the other one, the Instruction Follower, is to follow it on her own map. Their maps differ with respect to names of some locations, certain features (potential waypoints), and missing or displaced labels. Interlocutors were in the same room, while in Switchboard they used a telephone connection.

Syntactic priming as an instance of general priming or pre-activation is an almost universal and mechanistic effect. We accept that some control may be exerted by the conditions of the dialogue and possibly by speakers tailoring their utterances to match the needs of their audience. Still, we would expect to find syntactic priming in any genre, including the task-oriented dialogue of Map Task.

Again, a GLMM was built to correlate priming condition with the set of factors

Covariate	β	*SE*	$p(>\lvert z\rvert)$
Intercept	-1.024	0.025	< 0.0001 ***
ln(DIST)	-0.065	0.011	< 0.0001 ***
ln(FREQ)	0.571	0.025	< 0.0001 ***
PRIMETYPE$_{CP}$	-0.629	0.039	< 0.0001 ***
ln(DIST):ln(FREQ)	0.052	0.011	< 0.0001 ***
ln(DIST):PRIMETYPE$_{CP}$	-0.039	0.018	< 0.05 *
ln(FREQ):PRIMETYPE$_{CP}$	0.214	0.021	< 0.0001 ***

Table 2.3: The model of rule repetition in Map Task. Prime-target distance in utterances.

and predictors.

2.3.2 Results

The minimal model shows a reliable effect of ln(DIST) ($\beta = -0.065$, $p < 0.0001$), indicating that repetition becomes less likely as the distance between prime and target increases. This decay in repetition probability indicates priming.

ln(FREQ) interacts reliably with ln(DIST) ($\beta = 0.052$, $p < 0.0001$), which suggests that repetition probability decreases less quickly for rules with high frequencies. That is, we find less priming for more common rules.

PRIMETYPE$_{CP}$ also interacts reliably with ln(DIST) ($\beta = -0.039$, $p < 0.05$), which suggests that repetition probability decreases more quickly for the CP case, that is, comprehension-production priming (between speakers). That is, we find stronger priming between speakers than within speakers in Map Task.

Further effects are shown in the full model specification in Table 2.3. To produce this model, a three-way interaction between the three covariates discussed was initially fitted, found to be non-reliable ($\beta = 0.036$, $p = 0.132$) and removed, before the model was re-fitted.

We see a reliable decay of repetition probability, which we analyze as syntactic priming.

2.3.3 Discussion

Once again we find that repetition is more likely the shorter the distance between prime and target utterances is. Unlike in Switchboard, interlocutors repeat one another's syntactic structures more readily and more similarly to the way they repeat their own structures.

This finding confirms experimental results by Bock and Griffin (2000) and Branigan et al. (1999), who find syntactic priming over longer distances, even though the effect decays.[5]

It is remarkable that the priming effect decays very rapidly, reaching levels indistinguishable from the prior after about $5-6$ seconds. At first sight, this contrasts with Szmrecsanyi's (2006, p. 188) results, who finds that future marker choices (*will* vs. *going to*) decay only after 140 words (which would be approximately 45 seconds at a speech rate of 180 words/min). However, as Szmrecsanyi points out, due to the logarithmic nature of the forgetting function, most of the priming effect "declines within an interval of 10 words (...), equivalent to ca. 5 seconds of speech."

2.4 Measuring long-term adaptation

In the following, we widen the examination of priming with a second class of repetition effect: long-term adaptation.

2.4.1 Recent work

For structural priming[6], two repetition effects have been identified. Classical priming effects are strong: around 10% for syntactic rules (Reitter et al., 2006d). However, they decay quickly (Branigan et al., 1999) and reach a low plateau after a few seconds, which likens to the effect to semantic (similarity) priming. What complicates matters is that there is also a different, long-term adaptation effect that is also commonly called (repetition) priming.

Adaptation has been shown to last longer, from minutes (Bock and Griffin, 2000) to several days. Lexical boost interactions, where the lexical repetition of material

[5]The effect of PRIMETYPE on bias may be related to general levels of speaker idiosyncrasies, i.e. increased chance repetition within speakers. Fitting the main effect controls for that.

[6]in production and comprehension, which we do not distinguish further for space reasons. Our data are (off-line) production data.

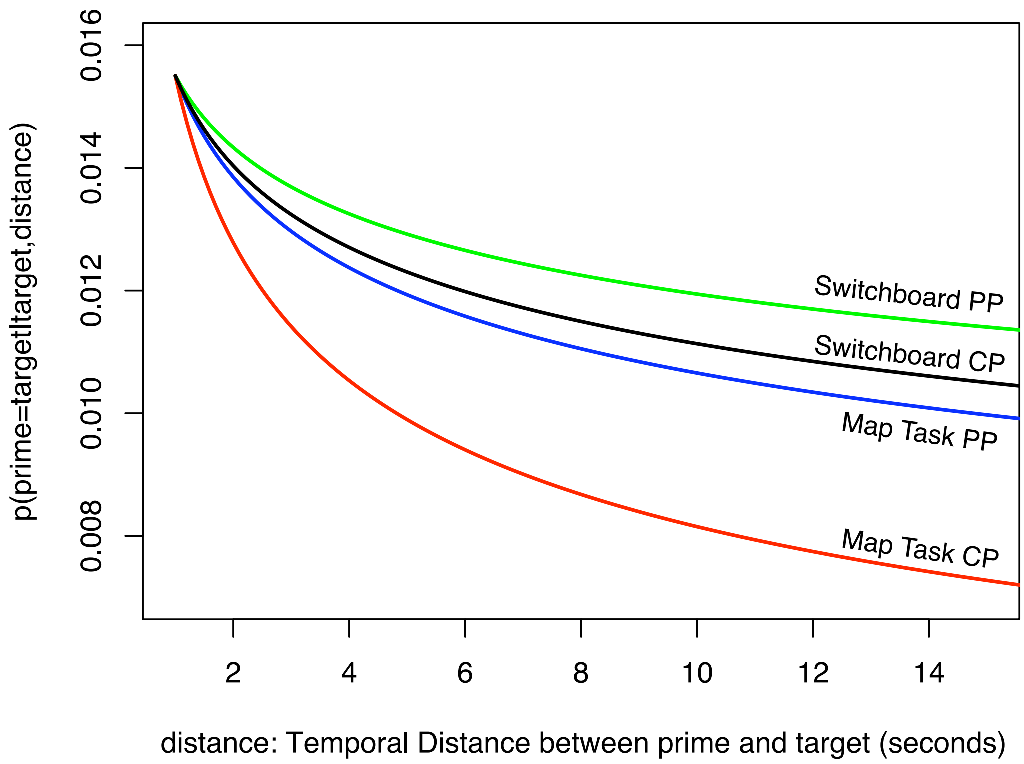


Figure 2.2: Decaying repetition probability estimates depending on the increasing distance between prime and target, contrasting different situations of SOURCE: spontaneous conversation in Switchboard, and task-oriented dialogue in Map Task, and PRIMETYPE: Comprehension-Production (CP) and Production-Production (PP) priming. (Experiment 5)

within the repeated structure strengthens structural priming, have been observed for short-term priming, but not for long-term priming trials where material intervened between prime and target utterances (Konopka and Bock, 2005). Thus, short- and long-term adaptation effects may well be due to separate cognitive processes, as recently argued by V. Ferreira (2006). Section 2.1 deals with decay-based short-term priming, Section 2.4 with long-term adaptation.

Church (2000) proposes adaptive language models to account for lexical adaptation. Each document is split into *prime* and *target* halves. Then, for selected words w, the model estimates

$$P(+adapt) = P(w \in target | w \in prime)$$

$P(+adapt)$ is higher than $P_{prior} = P(w \in target)$, which is not surprising, since texts are usually about a limited number of topics.

This method looks at repetition over whole document halves, independently of decay. To measure long-term adaptation in dialogue, we apply a similar technique to syntactic rules, where we expect to estimate structural priming effects of the long-term variety.

Rather than directly estimating conditional probabilities, we use linear models, as they allow us to take potential confounds into account, build a model for repeated measures and analyze non-discrete predictors including interactions, such as frequencies and temporal distances.

2.4.2 Method

After the initial few seconds, structural repetition shows little decay, but can be demonstrated even minutes or longer after the stimulus. To measure this type of adaptation, we need a different strategy to estimate the size of this effect.

While short-term priming can be pin-pointed using the characteristic decay, for long-term priming we need to inspect whole dialogues and construct and contrast dialogues where priming is possible and ones where it is not. Factor SAMEDOC distinguishes the two situations: 1) Priming can happen in contiguous dialogues. We treat the first half of the dialogue as priming period, and the rule instances in the second half as targets. 2) The control case is when priming cannot have taken place, i.e., between unrelated dialogues. Prime period and targets stem from separate randomly sampled dialogue halves that always come from different dialogues.

Thus, our model estimates the influence of priming on rule repetition. From a Bayesian perspective, we would say that the second kind of data (non-priming) allow the model to estimate a prior probability for rule repetitions. The goal is now to establish a correlation between SAMEDOC and the existence of repetition. If and only if there is long-term adaptation would we expect such a correlation.

Analogous to the short-term priming model, we define repetition as the occurrence of a prime within the first document half (PRIME), and sample rule instances from the second document half. To exclude short-term priming effects, we drop a 10-second portion in the middle of the dialogues.

2.5 Critical discussion

The methodology presented here departs from previous experimental work in several aspects.

2.5.1 Corpus data

The two measures of priming effects introduced apply to data in corpora. The collection of such corpus data does not require the design of an artificial situation, in which subjects necessarily become aware of their own linguistic output. Instead, they can concentrate the task at hand, or on the semantics of the conversation.

Corpora generally suffer from a lack of control w.r.t. to any specific hypothesis. This is certainly true for the Switchboard corpus, and even in the Map Task corpus, where environmental conditions and the exact task given to the subjects were carefully planned, we find no normalization of the language used by either interlocutor. The lack of control introduces two issues. *Noise* is random variation in the data. In any data analysis, this issue is dealt with statistically. *Confounds* can be addressed by introducing controls and, for known confounds, explicit random effects.

Corpora are typically larger than datasets gained from controlled experiments designed to examine just one hypothesis. Thus, we have an opportunity to investigate questions that involve small effects and multiple interactions. But it should be noted that data points gained from linguistic corpora are never independent samples of language or communication (Kilgarriff, 2005). For instance, a single utterance will typically yield multiple syntactic data points, but of course, the choices of syntactic constructions in a sentence depend heavily on each other. In a hypothetical, basic experimental setting, an utterance would count as a single trial, with filler sentences intervening in order to make each trial independent from the previous one. In the corpus study presented here, care is taken to group such linguistic interdependencies in the (random effects) models.

2.5.2 Speaker- and domain-specific differences in sub-languages

The use of a control (or, in Bayesian terms, a prior), is what allows us to address another concern. Is the repetition due to a particular sub-language established by a speaker, or established to accomplish the task responsible for the priming effects?

Again, this repetition due to the reduction of syntactic choices in a sub-language will appear in the control. Thus, it will be factored out.

2.5.3 Size of the prime period

The response variable used to determine priming encodes whether repetition occurred. Repetition is defined as the occurrence of a given syntactic structure (rule) within a period of time (*prime window*). For short-term priming, this prime window is held constant at one second. However, the a-priori probability of repetition occurring anywhere in the prime window also depends on the overall number of rule instances that occur in it. In other words: a fast speaker will show more overall repetition.

If language showed a clustering effect, it would possibly prompt speakers to significantly alter their speech production rates. Then, syntactic rules would cluster, causing higher chance repetition probabilities at short prime-target intervals. This represents a potential confound.

Tests were conducted on both datasets used throughout the experiments, using the original phrase structure rule annotation. We noted the number of syntactic rules in the prime window with each data point, #PRIMERULES. A GLMM was fitted using the technique described in this Chapter. Unsurprisingly, the size of the window correlates with *Primed*, the response variable, which indicates whether a given rule occurred in the prime window (e.g., for Map Task, #PRIMERULES, $\beta = 0.029, p < 0.0001$, and similarly for the log-transformed ln#PRIMERULES to reduce non-normality, and also similar for the joint dataset of Map Task and Switchboard). Crucially, however, the correlation does not impede the decay measurement that indicates short-term priming.

The Map Task speech showed no such clustering effect (#PRIMERULES:ln(DIST)), $p = 0.43$). The speech in Switchboard had a very small interaction in the opposite direction and thus could not present a confound (#PRIMERULES:ln(DIST), $\beta = 0.00069, p < 0.005$). These two models include the decay effect (*Dist*) as well as frequency (*Freq*) and the distinction between production/production and comprehension/production priming (PRIMETYPE) and their interactions. All effects remain significant (at $p < 0.05$ or better).

We conclude that with the given dataset, the varying number of syntactic rules in the one-second prime window does not represent a practical confound of the priming effect observed.

2.5.4 Hypothesis testing in repeated-measurement models

The significance values presented here are derived from t-tests on the parameter estimates (or z tests where appropriate). Recently, concerns have been raised about this common method. Baayen et al. (2008) suggest that the degrees of freedom assumed for the calculation of the p-values may be an overly optimistic upper bound. More importantly, they criticise t-tests for their failure to account for random effects, which are important in the context of the methodology here, as we analyze effects on a by-utterance basis (with an utterance-based grouping factor) to account for non-independence of the samples.

To evaluate and address these concerns, we replicated Experiments 1 and 2 using Markov-Chain Monte-Carlo (MCMC) sampling. For Switchboard, the method gives a 95% confidence interval for the $\ln(\text{DIST})$ parameter of $[-0.116, -0.109]$, i.e. well outside the magnitude of a null effect (0), and similarly for the interaction effect $\ln(\text{DIST})$:PRIMETYPE=CP ($[-0.042, -0.029]$), which applies for CP priming, and the interaction effect $\ln(\text{DIST})$:$\ln(\text{FREQ})$ ($[0.035, 0.047]$), indicating the influence of rule frequency on priming strength. A plot showed that the $\ln(\text{DIST})$ parameter appears to be normally distributed, which is generally expected for such parameters. For Map Task, highly significant effects (by t-tests) correspond to minimal confidence intervals around the fitted parameters.

The results show that effects judged to be significant, the model parameter estimates for one level lie outwith the confidence intervals estimated via MCMC for the contrasting level, as one would expect. Further, Baayen et al. (2008) note that the caveats are less relevant for large datasets such as ours.

2.5.5 Topic chains

A final concern in dialogue and other kinds of text is that text tends to be coherent. Adjacent utterances do not jump from topic to topic—instead, they form clusters. This is a potential confound: could clusters be responsible for the short-term priming effect? After all the short-term effect compares repetition levels shortly after a prime to those far away from the prime, and a topic cluster would produce exactly the effect interpreted as priming here.

The first answer is simple. Topic clusters primarily show up in lexical choices. Lexically repeated material is explicitly excluded from our data. The measures of structural priming consider only repetitions of syntactic structure, where the lexical material differs.

Secondly, excluding topic chains as a cause for increased repetition after a syntactic mention does not mean that semantic effects cannot play a role. Indeed, we would expect more semantic processing in task-oriented dialogue. This may lead to a preference for certain syntactic realizations. As Chapter 3 shows, we find that the short-term comprehension-production priming effect is stronger in spontaneous conversation.

So, while we exclude the possibility that lexical repetition of phrases causes speakers to also repeat syntactic rules, we consider semantic processing to be one of the possible causes of short-term priming effects. We revisit this explanation in Chapter 5.

Chapter 3

Structural Priming in Dialogue

3.1 Introduction

While humans are remarkably efficient, flexible and reliable communicators, we are far from perfect. Our dialogues differ in how successfully information is conveyed. In task-oriented dialogue, interlocutors communicate in order to solve a problem. Experiments can be constructed so that their success at the task depends on successful communication. Pickering and Garrod's (2004) Interactive Alignment Model assumes that priming holds the key to understanding how interlocutors build a common understanding of the situation, which then enables them to successfully communicate and solve a problem at hand. We examine this assumption by investigating the link between priming, the task-solving objective of the dialogue and the achieved success.

The Interactive Alignment Model (IAM) postulates that higher-level (semantic, situation-level) alignment is due to lower-level alignment (including syntactic priming). Priming leads to linguistic adaptation and grounding of situation models during speaker interaction. Priming in lower processing stages reinforces priming in higher ones, up to an alignment of a common situation model.

Some motivation for the IAM came from the *Maze Game* study by Garrod and Anderson (1987). There, participants were presented with a maze shown on a computer screen. The maze consisted of a grid layout, not unlike the one found classically in North American cities. Each participant was placed somewhere in this maze. The objective of the game was for each participant to move to set destinations.

Some of the "streets" could be blocked by gates, which could be temporarily opened or closed by moving to a "switch" at marked positions. Participants could only see their own version of the maze and had to coordinate, so that one participant would navigate to a switch, which would then allow the other participant to move to a given destination.

The maze game was designed to elicit communication between the participants, who would gradually build up a joint way of identifying positions in the maze. Indeed, the analysis of the dialogues showed that participants developed a common language. They converged with respect to the way they identified positions in the maze, for instance by saying *I'm at C 4*, or by saying *I'm one up on the diagonal from the bottom left to top right.*

In Garrod and Anderson's (1987) dialogues, there was no explicit negotiation of a scheme for references. Instead, participants implicitly coordinated their in- and output language. This led the authors to conclude that alignment may be based on more local linguistic effects rather than a grand strategic plan.

A contrasting assumption may be that alignment encodes information to be communicated. In particular, it may communicate agreement with, or respect for, the interlocutor. Indeed, alignment seems higher when certain psychological factors in the perceived relationship between the speakers are present. For example, in a "Wizard of Oz" experiment involving subjects interacting with what they perceived as either a basic or an advanced computer, or a human, subjects aligned deliberately more with the inferior "basic" computer, less with the "advanced" computer and least with the human (Pearson et al., 2004). It seems reasonable to assume a control mechanism in particular if alignment is an acquired communicative convention. Be it deliberate or otherwise controlled or neither, the question remains whether priming is the, or at least one, basis for alignment in dialogue.

In this chapter, we examine syntactic (structural) priming as one of the driving forces behind alignment. We choose syntactic over lexical priming for two reasons. Lexical repetition due to priming is difficult to distinguish from repetition that is due to interlocutors attending to a particular topic of conversation, which, in coherent dialogue, means that topics are clustered. Lexical choice reflects those topics, hence we expect clusters of particular terminology. Secondly: the maps used to collect the dialogues in the Map Task corpus contained landmarks with labels. It is only natural (even if by way of cross-modal priming) that speakers will identify landmarks using the labels and show little variability in lexical choice. We measure repetition of syntactic rules, whereby word-by-word repetition (topicality effects, parroting) is explicitly excluded.

3.2 Background: Priming in different dialogue types

The IAM, together with the idea that priming levels differ between speakers and between dialogue situations, predicts more priming in task-oriented dialogue than in spontaneous conversation, because situation-model level alignment is typically required to perform a given task. In this chapter, we test this prediction by comparing priming in spontaneous and task-oriented dialogue in a first set of experiments (Sections 3.5 through 3.6). Based on available data, we apply and then improve the measure of priming introduced in Chapter 2. Then, we strengthen the test of the IAM by correlating priming and long-term adaptation levels to the degree to which speakers were successful at carrying out a given task (Section 3.7).

The IAM is a relatively novel theory, which, at the point of writing, is still being specified further. In particular, concrete tests of the hypothesized relationship between priming and higher-level alignment are needed. The experiments in this chapter may be seen as a contribution in this respect.

3.3 Experiment 3: Comparing corpora

With their Interactive Alignment Model (IAM), Pickering and Garrod (2004) argue that the situation-model alignment of speakers is due to lower-level priming effects. In task-oriented dialogue, and in the task carried out by participants in Map Task, speakers need to align in order to successfully complete their tasks. Thus, the theory would predict that syntactic priming between speakers (CP) is greater in task-oriented dialogue.

To determine whether there is a significant influence of dialogue type on priming, we compare the effects we have seen in Experiments 1 and 2. To do so, we built a further model, aggregating the two datasets.

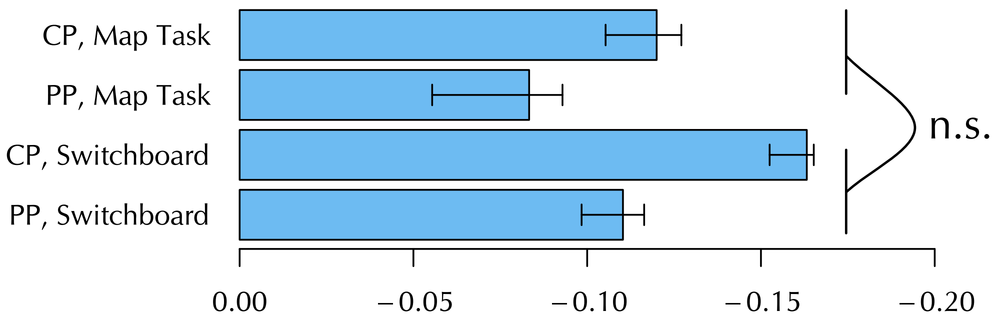


Figure 3.1: Decay effect sizes for ln(DIST) with different combinations of PRIMETYPE and SOURCE factors. Effects are given in logits, prime-target distance (ln(DIST)) measured by number of utterances (Experiment 3). Longer bars indicate stronger decay and priming. 95% confidence intervals were estimated with MCMC sampling. Note that the model (see text) does not suggest a reliable interaction effect of SOURCE and ln(DIST).

3.3.1 Method

We test the hypothesis suggested by the IAM by fitting a model of the joint dataset with SOURCE as a binary factor to indicate whether a repetition stems from Map Task (task-oriented) or Switchboard (not task-oriented). In order to match the voice-only modality in Switchboard, only Map Task dialogues in which interlocutors could not see one another were included.

The method to detect priming via the DIST variable and GLMMs is as in previous experiments. We use an interaction of DIST and SOURCE to determine whether priming levels differ between the two datasets.

3.3.2 Results

We find an effect of ln(DIST) ($\beta = -0.134$, $p < 0.0001$), indicating decay. (Note that this magnitude applies to the baseline condition, i.e., Switchboard and PP priming.)

The model estimates no reliable interaction effect of ln(DIST) with SOURCE (i.e., for $\text{SOURCE}_{MapTask}$: $\beta = -0.03$, $p = 0.096$). If this was significant, it would have resulted in a lower resulting estimate for ln(DIST) in the Map Task condition, suggesting that decay in Map Task was stronger than in Switchboard.

We find an interaction of PRIMETYPE_{CP} with ln(DIST) ($\beta = -0.042$, $p < 0.0001$), indicating stronger CP than PP priming. A three-way interaction of ln(DIST),

Covariate	β	SE	$p(>\|z\|)$
Intercept	0.584	0.016	< 0.0001 ***
ln(DIST)	−0.134	0.007	< 0.0001 ***
ln(FREQ)	0.831	0.006	< 0.0001 ***
PRIMETYPE$_{CP}$	−0.299	0.015	< 0.0001 ***
SOURCE$_{MapTask}$	0.139	0.042	< 0.001 ***
PRIMETYPE$_{CP}$: SOURCE$_{MapTask}$	−0.474	0.026	< 0.0001 ***
ln(DIST): ln(FREQ)	0.034	0.003	< 0.0001 ***
ln(DIST): PRIMETYPE$_{CP}$	0.042	0.007	< 0.0001 ***
ln(DIST): SOURCE$_{MapTask}$	−0.03	0.018	0.096

Table 3.1: The regression model for the joint dataset of Switchboard and Map Task (Experiment 3) with distance measured in utterances.

PRIMETYPE$_{CP}$ and SOURCE$_{MapTask}$ ($\beta = 0.011$, $p < 0.738$) was not reliable and removed from the model before the final reduced model presented here was fitted.

The contrastive analysis (Figure 3.1) produced by Markov-Chain Monte-Carlo sampling recasts the resulting effect sizes for ln(DIST) in terms of the different factor combinations. It shows that priming effects are comparable across the conditions except for comprehension-production priming, which appears stronger in Switchboard when distance is measured by utterances.

Frequency is correlated with ln(DIST) ($\beta = 0.034$, $p < 0.0001$), indicating stronger priming for lower-frequency rules.

Table 3.1 specifies the complete model after step-wise reduction to the significant covariates. Note that the fitted effects diverge from the contrastive analysis (Figure 3.1), which was produced from the full model including the non-significant interactions.

3.3.3 Discussion

As seen in the previous experiments, it can make a difference whether a speaker primes themself or is primed by their interlocutor. See also Figure 3.1 which provides the resulting priming strength estimates for the four factorial combinations of PRIMETYPE and SOURCE. Also, priming is stronger for less frequent rules.

Finding the marked difference between CP and PP priming, and also a clear PP priming effect in spontaneous conversation represents an advancement compared to Dubey et al. (2005), who do not find reliable evidence of adaptation within speakers in Switchboard for selected syntactic rules in coordinate structures.

The present data show little support for the hypothesis that semantic alignment in dialogue is based on lower-level (syntactic) priming. We would expect the difference to apply primarily to priming between speakers (CP), and not to priming within a speaker (PP). The present results cannot support this prediction: the reduced model yields unreliable effects for the relevant PRIMETYPE factor, and the contrastive analysis suggests an effect in the opposite direction.

When comparing data across corpora, we need to be careful to ensure that differences in genre and annotation are not the primary cause of the effect at hand. The coefficient for pre-activation decay is sensitive to utterance length, which becomes an issue when, for instance, utterances are not consistently marked or if decay occurs over time and not with utterances. Indeed, most utterances in Switchboard are actually *dialogue turns*, and given the genre, they are usually longer than those in Map Task. Even if priming decay takes place with linguistic activity, utterances do not serve as a sensible measure given the difference in utterance length.[1]

Therefore, it makes sense to re-address the hypothesis using *time* as the relevant decay dimension. We do so in Experiments 4 and 5.

3.4 Experiment 4: Decay over time, or with each utterance?

While the previous experiments have shown that repetition probability decays soon after any stimulus, it is unclear whether the pre-activation diminishes with time, or with actual linguistic activity. To some extent, corpora can help to make that distinction.

The differences between conversational and task-oriented dialogue that we pointed out (Experiment 3) are founded on the correlation of distance between prime and target and the likelihood of repetition. As stated before, this correlation is likely to be sensitive to the *scale* of ln(DIST). As an alternative, we can use the delay between the left boundaries of the priming and target phrases as the relevant

[1] On the other hand, CP and PP priming strengths are similar across the two corpora, which speaks again for the quality of the model.

predictor.

The models discussed measure the distance between prime and target in *utterances*. In this experiment, we fitted a second regression model, estimating decay over *time*.

3.4.1 Methods

To compare the two (obviously co-linear) predictors $\textsc{Dist}_{Time}$ and $\textsc{Dist}_{Utts}$, we estimated two simple linear regression models, one for time, the other one for number of utterances as predictor. Such simple linear regression models can, as opposed to GLMMs, produce a meaningful R^2 measure. In these models, we include the maximum-likelihood estimate of the number of chance repetitions, which is calculated from the overall frequency of each syntactic rule (this is in addition to the covariates discussed before). The response variable here is not binary, as in the other experiments, but a count of actual rule repetitions. The complete interaction term is $rep \sim ln(\textsc{Dist}_{Utts}) * \textsc{PrimeType} * \textsc{Source} + \textsc{Expected}$.[2]

The goodness-of-fit measure R^2 helps us determine how much of the variance in our data is explained by the model.

3.4.2 Results

For distance over utterances, R^2 is 0.91, for time (in 1-second buckets) it is 0.89, a similar size.

3.4.3 Discussion

Thus, there is no compelling empirical evidence to assume either time or utterances as the scale for decay.

While we cannot reasonably opt for one of the alternatives based on the present empirical result, there are arguments for a time-based measure. The unit of an *utterance* is not well-defined, especially for spoken language. Even if we take *turns* to be utterances, they differ greatly between corpora or even between speaker dyads. Both utterances and turns are delimited by human annotators, who depend on

[2]These models assume a normal distribution as opposed to the appropriate Poisson one. We therefore do not make claims based on the effect size estimates, but believe that the two models are commensurable.

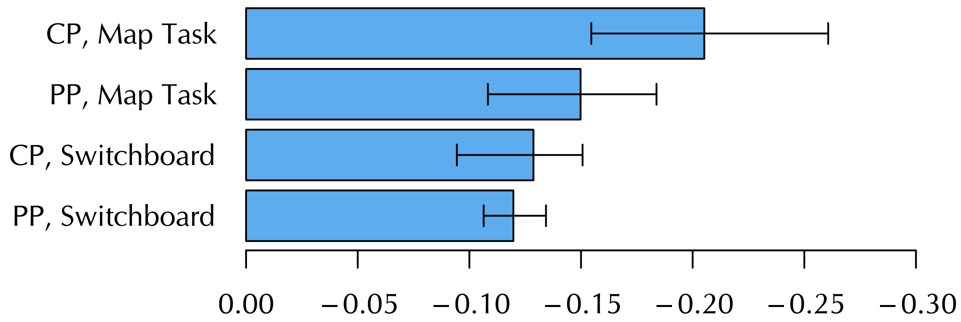


Figure 3.2: Decay effect sizes for ln(DIST) with different combinations of PRIMETYPE and SOURCE factors. Effect sizes for ln(DIST) in logits, prime-target distance is measured in seconds. (Experiment 5). Longer bars indicate stronger decay and priming. 95% confidence intervals were estimated with MCMC sampling.

clear instructions to be reliable. Time, on the other hand, is an objective and precise measure that can be obtained for many corpora.

We re-evaluate the effect of corpus choice seen in Experiment 3, this time using *time* as the decay scale. Time is also used as measure of distance in further experiments presented in Chapter 4.

3.5 Experiment 5: Priming over time

While time- and utterance-based models fit their respective data similarly well, *time* is a theoretically attractive measure of distance, in particular because the *utterance* is difficult to delineate in the context of speech.

3.5.1 Method

The methodology of this experiment is as it was in Experiment 3, except that DIST_{Time} is the distance predictor, instead of the DIST_{Utts} used previously.

3.5.2 Results

We find an effect of ln(DIST) ($\beta = -0.128$, $p < 0.0001$). This indicates priming in Switchboard. We also find an interaction of ln(DIST) with $\text{SOURCE}_{MapTask}$ ($\beta = -0.045$, $p < 0.05$), indicating stronger priming in Map Task.

Covariate	β	SE	$p(>\lvert z \rvert)$
Intercept	0.551	0.016	< 0.0001 ***
ln(DIST)	−0.128	0.007	< 0.0001 ***
ln(FREQ)	0.572	0.012	< 0.0001 ***
PRIMETYPE_{CP}	−0.157	0.012	< 0.0001 ***
$\text{SOURCE}_{MapTask}$	0.066	0.047	0.154
PRIMETYPE_{CP}: $\text{SOURCE}_{MapTask}$	−0.197	0.034	< 0.0001 ***
ln(DIST): ln(FREQ)	0.097	0.006	< 0.0001 ***
ln(DIST): $\text{SOURCE}_{MapTask}$	−0.045	0.019	< 0.05 *

Table 3.2: The regression model for the joint dataset of Switchboard and Map Task (Experiment 5), distance measured in seconds. This is the minimal model without unjustified covariates. (Laplace fit. Random intercept, grouped by utterances.)

An interaction between ln(DIST), PRIMETYPE_{CP} and $\text{SOURCE}_{MapTask}$ did not show a reliable effect ($\beta = -0.044$, $p = 0.300$) and was removed from the model. Subsequently, PRIMETYPE reliably interact with ln(DIST) ($\beta = -0.014$, $p = 0.345$). The reduced model reported here does not contain these covariates, except for the contrasts shown in Figure 3.2.

As before, ln(DIST) interacted with ln(FREQ) ($\beta = 0.097$, $p < 0.0001$), i.e., priming is stronger for less frequent rules.

Table 3.2 provides the reduced fitted model in full. Figure 3.2 (p. 59) shows the contrasts analysis with resulting priming strength estimates for the four factor combinations of PRIMETYPE and SOURCE. The confidence intervals, calculated with the more conservative Markov-Chain Monte-Carlo sampling suggest that PRIMETYPE is still a reliable factor within Switchboard.

Also, refer back to Figure 2.2 (p. 46), which illustrates the repetition probability as it decays with time for the four combinations of PRIMETYPE.

3.5.3 Discussion

The model based on temporal distance makes stronger predictions than the comparison based on utterances. The basic result from Experiments 1, 2 and 3 hold: there is priming in both corpora.

While Experiment 3 could not find reliable evidence for stronger priming in task-oriented dialogue (nor against it), this experiment now lends support to our initial hypothesis. Priming appears to be stronger in the Map Task corpus.

We return to these results in the Conclusion (Section 3.11) and again in Chapter 5, where we propose an alternative explanation.

3.6 General discussion

Both corpora of spoken dialogue which we investigated showed an effect of distance between prime and target on syntactic repetition probability, thus providing evidence for a structural priming effect for arbitrary syntactic rules. In both corpora, we also found reliable effects of both production-production (PP) priming (self-priming) and comprehension-production priming. But in the Map Task, a corpus of task-oriented dialogue, we find evidence for stronger overall priming than in Switchboard.

A possible explanation for these results is the reduced cognitive load that we can reasonably assume for spontaneous, everyday conversation (as in the Switchboard corpus). Pickering and Garrod (2004) suggest that interlocutors reduce their workload by aligning their linguistic and semantic representations, as re-using structure is easier than creating it. As cognitive load in non-task oriented, spontaneous conversation is low, speakers reduce the amount of priming that is required in dialogue that related to a difficult task. The fact that we consistently see stronger priming for less frequent syntactic rules supports the cognitive-load explanation: frequently used rules are more accessible, hence their representations need less preactivation.

Another reason may simply be that interlocutors in Switchboard (as in all spontaneous dialogue) switch topics frequently, engaging in longer turns in between. The length of the turns should influence the priming effects only if all turns are taken as one unit, but not if the distance (lag) between prime and target is measured temporally, in which case the turn length can be considered controlled for. It turned out that the difference in structural priming was evident even in the analyses over time. Such a sequence of monologues may, in general, be less affected by priming. The hypothesis that topic switches reduce priming may be tested in a future study.

On the other hand, one could expect that the narrow-bandwidth single channel (phone line in Switchboard) leads speakers to make an effort to at least accept more self-priming (PP), designing their message so that they could be easily understood. Such *audience design* would be in line with work by Pearson et al. (2004), who found that speakers use less alignment (or priming) when talking to an (artificial) interlocutor that was perceived to have better linguistic capabilities. However, we see little actual evidence of speakers having difficulty understanding each other over the phone line.

At the same time, speakers may also have had more difficulty in producing speech, lacking the visual feedback that a direct conversation offers. Pickering and Garrod (2004) actually foresee this possibility (monologue is more difficult than dialogue). However, visual feedback in Map Task was limited by experimental setup, for example because participants where looking at their maps. The fact that both participants were in the same room during the Map Task experiments gave them a richer communication channel, which may have affected their predisposition to temporarily adapt to each other. The data used in the comparative experiments was constrained to a condition in Map Task where participants were separated by a screen, so no eye-contact was possible. Furthermore, we have tested the hypothesis that syntactic priming is sensitive to visual contact in a further experiment. There, we compared short-term syntactic priming in a condition that allowed Map Task subjects to see each other to priming in a condition where subjects were separated using a screen. We did not find evidence for an influence of eye-contact on syntactic priming.

3.7 Preliminary conclusions

Reliable syntactic priming effects can be detected in dialogue even when the full range of syntactic rules is taken into account instead of selected constructions with known strong priming effects. We have modelled syntactic priming as the decay of repetition probability of syntactic rules, either in the course of linguistic activity, or over time.

The parameters of priming vary with the setting of the conversation. In particular, we believe that the task-orientedness of the dialogue and increased cognitive load may boost alignment between speakers. The Interactive Alignment

Model (Pickering and Garrod, 2004) provides a viable explanation for the different effects that the two corpora expose. What we observe is the reciprocal boosting of syntactic priming and the alignment of the situation models present in task-oriented dialogue. The interaction partners synchronize their situation models in the task-oriented setting, which co-occurs with cross-speaker priming (CP) on other communicative levels. While self-priming may have to do with reduced cognitive load in production, the CP priming appears to be enhanced by sharing a situation model.

Up to now, we have found that priming levels differ between spontaneous conversation in one corpus, and task-oriented dialogue in another corpus. The difference is particularly marked for priming between speakers.

This difference, on its own, is in line with the interactive alignment model. However, the IAM is not the only possible explanation. The dialogues in the two corpora differ greatly with respect to the overall goals of the speakers, their mode of interaction, the durations of their turns, their language register and their linguistic variability. While the underlying methodology can be expected to be robust with respect to differences in language, it is still unclear as to whether confounding factors could have affected actual priming levels. Furthermore, the correlation between dialogue type and priming is just that: a correlation, and not a statement of cause and effect.

The next experiments address these potential concerns. We examine only data from the Map Task corpus, collected under well-controlled conditions. We also broaden our view to distinguish short-term and long-term adaptation (see Chapter 2), and to evaluate to what extent task success can be predicted and estimated based on lexical and syntactic adaptation.

3.8 Experiment 6: Task success and short-term priming

In this section, we attempt to detect differences in the strength of short-term priming in successful and less successful dialogues. To do so, we use the measure of priming strength established in the previous sections of this chapter, which then allows us to test whether priming interacts with task success. Under the assumptions of IAM we would expect successful dialogues to show more priming than unsuccessful ones.

Obviously, difficulties with the task at hand may be due to a range of problems that the subjects may have, linguistic and otherwise. But given that the dialogues contain variable levels of syntactic priming, one would expect that this has at least some influence on the outcome of the task.

3.8.1 Methods

3.8.1.1 Data

We used the 128 dialogues from the HCRC Map Task corpus (Anderson et al., 1991). To understand how dialogue success is measured in Map Task, consider the design of the corpus collection experiment. Participants were given two slightly different maps depicting the same (imaginary) landscape. One participant was to give directions for a predefined route to another subject, who followed them, drawing a route on their own map. The spoken interactions were recorded, transcribed and syntactically annotated with phrase structure grammar.

The Map Task provides us with a precise measure of success, namely the deviation of the predefined and followed route. Success can be quantified by computing the inverse deviation between subjects' paths. Both subjects in each trial were asked to draw "their" respective route on the map that they were given. The deviation between the respective paths drawn by interlocutors was then determined as the area covered in between the paths (PATHDEV).

In this experiment, the short-term priming method described in Chapter 2 was used to correlate the priming effects established earlier (see Experiments 2 and 5) with path deviation by way of an interaction of DIST and PATHDEV.

Prime-target distance $\ln(\text{DIST})$ is measured in time (seconds).

3.8.2 Results

As before short-term priming is reliably correlated (negatively) with $\ln(\text{DIST})$, hence we see a decay and priming effect ($\ln(\text{DIST})$, $\beta = -0.164$, $p < 0.0001$).

Notably, path deviation and short-term priming did not correlate. The model showed there was no such interaction ($\ln(\text{DIST})$:PATHDEV, $\beta = 0.0001$, $p = 0.586$).

We also tested for an interaction with an additional factor indicating whether prime and target were uttered by the same or a different speaker (comprehension-production vs. production-production priming). This interaction did not approach

Covariate	β	SE	$p(>\|z\|)$
Intercept	-1.096	0.043	< 0.0001 ***
ln(DIST)	-0.164	0.018	< 0.0001 ***
ln(FREQ)	0.509	0.03	< 0.0001 ***
PATHDEV	0.000	0.001	0.831
ln(DIST):ln(FREQ)	0.077	0.012	< 0.0001 ***
ln(DIST):PATHDEV	0.000	0.000	0.586

Table 3.3: The full regression model for the Map Task dataset (Experiment 6).

reliability (ln(DIST):PATHDEV:PRIMETYPE$_{CP}$, $\beta = -0.0004$, $p = 0.60$).

We also tested whether priming changes over time over the course of each dialogue. There were no reliable interaction effects of centered prime/target times (ln(DIST):ln(STARTTIME), $\beta = 0.011$, $p =$,0.75; ln(DIST): PATHDEV:ln(STARTTIME), $\beta = -0.0002$, $p = 0.63$). Reducing the model by removing unreliable interactions did not yield any reliable effects. Table 3.3 specifies the full model.

3.8.3 Discussion

We have shown that while there is a clear priming effect in the short term, the size of this priming effect does not correlate with task success.

Does this indicate that there is no strong functional component to priming in the dialogue context? There may still be an influence of cognitive load due to speakers working on the task, or an overall disposition for higher priming in task-oriented dialogue: Experiment 5 points to stronger priming in such situations. Our results are difficult to reconcile with the model suggested by Pickering and Garrod (2004), if we take short-term priming as the driving force behind IAM.

Short-term priming decays within a few seconds. Thus, to what extent could syntactic priming help interlocutors align their situation models? In the Map Task experiments, interlocutors need to refer to landmarks regularly—but not every few seconds. It would be sensible to expect longer-term adaptation (within minutes) to drive dialogue success.

3.9 Experiment 7: Task success and long-term adaptation

Long-term adaptation is a form of priming that occurs over minutes and could, therefore, support linguistic and situation model alignment in task-oriented dialogue. IAM could be based on such an effect instead of short-term priming. Analogous to the previous experiment, we hypothesize that more adaptation relates to more task success.

Pickering and Garrod (2004) do not make the type of priming supporting alignment explicit. Should we find differences in the way task success interacts with different kinds of repetition effects, then this would be a good indication about which of the effects supports IAM. More concretely, we could tell whether alignment is due to the automatic, classical *priming* effect, or whether it is based on a long-term effect that is possibly related to implicit learning (Bock and Griffin, 2000; Chang et al., 2006).

3.9.1 Method

Simple, exploratory data analysis (Figure 3.3) shows a correlation between repetition probability in the Map Task dialogues and task success: as path deviation increases, repetition probability increases. Because path deviation is seen as negative task success, the correlation appears to be positive.

To analyze the data, we turn to the long-term adaptation measure introduced in Section 2.4, which uses pairs of dialogue halves. We distinguish two conditions with a SAMEDOC factor: pairs of dialogue halves that stem from the same document, in which case we can expect an adaptation effect, and pairs of dialogue halves stemming from randomly chosen dialogues (the control). PRIME is a binary coding for an instance of rule repetition, which is true if a given instance of a syntactic rule taken from the second dialogue half has occurred anywhere in the first dialogue half. (This coding is intended to be similar to the short-term priming analysis.)

Using the same data as in Experiment 6, task success is inverse path deviation PATHDEV as before, which should, under IAM assumptions, interact with the effect estimated for SAMEDOC. Thus, we fit the model PRIME $\sim$ SAMEDOC $*$ PATHDEV.

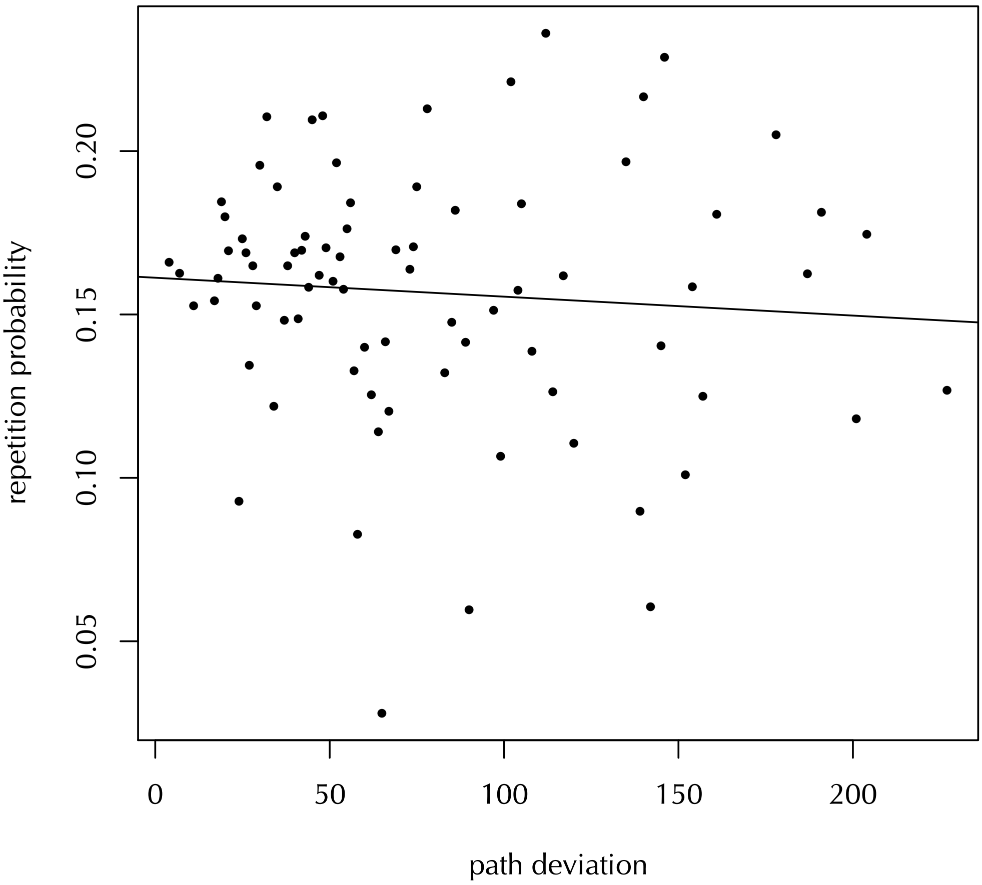


Figure 3.3: Maximum likelihood rule repetition probability for each Map Task dialogue over path deviation (PATHDEV) (negative task success). Trend: linear correlation between probability and path deviation.

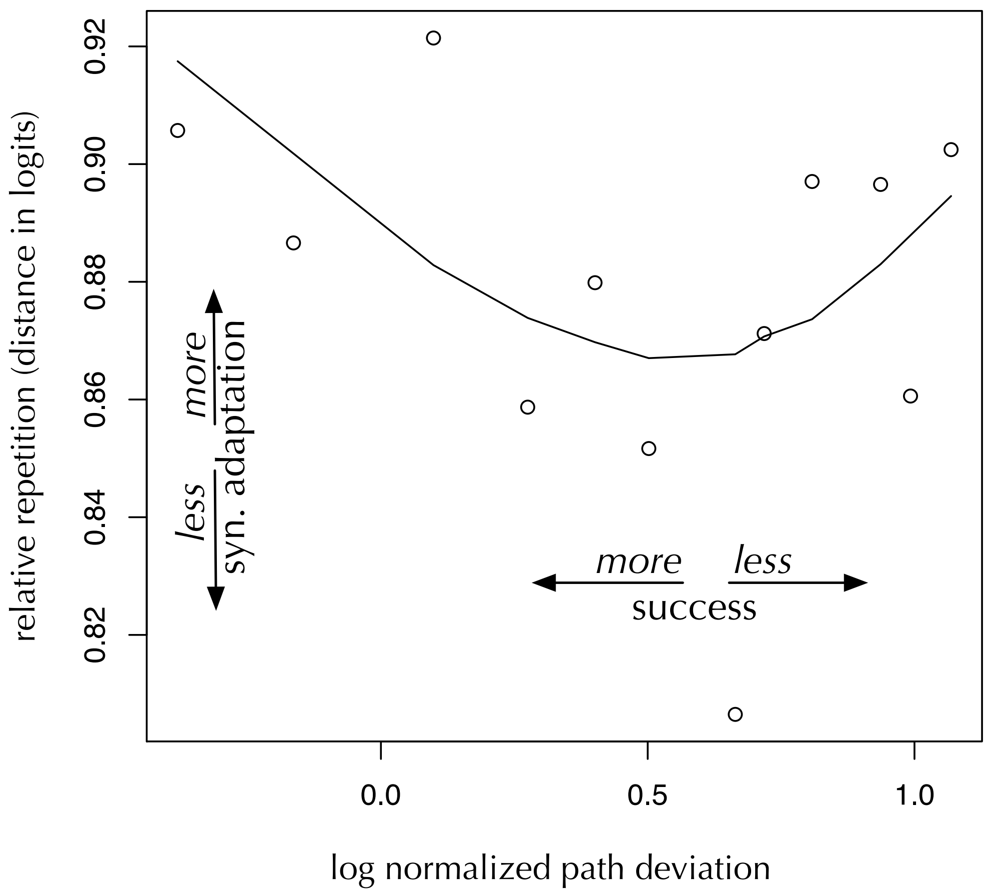


Figure 3.4: Log-odd ratios of rule repetition probability over normalized task success. Here, we show the relative rule repetition probability, which chance repetition excluded. Probabilities are aggregated over dialogues with similar path deviations. Task success is shown as (negative) normalized path deviation ($d/\mu(d)$) on a log scale. Trend: moving average.

3.9.2 Results

SAMEDOC showed a reliable, positive effect ($\beta = 3.303$, $p < 0.0001$), which means we see long-term repetition. This generalizes previous experimental priming results in long-term priming.

The effect interacted reliably with the path deviation scores (SAMEDOC:PATHDEV, $\beta = -0.624$, $p < 0.05$). Thus, we find a reliable correlation of task success and syntactic priming. Stronger path deviations relate to weaker priming.

The normalized rule frequency ln(FREQ) did not interact with SAMEDOC ($\beta = -0.044$, $p = 0.35$). The interaction was removed for all other parameters reported.[3]

3.9.3 Discussion

The more syntactic priming speakers show, the better do they perform at synchronizing their routes on the maps. This is exactly what one would expect under the assumption of IAM. Also, there is no evidence for stronger long-term adaptation of rare rules, which may point out a qualitative difference to short-term priming.

Of course, this correlation does not necessarily indicate a causal relationship. Still, participants in Map Task did not receive an explicit indication about whether they were on the "right track". Mistakes, such as passing a landmark on its East and not on the West side, were made and went unnoticed. Thus, it is not very likely that task success caused alignment to improve at large. We suspect such a possibility, however, for very unsuccessful dialogues. A closer look at the correlation (Figure 3.4) reveals that while adaptation indeed decreases as task success decreases, adaptation increased again for some of the least successful dialogues. It is possible that here, miscoordination became apparent to the participants, who then tried to switch strategies. Or, simply put: too much alignment (and too little risk-taking) is unhelpful. Further, qualitative, work needs to be done to investigate this hypothesis.

From an applied perspective, the correlation shows that the repetition effect that contributes to prediction accuracy is long-term syntactic adaptation as opposed to the more automatic short-term priming. We take this as an indication to include adaptation rather than just priming in a model of alignment in dialogue.

[3] Such an interaction also could not be found in a reduced model with only SAMEDOC and ln(FREQ).

3.10 Application: Predicting task success

An automatic measure of task success would be useful for evaluating conversations among humans, e.g., for evaluating agents in a call center. In human-computer dialogues, predicting the task success after just a first few turns of the conversation could avoid disappointment: if the conversation isn't going well, a caller may be passed on to a human operator, or the system may switch dialogue strategies. As a first step, we focus on human-human dialogue, since current spoken dialogue systems do not yet yield long, syntactically complex conversations.

In this section, we use syntactic and lexical features to predict task success in an environment where we assume no speaker model, no semantic information and no information typical for a human-computer dialogue system, such as the confidence reported by the automatic speech recognizer. The features we use link alignment between dialogue participants to low-level syntactic priming.

3.10.1 Previous approaches

Prior work on predicting task success has been done in the context of human-computer spoken dialogue systems. Features such as recognition error rates, natural language understanding confidence and context shifts, confirmations and re-prompts (dialogue management) have been used to classify dialogues into *successful* and *problematic* ones (Walker et al., 2000). With these automatically obtainable features, an accuracy of 79% can be achieved given the first two turns of "How may I help you?" dialogues, where callers are supposed to be routed given a short statement from them about what they would like to do. From the whole interaction (very rarely more than five turns), 87% accuracy can be achieved (36% of dialogues had been hand-labeled "problematic"). However, the most predictive features, which related to automatic speech recognition errors, are neither available in the human-human dialogue we are concerned with, nor are they likely to be the cause of communication problems there.

Moreover, failures in the Map Task dialogues are due to the actual goings-on when two interlocutors engage in collaborative problem-solving to jointly reach an understanding. In such dialogues, interlocutors work over a period of about half an hour. To predict their degree of success, we leverage the phenomenon of *persistence*, or *priming*.

3.10.2 Experiment 8: The success prediction task

In the following, we define two variants of the task and then describe a model that uses repetition effects to predict success.

Task 1: *Success is estimated* when an entire dialogue is given. All linguistic and non-linguistic information available may be used. This task reflects post-hoc analysis applications, where dialogues need to be evaluated without the actual success measure being available for each dialogue. This covers cases where, e.g., it is unclear whether a call center agent or an automated system actually responded to the call satisfactorily.

Task 2: *Success is predicted* when just the initial 5-minute portion of the dialogue is available. A dialogue system's or a call center agent's strategy may be influenced depending on such a prediction.

3.10.3 Method

To address the tasks described in the previous Section, we train Support Vector Machines (SVM) to predict the task success score of a dialogue from lexical and syntactic repetition information accumulated up to a specified point in time in the dialogue.

The primary idea in this applied approach is to solve the task as well as we can with a given, limited set of features. While we hope to confirm the priming—task success link demonstrated with linear models, the emphasis here is on task performance rather than on the model's parsimony or on estimates than can be interpreted with respect to the initial hypothesis. It is for this reason that we choose to address the task with a high-performance machine-learning algorithm, whose fitting algorithm and model representation attempts to maximize classification performance.

Repetition is measured on a lexical and a syntactic level. To do so, we identify all constituents in the utterances as per phrase structure analysis. *[Go [to [the [[white house] [on [the right]]]]]]* would yield 11 constituents. Each constituent is licensed by a syntactic rule, for instance $VP \rightarrow V\ PP$ for the top-most constituent in the above example.

For each constituent, we check whether it is a lexical or syntactic repetition, i.e., if the same words occurred before, or if the licensing rule has occurred before in the same dialogue. If so, we increment counters for lexical and/or syntactic repetitions, and increase a further counter for string repetition by the length of the phrase (in characters). The latter variable accounts for the repetition of long phrases.

We include a data point for each 10-second interval of the dialogue, with features reporting the lexical (LEXREP), syntactic (SYNREP) and character-based (CHARREP) repetitions up to that point in time. A time stamp and the total numbers of constituents and characters are also included (LENGTH). This way, the model may work with repetition proportions rather than the absolute counts.

We train a Support Vector Machine for regression with a radial basis function kernel ($\gamma = 5$), using the features as described above and the PATHDEV score as output.

3.10.4 Evaluation

We cast the task as a regression problem. To predict a dialogue's score, we apply the SVM to its data points. The mean outcome is the estimated score.

A suitable evaluation measure, the classical R^2, indicates the proportion of the variance in the actual task success score that can be predicted by the model. All results reported here are produced from 10-fold cross-validated 90% training / 10% test splits of the dialogues. No full dialogue was included in both test and training sets.

Task 1 was evaluated with all data. The Task 2 model was trained and tested on data points sampled from the first 5 minutes of the dialogue.

For Task 1 (full dialogues), the results (Table 3.4) indicate that ALL repetition features together with the LENGTH of the conversation, account for about 17% of the total score variance. The repetition features improve on the performance achieved from dialogue length alone (about 9%).

For the more difficult Task 2, ALL features together achieve 14% of the variance. (Note that LENGTH is not available.) When the syntactic repetition feature is taken out and only lexical (LEXREP) and character repetition (CHARREP) are used, we achieve 6% in explained variance.

The baseline is implemented as a model that always estimates the mean score. It should, theoretically, be close to 0.

	Task 1	Task 2
SYNREP, CHARREP and LENGTH	**0.17**	**0.14**
ALL w/o SYNREP	0.15	0.06
ALL w/o LEX/CHARREP	0.09	0.07
LENGTH ONLY	0.09	n/a
Baseline	0.01	0.01

Table 3.4: Portion of variance explained (R^2)

3.10.5 Discussion

Obviously, linguistic information alone does not explain the majority of the task-solving abilities. Apart from subject-related factors, communicative strategies also play a role.

That said, linguistic repetition serves as a good predictor of how well interlocutors will complete their joint task. The features used are relatively simple: provided there is some syntactic annotation, rule repetition can easily be detected. Even without syntactic information, lexical repetition already goes a long way.

The application-oriented results strengthen our initial hypothesis of the link between the tendency to repeat structural choices in language production and the success of the communicative process as a whole. They do not point at a link of short-term priming effects and task success.

Especially for Task 2, syntactic repetition made a substantial individual contribution to the performance of the model. This is compatible with a view that sees a predisposition in speakers to adapt to one another more or less, and that this adaptation ultimately leads to task success. Such adaptation is visible early on in the dialogues, obviously more so than lexical adaptation.

3.11 Conclusion

Task success in human-human dialogue is predictable—the more successfully speakers collaborate, the more they show linguistic adaptation. This confirms our IAM hypothesis. In the applied model, knowledge of lexical and syntactic repetition helps to determine task success even after just a few minutes of the conversation.

We suggested two application-oriented tasks (estimating and predicting task success) and an approach to address them. They now provide an opportunity to explore and exploit other linguistic and extra-linguistic parameters.

The primary contribution is a closer inspection of structural repetition, which showed that it is long-term adaptation that varies with task success, while short-term priming appears largely autonomous. Long-term adaptation may thus be a strategy that aids dialogue partners in aligning their language and their situation models.

While long-term adaptation is correlated with task success, we have also shown that short-term priming is not. This is not necessarily surprising from the point of view of IAM, given that priming decays so rapidly that it can hardly influence referential expressions and other relevant communicative means, which occur too infrequently to be affected by an effect lasting just five seconds.

The fact that short-term priming and long-term adaptation differ qualitatively is relevant from an architectural viewpoint. It suggests that there is more than one cognitive basis for these repetition effects: if there was only one, we would expect short-term priming and long-term adaptation to co-vary with variables such as task success. We consider this issue again in Chapter 5.

With the task success correlation in mind, we can also re-evaluate the results obtained in Experiments 3 and 5. If short-term priming does not influence task success, why would there be more short-term priming in task-oriented dialogue than in spontaneous conversation? The explanation we offer depends on the more intense semantic processing activity we can expect to find in task-oriented dialogue. In the Map Task experiments, listeners processed actively what was being said, because the task demanded just that. In the conversations recorded in the Switchboard corpus, interlocutors were not required to remember or process much of the content discussed. We propose a mechanism for short-term priming (see Chapter 5) that depends on spreading activation of lexical (and thus also semantic) material. We suggest that more intense semantic processing leads to more lexical material being retained in working memory, spreading activation to associated syntactic structures. This is what causes strong priming in task-oriented dialogue, and presumably quite generally in "engaged" dialogue.

Chapter 4

Priming as Evidence for Syntactic Structure

4.1 Introduction

When humans speak or write, they convert conceptual representations of the message to be conveyed into sequences of sounds or letters. This task of *language production* is often analyzed in terms of a processing chain which includes conceptualization, formulation, and articulation (Levelt, 1989). The conceptualization module selects concepts to express, and the formulation module decides how to express them. Formulation involves determining the lexical, syntactic, and semantic representation of the utterance. Syntax determines the systematic relationship between meaning and form of an utterance, without which language could not be produced.

Given the central role of syntax in language production, it is not surprising that a significant amount of recent research has tried to establish the exact nature of the syntactic representations that underlie the production process. As syntactic structures cannot be observed directly, a number of indirect ways have been developed to investigate them. An important one is the study of *structural priming*, which is the preference of the language processor to re-use previous syntactic choices. As an example, consider the verb *give*, which can occur in either a prepositional object (PO) construction (see (1-a)) or in a double object (DO) construction (see (1-b)):

(1) a. The policeman gives a gun to the magician.
    b. The policeman gives the magician a gun.

As described in Chapter 1, experimental results (e.g., Bock 1986b) show that participants who have a choice between producing the DO and the PO construction (e.g., in a picture naming task) are more likely to choose that construction which they (or their interlocutor) have produced previously.

Priming results such as this one give us a handle on syntactic representations: priming is only expected between constructions that share the same representation, therefore the presence or absence of priming can be used as a diagnostic for whether two constructions involve identical representations or not. Using examples such as (1), it has been argued that priming takes place on the level of syntactic rules (though this can also be interpreted as priming of sequences, as discussed below). There is also evidence for the priming of attachment decisions (Scheepers, 2003), and for the priming of sequences of constituents (Scheepers and Corley, 2000).

Using priming effects to inform syntactic theory is a relatively novel idea, especially in combination with data-oriented methodology. Previous corpus-based priming studies have considered only uncontroversial classes of constructions (e.g., passive/active). Our contribution is to overcome this limitation by defining a computational model of priming with a clear interface to a particular syntactic framework. The general assumption we make is that priming is a phenomenon relating to grammatical constituents—these constituents determine the syntactic choices whose repetition can lead to priming. Crucially, grammatical frameworks differ in the grammatical constituents they assume, and therefore predict different sets of priming effects.

We require the following ingredients to pursue our approach: a syntactic theory that identifies a set of constituents, a corpus of linguistic data annotated according to that syntactic theory, and a statistical model that estimates the strength of priming based on a set of external factors. We can then derive predictions for the influence of these factors from the syntactic theory, and test them using the statistical model.

Chapter 2 reinforced the structure-based view of priming. We have demonstrated that priming can occur for arbitrary syntactic rules in a large corpus of spoken dialogues. This is an important generalization of results from experimental work, which has only investigated priming of syntactic alternatives (such as (1) above), not for arbitrary rules.

In this Chapter, we first use the same regression models to quantify structural priming effects that have been applied to dialogue in Chapter 2. We verify predictions made by Combinatory Categorial Grammar (CCG, Steedman (2000)), a syntactic framework that has the theoretical potential to elegantly explain some of the phenomena discovered in other priming experiments.

Structure-based priming has been challenged by Chang et al. (2006), who propose a Simple Recurrent Network model that captures priming in the repetition of sequences of abstract lexical types, such as parts of speech. In this model, syntactic priming does not involve syntactic rules, but is explained simply as the learning of lexical or sub-lexical sequences.

Such a decidedly non-structural account of syntax would not just explain the phrase-structure based, but also the CCG analysis of syntactic structure, and the process by which it is generated. We therefore take a step back from CCG to investigate the question: do arbitrary sequences of lexical types prime?

Our corpus data make it possible to directly compare the structure-based and the sequence-based view of priming. The key idea is to compare priming effects for *constituents* (i.e., linguistic units generated by structural syntactic rules) with priming effects for *distituents* (i.e., sequences of parts of speech that cannot form a linguistic unit, even in the wide range of phrase structure derivations we find in a large corpus). Only under the sequence-based account of syntax do we predict the equal priming of distituents.

This chapter is organized as follows. In Section 4.2, we start with a broad question: Does priming apply to structure as opposed to transitions from one syntactic surface category to the next one? The experiments in Sections 4.3 through 4.5 test just this: that is, whether short-term and also long-term priming effects are sensitive to structural boundaries. We follow up with an investigation of the type of structure that can account for priming. In Section 4.6, we explain the relationship between structural priming and CCG, which leads to a set of specific predictions, detailed in Section 4.7. The data used for the following experiments are presented in Section 4.8. In the following, we test predictions arising from CCG. Sections 4.9 and 4.10 examine these. Sections 4.11 and 4.12 provide a discussion of the implications of the findings.[1]

[1]This chapter contains material first published in Reitter and Keller (2007) (Distituents) and in Reitter et al. (2006a) (CCG). The author is indebted to Julia Hockenmaier for her comments and also her assistance with producing and describing the CCG-based version of Switchboard.

4.1.1 Structural priming

Previous studies of structural priming (Bock, 1986b; Branigan et al., 2000a) have made few theoretical assumptions about syntax, regardless of whether the studies were based on planned experiments or corpora. They leverage the fact that alternations such as *He gave Anna the car keys* vs. *He gave the car keys to Anna* are nearly equivalent in semantics, but differ in their syntactic structure (double object vs. prepositional object). In such experiments, subjects are first exposed to a *prime*, i.e., they have to comprehend or produce either the double object or the prepositional object structure. In the subsequent trial, the *target*, they are the free to produce or comprehend either of the two structures, but they tend to prefer the one that has been primed. In corpus studies, the frequencies of the alternative constructions can be compared in a similar fashion (Gries, 2005; Szmrecsanyi, 2005).

Traditionally, syntactic priming has been explained in terms of the activation of structural representations in the language production system (Bock, 1986b; Branigan et al., 1999). In order to generate an utterance, a syntactic structure of this utterance has to be built, and this process involves the activation of syntactic frames, such as the double object frame of the verb *give* in (1-b). This activation decays over time, and when the production system has to generate another utterance, it is more likely to utilize a syntactic frame that has been pre-activated, i.e., that has been used in the recent past. This then leads to the priming effect, e.g., in the case of (1-b), the production system is more likely to generate another double object construction (rather than the alternative prepositional object construction in (1-a)).

In Chapter 2, we presented a different method to examine priming effects in the general case. Rather than selecting specific syntactic alternations, general syntactic units are identified. This method detects syntactic repetition in corpora and correlates its probability with the distance between prime and target, where at great distance, any repetition can be attributed to chance. The size of the priming effect is then estimated as the difference between the repetition probability close to the prime and far away from the prime. This is a way of factoring out chance repetition (which is required if we do not deal with syntactic alternations). By relying on syntactic units, the priming model includes implicit assumptions about the particular syntactic framework used to annotate the corpus under investigation.

A key characteristic of syntactic priming is the *lexical boost*, i.e., the fact that priming is enhanced if the prime and the target share lexical material (such as the verb) in addition to sharing syntactic structure (such as the DO frame). This effect can be explained fairly naturally by the activation-based view of priming: the more aspects of the representations of the prime and the target are shared, the more residual activation from the prime biases the syntactic choices made during the target elicitation, leading to an increased priming effect.

The exact nature of the syntactic representations (syntactic frames, etc.) that underlie priming has been the subject of some debate. Recently, a series of corpus studies have provided evidence for *syntactic structure* as the correct level of representation. These studies provide corpus evidence for the repetition of syntactic rules in corpus data consistent with experimental results on syntactic priming. This includes evidence for the priming of specific constructions (Gries, 2005; Szmrecsanyi, 2005; Dubey et al., 2005) as well as evidence for a generalized priming effect that applies to arbitrary rules (Chapter 2) and does not have to involve the alternation of semantically equivalent syntactic realizations (as in example (1)).

These corpus studies also constitute important corroborating evidence for the activation-based view, as they replicate the central characteristics of the experimental results on priming, including the rapid, exponential decay of the effect and the increased priming if head words are repeated (lexical boost) (Bock, 1986b; Branigan et al., 1999).

4.1.2 Syntactic theories and cognitive reality

Syntactic priming can be demonstrated for the structures that are assumed to underlie syntactic units. A classical view of syntactic analysis assigns a tree structure, with larger constituents subsuming smaller ones (in sub-trees), where each branch is licensed by a syntactic rule. Such rules are subject to priming, as shown previously.

This framework of syntax is not likely to be an accurate reflection of the cognitive units of language production and comprehension. The phrase structure framework lacks in abstraction and encodes constraints in the language-specific grammar—constraints that apply to all languages, and the human language facility in general. Still, phrase structure rules are a necessary, reasonable and widely agreed simplification, serving as a suitable starting point for a generalization of priming (and processing) models.

We see a syntactic rule as a unit that represents a (set of) nodes which are acti-

A: And$_{CC}$ all$_{DT}$ of$_{IN}$ a$_{DT}$ sudden$_{JJ}$ he$_{PRP}$'s$_{HVS}$ got$_{VBN}$ a$_{DT}$ hang$_{NN}$ glider$_{NN}$

B: I$_{PRP}$ do$_{VBP}$ n't$_{RB}$ even$_{RB}$ heard$_{VBN}$ of$_{IN}$ that$_{DT}$ show$_{NN}$

A: You$_{PRP}$ have$_{VBP}$ n't$_{RB}$

B: It$_{PRP}$'s$_{BES}$ called$_{VBN}$ McGyver$_{NNP}$?

A: He$_{PRP}$'s$_{BES}$ like$_{UH}$ a$_{DT}$ semigovernment$_{JJ}$ type$_{NN}$ agent$_{NN}$ who$_{WP}$ the$_{DT}$ Phoenix$_{NNP}$ Foundation$_{NNP}$ supposedly$_{RB}$ ...

Figure 4.1: Excerpt from the tagged Switchboard data.

vated during the processing of a sentence. We assume a partially shared apparatus for comprehension and production, in which previously activated nodes retain some of their activation and thus, are preferentially activated for a limited time period. Lower activation thresholds may be less effective when activation can be reached more easily, as in the case of high-frequency links. Linearization and hierarchical syntactic production are, in our model, closely related rather than separate.

4.2 Is syntactic priming at all structural?

Syntactic structure does not necessarily imply the presence of phrase structure rules as used in the previous experiments. A recent model of sentence production has suggested priming operates on sequences of abstract lexical categories rather than on rules (Chang et al., 2006). Under this view, well-established structural priming effects could be seen as epiphenomena, resulting from the priming of pre-lexical sequences such as parts of speech (POS).

Many known priming effects can be explained in this way, e.g., the fact that (2-a) primes (2-c) could be due to the shared POS sequence NN IN in both sentences.

(2) a. The$_{DT}$ doctor$_{NN}$ gives$_{VBZ}$ some$_{DT}$ flowers$_{NN}$ to$_{IN}$ his$_{PRP\$}$ girl$_{NN}$
    b. The$_{DT}$ doctor$_{NN}$ gives$_{VBZ}$ his$_{PRP\$}$ girl$_{NN}$ some$_{DT}$ flowers$_{NN}$
    c. The$_{DT}$ policeman$_{NN}$ gives$_{VBZ}$ a$_{DT}$ gun$_{NN}$ to$_{IN}$ the$_{DT}$ magician$_{NN}$
    d. The$_{DT}$ policeman$_{NN}$ gives$_{VBZ}$ the$_{DT}$ magician$_{NN}$ a$_{DT}$ gun$_{NN}$

CC	coordinating conjunction (and, or)
DT	singular determiner/quantifier (this, that)
IN	preposition
JJ	adjective
MD	modal auxiliary (can, should, will)
NN	singular or mass noun
NNS	plural and/or possessive noun
PRP	personal pronoun
RB	adverb
UH	hesitation
VBZ	verb, 3rd. singular present
VBP	verb, present tense, other than 3rd singular
WDT	wh- determiner (what, which)
WP$	possessive wh- pronoun (whose)
WRB	wh- adverb (how, where, when)

Table 4.1: Common Brown/Switchboard part-of-speech tags.

Sentence (2-b), on the other hand, contains a different POS sequence (NN DT NN) and therefore is expected to prime (2-d), but not (2-c), consistent with experimental results on the priming of prepositional object and double object constructions. (See Table 4.1 for a subset of the part-of-speech categories used in this study, and Figure 4.1 for an excerpt from the corpus.)

Another example of priming that could be explained by sequence priming is Bock and Loebell's (1990) result, showing that a sentence with a locative prepositional argument, e.g., *The 747 was landing by the airport's control tower*, primes a passive sentence such as *The man is being stung by a bee*.

The sequencing view of priming is central to Chang et al.'s (2006) Dual-path Model, a connectionist model of sentence production that aims to account for results from both language acquisition and syntactic priming. At the core of the Dual-path Model there are two mechanisms. The first one is the *Sequencing System*, consisting of a Simple Recurrent Network (SRN, Elman, 1990) which generates sequences of words, compressing them to abstract parts of speech (POS) categories. As is common for SRNs, language production is essentially the task of predicting the next word given its left context, and an error-driven learning algorithm is used to train the model. The second mechanism in the Dual-path Model is the *Meaning System*, which maps meaning representations to words and vice verse. These representations consist of *what*- and *cwhat*-nodes (representing the lexical semantics of words in production and comprehension, respectively) and *where*- and *cwhere*-nodes (representing words' semantic roles in production and comprehension). Figure 4.2 gives a schematic view of the Dual-path Model. Note that the model contains a self-monitoring loop which connects the currently produced word with the comprehended version of the previously produced word (*cword* in the diagram). For the following investigation, we will concentrate on key ideas behind the Sequencing System. Note that rather than testing the model explicitly (it does not reach corpus coverage), we discuss its key idea of sequence priming.

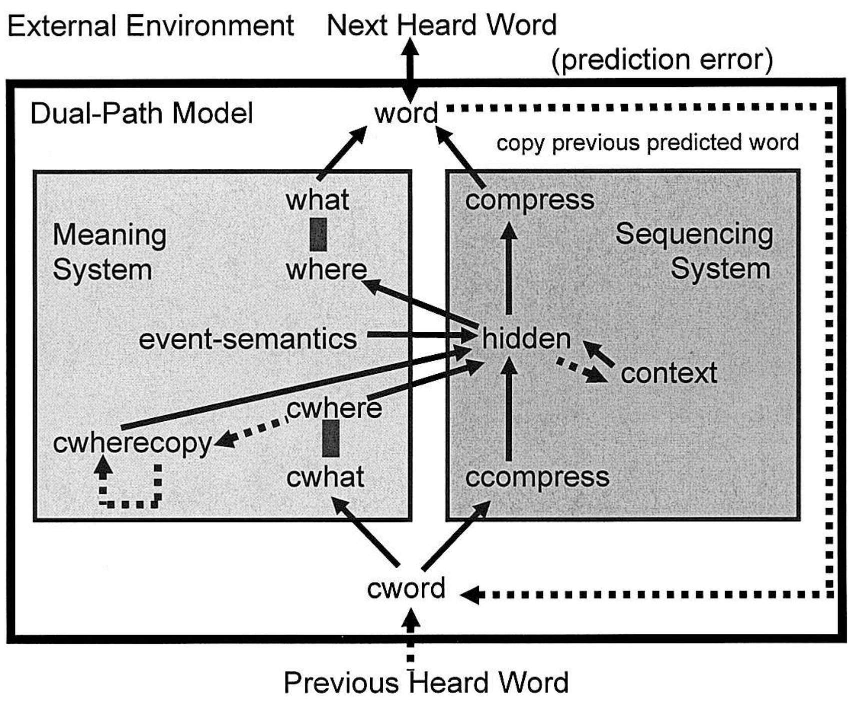


Figure 4.2: Schematic view of the Dual-path Model (figure from Chang et al. 2006)

The Dual-path Model accounts for a range of structural priming results, as well as for certain findings in the language acquisition literature (preferential looking and elicited production studies). The model makes two key assumptions: (1) language comprehension and production is based on abstract lexical (POS) sequences as the underlying representation, no hierarchical syntactic structures (and no syntactic rules) are explicitly modeled; (2) the same implicit learning processes underlie language production and acquisition, which means that short-term priming (which decays in a matter of seconds) and long-term priming (which can take days to decay) are accounted for by the same mechanism, i.e., an SRN trained to predict POS sequences. The first assumption is addressed in the following.

To make the distinction between the two accounts of syntactic priming clear, we use the term *structural* to refer to a view of syntax in which grammaticality of a word occurring at a position i is determined by more than the part of speech at position $i-1$. Practically, this means that syntax is governed by hierarchical dependencies or graphs.[2]

In order to distinguish structural from sequential priming, we use the notion of *distituents* from the grammar induction literature (e.g., Magerman and Marcus, 1990; Kuhn, 2004). Distituents are pairs of POS tags that cannot form a syntactic unit. All other pairs, i.e., the ones that occur in a syntactic unit, are deemed *constituent*. Crucially, such pairs are predicted to show decaying repetition due to priming under both assumptions, structural and sequential priming. *Distituent* pairs, however, will show an equal amount of priming only if sentence production is sequentially biased. Under the structure-based view, there should be less distituent priming, as distituents (by definition) cannot be generated by syntactic rules.

To define distituents more precisely, we refer to the POS categories and the tree-structured syntactic analysis of each sentence.

For each word, we extracted its terminal syntactic category, the *part-of-speech* (POS) tag, that is, "finite verb", or "determiner", "preposition", "common noun", and the like. Among the arbitrary sequences of word-bigrams extracted this way, we can distinguish sequences that also represent constituents. The syntax tree then defines *constituents* or subtrees. For example, in the syntax tree in (3), *the policeman*, among other phrases, forms a constituent.

(3) [The$_{DT}$ policeman$_{NN}$ [shows$_{VBZ}$ [$_i$ the$_{DT}$ girl$_{NN}$]
[$_j$ his$_{PRP\$}$ gun$_{NN}$]]]

[2] Note that, in the psycholinguistic literature, the term *structural priming* often refers to priming of more properties than just syntactic ones, i.e. priming of morphosyntactic properties, or thematic roles.

Some of the POS sequences cross constituent boundaries in a particular instance and never represent constituents elsewhere. This leads us to the definition of a *distituent*.

Distituent: A distituent is a POS pair that cannot be adjacent without crossing at least one constituent boundary. For example, NN PRP$ (noun, possessive pronoun) is a distituent in English, because there can be no constituent that directly combines a noun followed by a possessive pronoun. Of course, such a POS sequence will occur in the data as in (3), but for a distituent bigram, the two POS tags will always belong to at least two *different* constituents (in the above case two argument noun phrases *i*, *j*). To give another example, DT NN is not a distituent, because the determiner and the noun directly form a noun phrase. NN VBZ is not a distituent either: while it does cross constituent boundaries in (3), it appears without doing so (in its own constituent) in a verbal phrase with an intransitive verb elsewhere in the corpus (*before [school starts]*, common nouns are annotated as NN in the tagset used). (In cases of annotation errors, the conservative definition of distituents may result in an overly restrictive selection.)

Table 4.2 lists the most frequent distituents. An equivalent definition of distituency refers to dominance in the syntax tree.

Distituent (alternative): Two adjacent word tokens α, β are distituent if and only if for all adjacent word tokens α', β' in the corpus, whose POS tags are the same as those of α, β, there is no syntax node N such that N directly dominates α' and β'.

By making the qualification that there may be no non-distituent POS pair of the same type in the corpus, we adopt a rather strict notion of distituents.

If a syntactically annotated corpus is available, then the syntactic annotation can be used to identify distituents in the data as follows: for every sequence of two adjacent parts of speech (bigram) in the corpus, we determine whether it occurs inside a constituent without crossing constituent boundaries anywhere in the corpus. If and only if this is the case do we regard this sequence as a distituent. Note that distituents (contrary to constituents) do not have a hierarchical structure—they should be regarded simply as bigrams that cross constituent boundaries.

The distinction of distituent and constituent bigrams enabls us to contrast the two priming models. Under a sequence priming account, we would expect that all sequences of lexical abstracts would show approximately equal priming, regardless of their structural properties. Sequences that cross constituent boundaries should, if anything, show more priming, given that such sequences tend to be rare and priming effects are stronger for low-frequency items. Under a structural priming account, in which linguistic decisions are subject to priming, we would expect to see less priming for sequences crossing constituent boundaries.

Freq.	POS bigram	example	Freq.	POS bigram	example
1180	JJ TO	good to	45965	PRP VBP	you know
630	VBN TO	supposed to	27531	DT NN	a lot
392	NNS TO	people to	20247	IN DT	in the
335	NNS MD	people will	15517	NN IN	lot of
227	NNS VBZ	years is	14578	PRP VBD	they did
151	WRB TO	how to	14163	IN PRP	of it
151	NN PRP$	fact my	11920	JJ NN	little bit
127	VB PDT	have all	11629	CC PRP	and me
120	JJ VBZ	old is	11313	TO VB	to be
108	EX MD	there would	10309	DT JJ	each other
...			...		
2	WDT VBN	whatever needed	7	RP RBR	out more
2	NN RBS	country most	5	PRP NNPS	them Giants

Table 4.2: The most common *distituent* (left) and *constituent* (right) POS bigrams from the corpus as well as some uncommon ones.

4.3 Experiment 9: Short-term bigram priming

If Chang et al.'s (2006) sequencing view of priming is correct, then there should be no systematic difference between constituents and distituents. Therefore, his model predicts that in corpus data, we should find priming for both constituents and for distituents. On the other hand, if the rule-based view is correct, then priming should be confined to constituents, as distituents cannot be generated by syntactic rules, and therefore cannot be subject to priming. The present experiment tests these two alternative hypotheses for short-term priming, i.e., for structural or non-structural repetition that decays rapidly.

4.3.1 Method

4.3.1.1 Data

Distituents were identified in the Switchboard corpus following the definition given in the previous section. Bigrams including hesitations such as *like* and *uh*, or with POS tags not identified by the original annotation (marked XX), were excluded. This way, we extracted 378 different types of POS bigrams, 80 of which were distituents. See Table 4.2 for common distituent and constituent bigrams. Data points with rare POS bigrams (frequency $f \leq 10$) and unknown POS tags were discarded.

4.3.1.2 Statistical analysis

To analyze priming effects in our corpus data, we examine the repetition of POS bigrams. Whenever a POS bigram is repeated within a short time period more often than we would expect from chance repetition, we accept it as an example of structural priming.

As discussed before, short-term priming is subject to a swift decay. The increase in repetition probability is seen shortly after the stimulus, but less so a few seconds later. Therefore, we use the time elapsed after a stimulus to predict whether repetition will occur. An exploratory data analysis is shown in Figure 4.3, where the repetition probability is shown for various lags between prime and target bigram, from 1 to 14 seconds. Intuitively, the decay is less obvious for distituents than for constituents, but the following statistical analysis proves that this is indeed the case.

The statistical methodology follows the techniques developed to estimate rule-based priming levels. A logistic regression model was used to compute a correlation coefficient between repetition and the temporal distance d (as covariate $\ln(\textsc{Dist})$)[3].

For each occurrence of a POS bigram (*target*) at a time t, we examine the POS bigrams in the one-second time period $[t-d-0.5, t-d+0.5]$. If the POS bigram

[3]This $\ln(\textsc{Dist})$ (Distance) covariate should not be confused with the covariate distinguishing distituent from constituent bigrams, DISTITUENT.

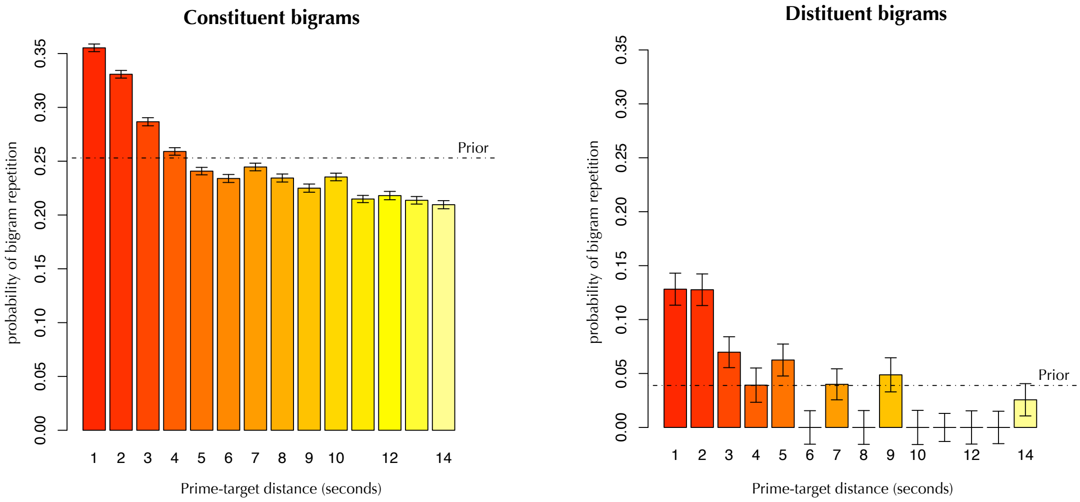


Figure 4.3: Constituent POS bigrams show a rapid decay within the first few seconds after a prime, while distituent bigrams do not seem to show a similar decay. 95% confidence intervals obtained by bootstrapping. Prior is calculated as mean repetition over all prime-target pairs.

re-occurs, we count the target occurrence as *repeated*, otherwise as control case. This is the predicted (dependent) binary variable. The models are designed to estimate the probability of repetition.

If there is no structural priming effect (null hypothesis), we would expect there to be no relationship between predicted repetition probability and ln(DIST). An interaction between this effect and the factor distinguishing distituent from constituent bigrams (DISTITUENT) would reveal differences in priming strength between constituents and distituents.

To account for frequency effects in priming as they have been reported previously, we include the normalized bigram frequency as a covariate ln(FREQ). A further factor PRIMETYPE distinguishes priming between speakers (comprehension-production priming, CP) from priming within a speaker (production-production priming, PP): only in the latter case were prime and target uttered by the same speaker.

To implement this logistic regression model, we use generalized linear mixed models with a logit link and random variables grouping bigrams from each utterance to reflect potential non-independence. Apart from the use of bigrams rather than syntactic rules, the methodology follows the one described in Chapter 2.

The dataset was re-sampled for balance with respect to the response variable in the respective experiment.[4] Interactions (and main effects) were removed where appropriate, i.e., where there was no significant coefficient and no dependent interaction.

4.3.2 Results

The results show a reliable main effect for $\ln(\textsc{Dist})$ ($\beta = -0.074$, $p < 0.0001$), indicating a baseline priming effect. The model also showed a reliable interaction of $\ln(\textsc{Dist})$ and DISTITUENT ($\beta = 0.209$, $p < 0.05$), indicating reliably less and probably a lack of priming for distituents (the sum of the two coefficients is positive, thus showing no decay: $-0.074 + 0.209 > 0$).

$\ln(\textsc{Dist})$ also interacts reliably with $\ln(\textsc{Freq})$ ($\beta = 0.156$, $p < 0.0001$), showing that higher-frequency POS bigrams receive less priming. See Table 4.3 for the full specification of the model (after having been reduced to significant or otherwise relevant terms only).

In Figure 4.4, a contrast model is shown, which contains the same covariates and the maximal set of interactions, but showing the effect sizes under different conditions of PRIMETYPE and DISTITUENT.

4.3.3 Discussion

The main effect is consistent with previous results. The resulting model also replicates the priming, frequency, and type effects found with other corpora, and using phrase structure rules (Chapter 2), as well as frequency effects found experimentally for relative clause attachment priming (Scheepers, 2003).

With respect to the hypothesis leading to this experiment, we found not only reliably less priming for distituents: no priming effect for distituents could be found. This provides evidence against a non-structural priming account.

[4] In an experimental design, we would control and balance dependent variables rather than the response, but here, where we are interested in the fitted interactions, the model fitting is more reliable with a balanced dataset.

Covariate	β	SE	$p(>\lvert z\rvert)$
Intercept	0.414	0.015	< 0.0001 ***
ln(DIST)	-0.074	0.008	< 0.0001 ***
ln(FREQ)	0.506	0.012	< 0.0001 ***
PRIMETYPE$_{CP}$	-0.234	0.030	< 0.0001 ***
DISTITUENT	-0.737	0.157	< 0.0001 ***
ln(DIST):ln(FREQ)	0.150	0.006	< 0.0001 ***
ln(DIST):PRIMETYPE$_{CP}$	0.054	0.015	< 0.0005 ***
ln(DIST):DISTITUENT	0.209	0.083	< 0.05 *

Table 4.3: The model for the repetition of part-of-speech bigrams including a factor distinguishing distituent and constituent bigrams. Lower β coefficients for ln(DIST) indicate stronger priming.

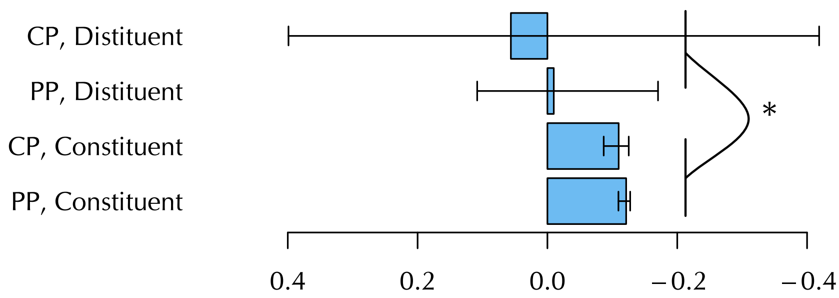


Figure 4.4: Effects of ln(DIST) in the short-term priming models. Further frequency effects (for each combination of conditions) were fitted i.e., controlled for. Effect sizes in logits. Longer bars (to the right) indicate stronger decay, hence stronger short-term priming. 95% confidence intervals via Markov-Chain Monte-Carlo sampling. Significance indications taken from the reduced model described in the text.

However, we caution against accepting these results alone as strong evidence for a lack of priming of non-structural syntactic sequences, in particular because we would have expected more reliable support for the crucial ln(DIST):DISTITUENT interaction, which indicates the relevance of the DISTITUENT factor to priming strength. We therefore seek confirmation of the results in the following experiment.

4.4 Experiment 10: Short-term bigram priming vs. path length

The previous experiment produced results that supported the initial hypothesis: distituent bigrams show less priming than constituent bigrams, if any at all. However, the robust contrastive analysis of the data showed relatively low confidence for the priming levels of distituents. The reason for this is the very restrictive definition of distituency, which left us with only few distituent bigrams (on the order of $3,000$).

The following study was designed to address this concern. Here, we will use a continuous measure to quantify the number of syntactic decisions taken in order to produce a POS bigram. For a constituent bigram, only one phrase-structure rule needs to be invoked in order to produce the bigram. To repeat a distituent bigram, more rules need to be repeated. If and only if structural decisions are primed, syntactic priming should be weaker, the more distinct decisions need to be primed. If only the transitions between POS tags are primed, syntactic priming should be indifferent to structural boundaries.

4.4.1 Method

Using the same data as in the previous experiment, we identified the *minimal path length* of each bigram. This is the length of the shortest route from the POS tag of the first word to the POS tag of the second word in each bigram in the phrase structure syntax tree. A route in this context is defined as a continuous of arcs, with arcs defined by the phrase structure tree (seen as a directed acyclic graph). A branch derived from the rule $NP \rightarrow DET\ N$ would define the arcs from DET to NP, and from NP to N. To illustrate this, consider the tree in Figure 4.5 (p. 105). The minimal path length of the bigram *I really* (POS bigram PRP-RB) is 4, corresponding to the path PRP, NP-SBJ, S, ADVP, RB. The minimal path length of the bigram *what I* is 5. The mean minimal path length in our data is 2.96. In the model descriptions,

Covariate	β	SE	$p(>\|z\|)$
Intercept	0.863	0.03	< 0.0001 ***
ln(DIST)	−0.209	0.014	< 0.0001 ***
ln(FREQ)	0.456	0.012	< 0.0001 ***
PRIMETYPE$_{CP}$	−0.062	0.031	< 0.05 *
PATHLEN	−0.120	0.009	< 0.0001 ***
ln(DIST):ln(FREQ)	0.145	0.006	< 0.0001 ***
ln(DIST):PRIMETYPE$_{CP}$	0.002	0.013	0.909
ln(FREQ):PRIMETYPE$_{CP}$	0.132	0.032	< 0.0001 ***
ln(DIST):PATHLEN	0.034	0.004	< 0.0001 ***
ln(DIST):ln(FREQ):PRIMETYPE$_{CP}$	−0.03	0.014	< 0.05 *

Table 4.4: The reduced model showing the interaction between ln(DIST) and PATHLEN. Lower β coefficients indicate stronger priming. Thus, the interaction with PATHLEN points to less priming for greater path lengths.

we denote the covariate as PATHLEN. The minimal path length is closely related to the notion of *connection path* (Lombardo and Sturt, 1999), which is the chain of syntactic nodes that must be constructed to connect a new word to its left context.

For priming to be sensitive to syntactic structure, we would expect priming to be stronger for small minimal path lengths. Longer minimal paths relate to greater sets of structural decisions involved in the production of the bigram. Distituent bigrams translate to longer minimal paths[5], so this hypothesis is consistent with the one examined in the previous experiment.

4.4.2 Results

We obtain a main effect of ln(DIST) ($\beta = -0.209$, $p < 0.0001$), showing decay, i.e., priming. Crucially, ln(DIST) interacts with PATHLEN ($\beta = 0.034$, $p < 0.0001$). This means that priming becomes weaker as the minimal path length increases.

The full model is specified in Table 4.4.

[5]To be precise, distituents have a path length greater than 2 and this also has to be the case for any other bigram with the same two POS in the corpus.

4.4.3 Discussion

Again, the results support the original hypothesis: priming is sensitive to syntactic structure. It is the structural units that show priming rather than arbitrarily chosen sequences of abstract lexical categories.

It is not surprising that repetition effects involving constituent bigrams approximate the underlying syntactic priming effect. In natural language processing, transition-based (n-gram) models are commonly used to approximate syntactic regularity. Thus, transition-based models such as White and Baldridge (2003) and Chang et al. (2006) can implement sequence learning and thereby emulate structural priming. Given the marked contrast between constituents and distituents, we argue that structure-based models provide a more convincing explanation of the effect than transition-based models. Our explanation posits structural representations at the heart of the human formulation mechanism.

There are further explanations for the effect. Distituent bigrams do not only cross syntactic boundaries. They are also more likely than constituent bigrams to cross discourse unit or clause boundaries and the transitions between semantic units. (No bigrams used in these experiments crossed utterance boundaries.) Where priming interacts with semantic processing, we would expect precisely the effect we obtain.

Alternative explanations involve prosodic units, which tend to coincide with syntactic constituents. Thus, distituent bigrams are more likely to cross prosodic boundaries. While there is, to our knowledge, no positive evidence for an interaction of priming and intionation, it remains a potential confound.

To produce (or comprehend) language, syntactic structure does not have to be retained once an utterance has been produced or understood. A processing model dealing with adaptation has to specify the units that are subject to learning or adaptation. In the present experiment, we have looked at very short-term processes. To determine whether there is an adaptation of syntactic structure beyond transition models, we investigate differences in long term priming effects for constituent and distituent bigrams.

4.5 Experiment 11: Long-term bigram priming

While short-term priming effects are strong, they also decay quickly. Adaptation is more similar to implicit learning in that it lacks this strong decay. If priming and adaptation are indeed two qualitatively different cognitive processes, then Chang's Dual-path Model may be able to account for the adaptation. Support for the model would come from data showing that learning applies to sequences rather than structures. Therefore, comparing the adaptation of constituent and distituent bigrams would shed light on this question. This is the aim of the present experiment.

4.5.1 Method

The dataset was the same as in Experiment 9.

While short-term priming can be pin-pointed using the characteristic decay, for long-term priming we need to inspect whole dialogues. As in Experiment 9, we use a binary response variable to reflect the repetition of a POS bigram. While we estimated repetition probability as a function of distance between prime and target in Experiment 9, with primes occurring in a one-second priming period at a set distance before the target, we now regard the first half of a dialogue as the priming period, testing all POS bigrams in the second half for repetition.

We contrast repetition in two conditions, which distinguish situations where priming can have taken place (SAMEDOC=1) from others (control), where repetition is only due to chance (SAMEDOC=0).

To do so, we split each dialogue into two equal halves, but exclude a 10-second portion in the middle to avoid short-term priming effects. The first half is designated as the *priming half*, the second half contains the *targets*. For each target POS bigram, we check whether it has occurred in the priming half (repetition).

For the priming condition SAMEDOC=1, we keep dialogues together: priming and target halves stem from the same original dialogue. For the non-priming control condition (SAMEDOC=0), priming and target halves are randomly chosen so that they stem from different dialogues.

We can then cast long-term adaptation as the differential between rule repetition in document halves of single dialogues, and repetition in dialogue halves sampled from different dialogues. The goal is now to establish a main effect of SAMEDOC for adaptation, and its interaction with DISTITUENT.

4.5.2 Results

The resulting model shows a number of reliable main effects and interactions. In the following, we do not only analyze significance, but also pay attention to effect sizes.

We find a reliable main effect of SAMEDOC ($\beta = -0.34$, $p < 0.0001$) and the interaction of ln(DIST) with SAMEDOC ($\beta = -0.15$, $p < 0.0001$). This indicates that at low bigram frequencies (ln(FREQ) < -2.27), repetition of constituents is greater in priming dialogues than in the control. Thus, we find positive adaptation of *constituent* bigrams.

Further, the model shows a reliable interaction of DISTITUENT with SAMEDOC ($\beta = -0.38$, $p < 0.05$) and with SAMEDOC:ln(FREQ) (triple interaction). This means that at similarly low bigram frequencies (ln(FREQ) < -2.56), again repetition of distituents is greater in priming dialogues than in the control. Thus, we find positive adaptation of *distituent* bigrams.

Centered and transformed bigram frequencies range from -6.67 to 1.50 and average at μ(ln(FREQ)) $= -0.81$, standard deviation σ(ln(FREQ)) $= 1.48$, with the lower quartile at -1.7160. The above adaptation effects apply to the 13% of bigrams with the lowest frequencies.[6]

The model shows positive adaptation for low-frequency bigrams, both in the cases of constituents and distituents. This evidence is supported further by a simplified model, where the triple interaction involving the POS frequency is removed. In this simplified model, no reliable interaction effect of DISTITUENT and SAMEDOC can be found ($p = 0.38$).

We conclude that there is no evidence for a difference in long-term adaptivity between constituents and distituents.

4.5.3 Discussion

Short-term priming, decaying within a few seconds, and long-term adaptation lasting minutes and in some cases even days, differ substantially (see V. Ferreira, 2006). Our data show both kinds of repetition effects. However, syntactic structure mattered only for short-term processing effects: long-term adaptation may well operate

[6]Further coefficients were fitted which are irrelevant to our purposes because they describe effects on chance repetition: ln(FREQ) ($\beta = 1.73$, $p < 0.0001$), DISTITUENT ($\beta = -1.02$, $p < 0.0001$), ln(FREQ):DISTITUENT ($\beta = -0.45$, $p < 0.0001$).

on abstract lexical sequences rather than syntactic structure.

A model where sequences of parts of speech, or lexemes, are memorized as procedures would explain the findings. Effectively, this likens long-term adaptation to a procedural memory effect. Stored procedures can certainly help speakers to produce and listeners to understand language. They may support alignment effects in dialogue (Pickering and Garrod, 2004). Moreover, they are consistent with Chang et al.'s (2006) model. So while we argue against the sequence- or transition-based account for priming, we believe it to be plausible for long-term adaptation processes.

The syntactic processor in Chang et al.'s (2006) model has a few theoretical shortcomings. First, it revives the notion of language as a Markov process rather than a system of hierarchical rules forming syntactic dependencies. Markov processes have been criticized as inadequate for modeling natural language syntax by a number of authors, ranging from Chomsky (1957) to Steedman (1999). Second, while Chang et al.'s (2006) model is able to explain a range of effects found in experimental studies, it has not been evaluated on naturalistic data such as those drawn from corpora. This contrasts with the wealth of corpus evidence for rule-based priming.

Our finding is not to be taken as an argument against Simple Recurrent Networks, in general, as a model of syntactic learning and processing. If a hierarchy of multiple layers of transitions is acquired, the hierarchy implements a notion of structural constituents that would conform with our results (cf. Elman, 1990).

From an applied point of view, approximating priming using POS sequences can still be useful for practical applications. N-gram models are wide-spread in computational linguistics (e.g., Kuhn and Mori, 1990; Brown et al., 1992) and have been shown to cover a broad variety of linguistic data. Adapting probabilities in these models to emulate priming has been demonstrated in the domain of human-computer dialogue (White and Baldridge, 2003).

4.6 Priming and lexicalized grammar

As a preliminary conclusion, we can state that at least at the immediate syntactic level, i.e., where short-term effects matter, there is a clear relevance of *structure*. Syntactic processing is structural. But what is the nature of this structure? In the following, we investigate the psycholinguistic reality of one recent account of syntactic structure using priming effects.

Previous work has demonstrated that priming effects on different linguistic levels are not independent (Pickering and Branigan, 1998). Lexical repetition makes repetition on the syntactic level more likely. For instance, suppose we have two verbal phrases (prime, target) produced only a few seconds apart. Priming means that the target is more likely to assume the same syntactic form (e.g., a passive) as the prime. Furthermore, if the head verbs in prime and target are identical, experiments have demonstrated a stronger priming effect. This effect seems to indicate that lexical and syntactic representations in the grammar share some information (e.g., subcategorization information), and therefore these representations can prime each other.

Consequently, we treat subcategorization as co-terminous with syntactic type, rather than as a feature exclusively associated with lexemes. Such types determine the context of a lexeme or phrase, and are determined by derivation. Such an analysis is exactly what categorial grammars suggest. The rich set of syntactic types that categories afford may be just sufficient to describe all and only the units that can show priming effects during syntactic processing. That is to say that syntactic priming is categorial-type priming, rather than priming of phrase structure rules.

Consistent with this view, Pickering and Branigan (1998) assume that morphosyntactic features such as tense, aspect or number are represented independently from combinatorial properties which specify the contextual requirements of a lexical item. Property groups are represented centrally and shared between lexicon entries, so that they may—separately—prime each other. For example, the pre-nominal adjective *red* in *the red book* primes other pre-nominal adjectives, but not a post-nominal relative clause (*the book that's red*) (Cleland and Pickering, 2003; Scheepers, 2003).

However, if a lexical item can prime a phrasal constituent of the same type, and vice verse, then a type-driven grammar formalism like CCG can provide a simple account of the effect. In CCG, lexical and derived syntactic types have the same combinatorial potential, which is completely specified by the type. In structure-driven theories, this information is only implicitly given in the derivational process.

4.6.1 Combinatory Categorial Grammar

CCG (Steedman, 2000) is a mildly context-sensitive, lexicalized grammar formalism with a transparent syntax-semantics interface and a flexible constituent structure that is of particular interest to psycholinguistics, because it allows for the construction of incremental derivations. CCG has also enjoyed the interest of the NLP community, with high-accuracy wide-coverage parsers (Clark and Curran, 2004; Hockenmaier and Steedman, 2002) and generators (White and Baldridge, 2003) available.

In CCG, words are associated with lexical categories which specify their subcategorization behaviour, e.g., $((\mathsf{S[dcl]\backslash NP})/\mathsf{NP})/\mathsf{NP}$ is the lexical category for (tensed) ditransitive verbs in English such as *gives* or *send*. Such verbs expect two NP objects to their right, and one NP subject to their left. Complex categories $\mathsf{X/Y}$ or $\mathsf{X\backslash Y}$ are functions which yield a constituent with category X if applied to a constituent with category Y to their right (/Y) or to their left (\Y).

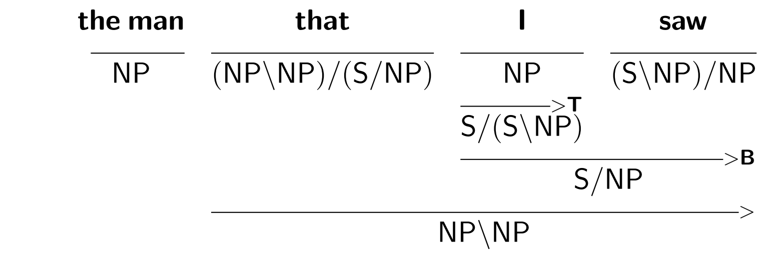
Constituents are combined via a small set of combinatorial rule schemata:

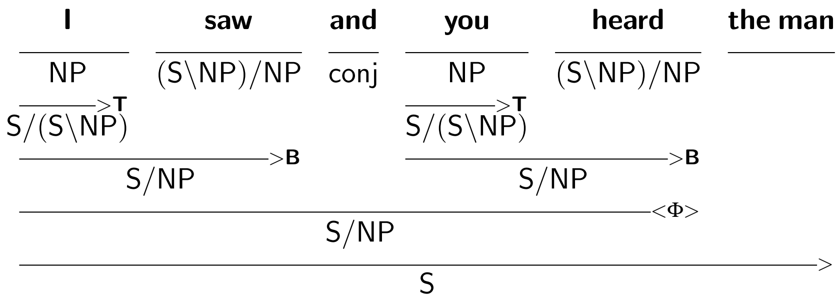
Forward Application:	$\mathsf{X/Y}$	Y	$\Rightarrow_{>}$	X
Backward Application:	Y	$\mathsf{X\backslash Y}$	$\Rightarrow_{>}$	X
Forward Composition:	$\mathsf{X/Y}$	$\mathsf{Y/Z}$	$\Rightarrow_{\mathbf{B}}$	$\mathsf{X/Z}$
Backward Composition:	$\mathsf{Y\backslash Z}$	$\mathsf{X\backslash Y}$	$\Rightarrow_{\mathbf{B}}$	$\mathsf{X\backslash Z}$
Backward Crossed Composition:	$\mathsf{Y/Z}$	$\mathsf{X\backslash Y}$	$\Rightarrow_{\mathbf{B_X}}$	$\mathsf{X/Z}$
Forward Type-raising:	X		$\Rightarrow_{\mathbf{T}}$	$\mathsf{T/(T\backslash X)}$
Coordination:	X conj	X	$\Rightarrow_{\mathbf{\Phi}}$	X

Function application is the most basic operation (and used by all variants of categorial grammar):

$$
\begin{array}{ccc}
\textbf{I} & \textbf{saw} & \textbf{the man} \\
\hline
\mathsf{NP} & \mathsf{(S\backslash NP)/NP} & \mathsf{NP} \\
\cline{2-3}
 & \multicolumn{2}{c}{\mathsf{S\backslash NP}\ \ >} \\
\hline
\multicolumn{3}{c}{\mathsf{S}\ \ <}
\end{array}
$$

Composition (**B**) and type-raising (**T**) are necessary for the analysis of long-range dependencies and for incremental derivations. CCG uses the same lexical categories for long-range dependencies that arise for example in wh-movement or coordination as for local dependencies, and does not require traces:





The combinatorial rules of CCG allow multiple semantically equivalent syntactic derivations of the same sentence. This *spurious ambiguity* is the result of CCG's flexible constituent structure, which can account for long-range dependencies and coordination (as in the above example), and also for interaction with information structure.

CCG parsers often limit the use of the combinatorial rules (in particular: type-raising) to obtain a single right-branching *normal-form* derivation (Eisner, 1996) for each possible semantic interpretation. Such normal-form derivations only use composition and type-raising where syntactically necessary (e.g., in relative clauses). Figures 4.6 and 4.7 (p. 106) show a case of multiple, semantically equivalent analyses, with two derivations: once as a normal-form, and then in a maximally incremental variant.

CCG is distinguished from most other grammatical theories by the fact that its rules are *type-dependent*, rather than structure-dependent like classical transformations. Such rules adhere strictly to the constituent condition on rules, i.e., they apply to and yield constituents. Moreover, the syntactic types that determine the applicability of rules in derivations are transparent to (i.e., are determined, though not necessarily uniquely, by) the semantic types that they are associated with. As a consequence, syntactic types are more expressive and more numerous than standard parts of speech: there are around 500 highly frequent CCG types, as compard to the standard 50 or so Penn Treebank POS tags. As we will see below, these properties allow CCG to discard a number of traditional assumptions concerning surface constituency. The CCG types also allow us to make a number of testable predictions concerning priming effects, most importantly: (a) that priming effects are type-driven and independent of derivation and, as a corollary; (b) that lexical and derived constituents of the same type can prime each other. These effects are not expected under more traditional views of priming as structure-dependent.

4.6.2 Incrementality

Models of syntactic processing differ in the extent of their incrementality. A language generator, for instance, could work top-down, driven only by semantics. In that case, the last word of a sentence, or the last phrase of a long utterance, could be generated first, and it would be stored before it is uttered. A fully incremental generator, on the other hand, can select and adjoin every word to the current syntactic representation as it is produced, and very little buffering is necessary.

Various studies have examined the degree of incrementality in comprehension and production. See F. Ferreira and Swets (2002) for a summary that formed the basis of part of this section.

To evaluate whether speaking begins before phonological planning has completed for the whole utterance, experimental designs manipulate the phonological complexity of words at the beginning and the end of utterances. Wheeldon and Lahiri (1997) tested incrementality in production. Their Dutch-speaking subjects were given a noun phrase (e.g., *het water*—the water) and a question (*Wat zoek je?*—what do you seek?). The subjects were to answer the question as quickly as possible in a full sentence. Wheeldon and Lahiri found that their subjects began their sentences earlier when the first word was phonologically less complex. In further experiments, participants were asked to plan their sentences carefully. Then, sentence production latencies depended on the complexity of the entire utterance—not just on the first word. Wheeldon and Lahiri conclude that speakers start speaking whenever possible.

On the syntactic level, incrementality can be tested by manipulating the set of choices that a speaker needs to consider before beginning to decide on a sentence-initial word. V. Ferreira (1996) found that production is faster when there is more syntactic choice: when a verb allows different complements, utterance onset latencies are lower than when the verb only allows one type of complement. In his experiment, participants were presented with beginnings of sentences with either one of two verbs: *I gave*, or *I donated*, and then two complements, shown in sequence: *toys, children*, or *children, toys*. In the case of *donate*, only a prepositional object with *to* is possible: *I gave/*donated the children the toys*, vs. *I gave/donated the toys to the children*. Under the incremental assumption, we expect production to be easier in the case of *I gave* than for *I donated*, when the complement sequence was *children, toys*, because the complements may be syntactically integrated as they come if the verb allows these forms. That is exactly what V. Ferreira (1996) found.

Finally, not even the semantics of the utterance need be known before speaking begins. Brysbaert et al. (1998) asked their subjects to calculate the sums of simple additions ($21 + 4$ in one condition, or $4 + 21$ in the other). Subjects were asked to respond with the result as soon as possible. Subjects speaking two languages were tested: *Dutch*, where the response would be formulated as *five and twenty*, and *French*, where the response is *vingt-cinq* (twenty-five). That is, in Dutch, the last digit (five) is needed first, while in French, the ten position (twenty) is needed before a response can be given.

Dutch participants were faster to realize 25 in the $4 + 21$ than in the $21 + 4$ order, while French participants came up with the result of $21 + 4$ more quickly. Brysbaert et al. explain this with an incremental planning and production strategy: all speakers began giving the response as soon as possible. The French speakers need to calculate the *ten* digit, in this case, *twenty*, so they were fastest in the condition where this digit was available early. Dutch speakers wait for *five*, which is calculated more easily in the $4 + 21$ condition. Thus, they can start the phonological realization process as soon as part of the number is known. F. Ferreira and Swets (2002) lend some support the incrementality hypothesis, reporting results from an experiment using a similar paradigm. In these experiments, the difficulty of adding "ones" and "tens" digits was manipulated in an English language task. Participants took, overall, longer for problems that made it difficult to calculate the "ones" (*five* in the above example), that is, when a *carry* operation was involved ($24 + 7$ as opposed to $24 + 5$), but crucially, they did *not* start speaking later. This was the case even when they were under time pressure to produce responses quickly. However, participants, when under time pressure, lengthened the pronunciation of the first digit when the "ones" were difficult to calculate and, again, took longer overall. This points to *some* pre-utterance and *some* incremental planning during speaking, which is a strategic decision rather than a general principle. Consistently, (English) speakers preferred to see the two-digit addend before the one-digit addend ($21+4$), which F. Ferreira and Swets interpret as a sign for a *preferred order of planning*. Speakers prefer to plan in the order the arguments are realized. (This would be compatible with models that see complexity as a result of keeping more arguments active or in a buffer, e.g., Gibson's (1998) Dependency Locality Theory.)

There is also evidence that seems to contradict an incremental account. Meyer (1996) used semantic distractors for noun subjects and objects that appeared post-verbally in Dutch sentences. Such distractors are either similar or dissimilar to the subjects and objects in a phonological or semantic way, and they manipulate the difficulty of lexical access for the subjects or objects they are meant to distract from. As an underlying assumption, distractors are assumed to facilitate or inhibit lexical access. The crucial question is, again, one of utterance onset: does it take longer to begin to speak when a subject or object distractor is present? If so, subjects carry out lexical access for items late in the sentence even before the first word is spoken.

Meyer's answer is: yes, even with post-verbal object distractors, the presence of distractors caused sentence onset times to lengthen. This suggests that not just the first argument (normally the subject) and the verb, but also other arguments need to be accessed, because post-verbal distractors could not have otherwise exerted an influence on the subject's lexical access. This finding is, however, related to the *semantic* encoding of the utterance. It does not affect *syntactic* planning.

Standard patterns of realization in the verb phrase are easier to produce than non-standard variants. Stallings et al. (1998) found that the likelihood of producing Heavy NP-shifted structures (e.g., *Mary introduced to Bill the new neighbor.*) differed for verbs, and that sentences had a longer preparation phase when they did not conform to the verb's common way of Heavy NP shifting. The longer preparation serves to show that some planning seems to be carried out before the first word is spoken.

F. Ferreira and Swets (2002) argue that practical language production is not perfectly fluent (citations theirs). Speakers tend to pause before major syntactic constituents (e.g., Clark and Wasow, 1998), before deep clauses (Ford, 1982) or before complement and relative clauses (Holmes, 1988). F. Ferreira and Swets provoke thorough planning and (some) incremental behavior in their participants, depending on how pressured the participants are to speak quickly. The solution they propose is that incrementality should not be seen as "architectural". Instead, speakers strike a balance between speaking quickly and planning accurately—and usually they plan more accurately than they need to. The balance can be adjusted to suit the circumstances.

A syntactic theory such as CCG supports such a view, as the degree of incrementality is flexible (see Section 4.6.1 and also Experiment 12). A crucial prediction in combination with a very restrictive model of working memory (such as ACT-R's) is that incremental production and comprehension are cheaper than planned, non-incremental processing.

The comprehension data suggest that a production model would at least use a syntactic formalism and lexicalized structures that are compatible with incremental analyses, as those elements can be assumed to be shared between comprehension and production.

4.7 Predictions

4.7.1 Priming effects

We expect priming effects to apply to CCG *categories*, which describe the type of a constituent including the arguments it expects. Under our assumption that priming manifests itself as a tendency for repetition, repetition probability should be higher at short distances from a prime (see Section 2.1.4 for details).

4.7.2 Lexical and phrasal nodes

In categorial grammar, lexical categories specify the subcategorization behavior of their heads, capturing local and non-local arguments. As words are combined with others to form phrases, these phrases then are typed using the same system of categories. So, phrasal constituents may have the same categories as lexical items. For example, the verb *saw* might have the (lexical) category $(\mathsf{S}\backslash\mathsf{NP})/\mathsf{NP}$, which allows it to combine with an NP to the right. The resulting constituent for *saw Johanna* would be of category $\mathsf{S}\backslash\mathsf{NP}$—a constituent which expects an NP (the subject) to its left, and also the lexical category of an intransitive verb. Similarly, the constituent consisting of a ditransitive verb and its object, *gives the money*, has the same category as *saw*. Under the assumption that priming occurs for these categories, we proceed to test a hypothesis that follows from the fact that categories encode *unsatisfied* subcategorized arguments.

Given that a transitive verb has the same category as the constituent formed by a ditransitive verb and its direct object, we would expect that both categories can prime each other, if they are cognitive units. More generally, we would expect that lexical and phrasal (non-terminal) categories of the same syntactic type may prime each other. The interaction of such conditions with the priming effect can be quantified in the statistical model.

Lexical nodes are types of words as they are retrieved from the lexicon. They are terminal nodes in the derivation when seen as a tree.

Phrasal nodes are types of any phrase that is in a non-terminal position in the tree. (These types encode partially categories that result from the combination of other nodes, possibly satisfying part of their subcategorization frames.)

4.7.3 Incrementality of analyses

Type-raising and composition allow derivations that are mostly left-branching, or *incremental*. Adopting a left-to-right processing order for a sentence is important, if the syntactic theory is to make psycholinguistically viable predictions (Niv, 1994; Steedman, 2000).

Pickering et al. (2002) present priming experiments suggesting that, in production, structural dominance and linearization do not take place in different stages. Their argument involves verbal phrases with a shifted prepositional object such as *showed to the mechanic a torn overall*. At a dominance-only level, such phrases are equivalent to non-shifted prepositional constructions *(showed a torn overall to the mechanic)*, but the two variants may be differentiated at a linearization stage. Shifted primes do not prime prepositional objects in their canonical position, thus priming must occur at a linearized level, and a separate dominance level seems unlikely (unless priming is selective). CCG is compatible with one-stage formulations of syntax, as no transformation is assumed and categories encode linearization together with subcategorization.

CCG assumes that the processor may produce syntactically different, but semantically equivalent derivations.[7] We can produce such variants computationally, and we examine two of them: an *incremental* (left-branching) one, and a *normal-form* analysis, corresponding to conventional phrase structure analyses. While neither the incremental nor the normal-form analyses represent the single correct derivations, they are two extremes of a spectrum of derivations. Based on this syntactic view, and on the empirical evidence presented in Section 4.6.2, we propose the following hypothesis:

Flexible Incrementality Hypothesis: Speakers may mix advance planning of syntactic structure with incremental language production. Speakers may do so either to fulfill processing constraints or as a communicative strategy.

As a consequence, we expect to find priming effects predicted on the basis of both, incremental and normal-form CCG analyses.

4.8 Corpus data

4.8.1 The Switchboard corpus

We have already found structural priming effects for Penn-Treebank style phrase structure rules in the Switchboard corpus (see Section 2.2.1 for a description of the data).

4.8.2 Disfluencies

Speech is often disfluent, and speech repairs are known to repeat large portions of the preceding context (Johnson and Charniak, 2004). The original Switchboard transcripts contain these disfluencies (Figure 4.5).

It is unclear to what extent these repetitions are due to priming rather than simple correction. In disfluent utterances, we therefore eliminate reparanda and only keep repairs. Hesitations (`uh`, etc.), and utterances with unfinished constituents are also ignored.

[7] In the theory, selectional criteria such as information structure and intonation allow distinction between semantically different analyses. These were not relevant to produce the CCG version of the syntax annotations in Switchboard.

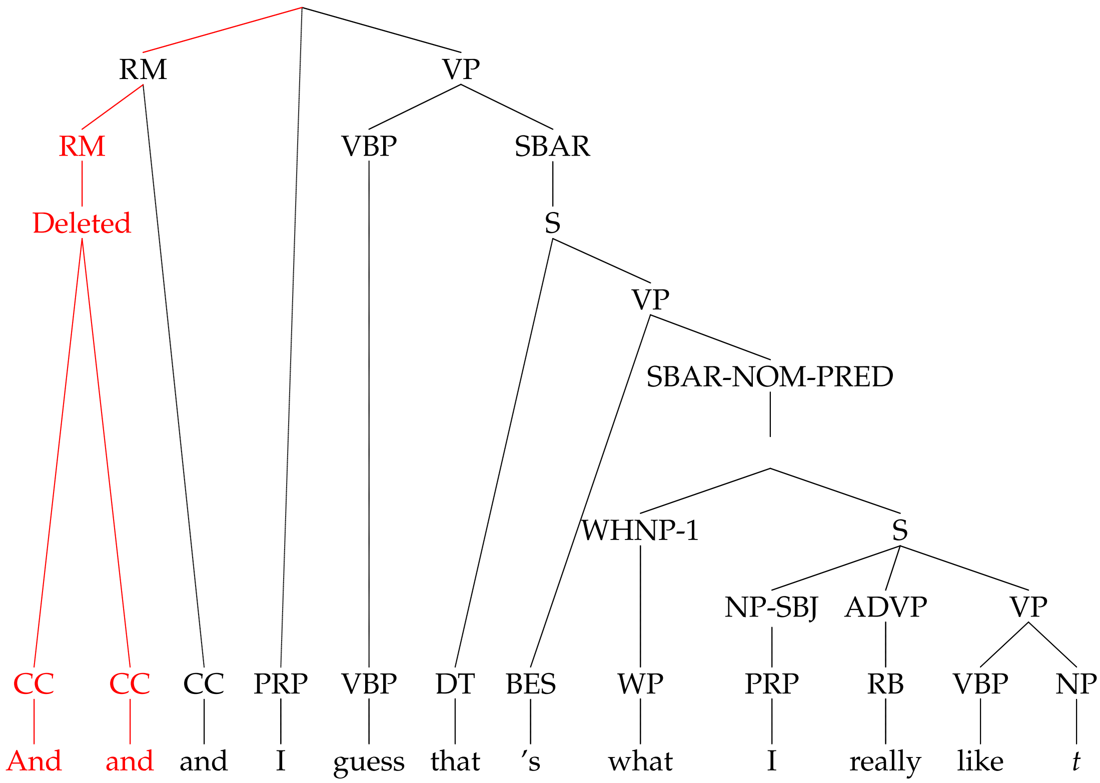


Figure 4.5: Disfluencies in Switchboard. Reparanda deleted in our analysis are marked in red. They include the subtree under "Deleted" and its immediate "RM" parent node, yielding *And and*.

4.8.3 Translating Switchboard to CCG

Since the Switchboard annotation is almost identical to the one of the Penn Treebank, we use a translation algorithm similar to the one used by Hockenmaier and Steedman (2007). We identify heads, arguments and adjuncts, binarize the trees, and assign categories in a recursive top-down fashion. Nonlocal dependencies that arise through wh-movement and right node raising are captured in the resulting derivation. Figure 4.6 shows the normal-form CCG derivation we obtain for the non-disfluent portion of the tree shown in Figure 4.5.

We then transform this normal-form derivation into the most incremental (i.e., left-branching) derivation possible, as shown in Figure 4.7. The transformation is implemented with a top-down recursive procedure which changes each subtree

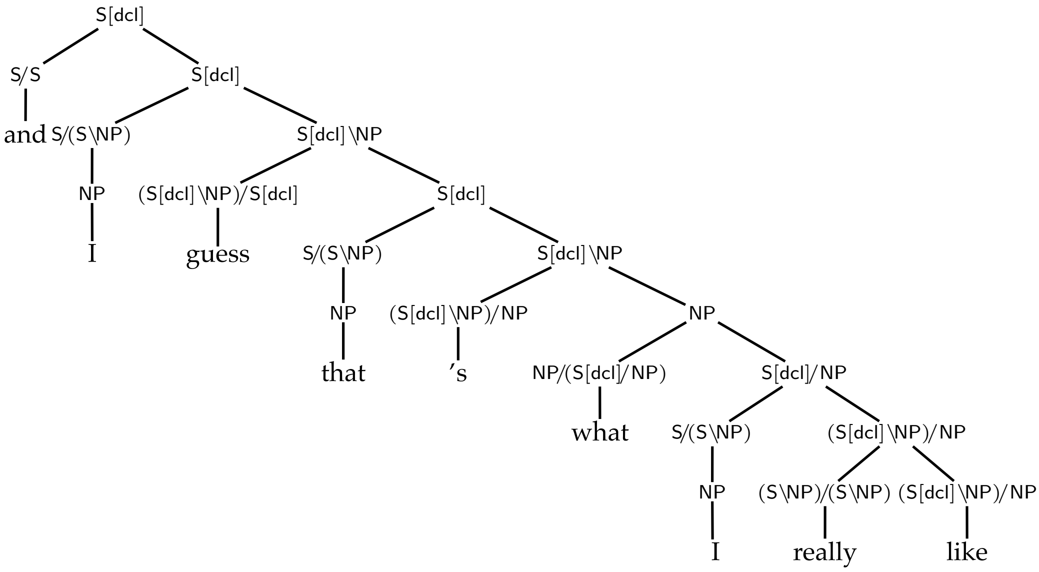


Figure 4.6: Normal-form CCG analysis of the sentence fragment *and I guess that's what I really like* from Switchboard.

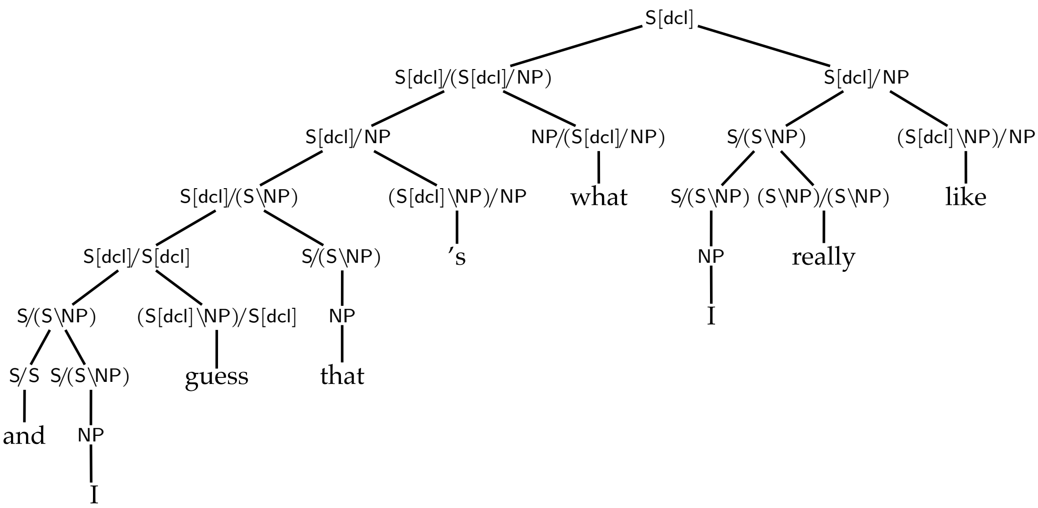


Figure 4.7: Incremental CCG analysis of the sentence fragment *and I guess that's what I really like* from Switchboard.

of depth two into an equivalent left-branching analysis if the combinatorial rules allow it. This procedure is run until no further transformation can be executed. It results in a maximally incremental derivation as far as allowed by CCG. The categories of the lexical nodes of the two derivations are identical.

4.9 Experiment 12: Priming within incremental and normal-form derivations

CCG assumes a multiplicity of semantically equivalent derivations with different syntactic constituent structures. Here, we investigate whether two of these, the normal-form and the most incremental derivation, differ in the strength with which syntactic priming occurs.

4.9.1 Method

CCG assumes a minimal set of combinatorial rule schemata. Much more than in those rules, syntactic decisions are evident from the *categories* that occur in the derivation.

Given the categories for each utterance, we can identify their repeated use. A certain amount of repetition will obviously be coincidental. But structural priming predicts that a target category will occur more frequently closer to a potential prime of the same category. Therefore, we can correlate the probability of repetition with the distance between prime and target. Generalized Linear Mixed Effects Models (GLMMs, Section 2.1.4) allow us to evaluate and quantify this correlation.

Every syntactic category is counted as a potential prime and (almost always) as a target for priming. Because interlocutors tend to stick to a topic during a conversation for some time, we exclude cases of syntactic repetition that are a result of the repetition of a whole phrase.

We include the log-transformed frequency of the syntactic category in Switchboard (ln FREQ) to estimate and account for the effect that frequency has on accessibility of the category.[8]

[8]Previous work points out that priming is sensitive to frequency (see Scheepers (2003) for high/low relative clause attachments and Chapter 2 for phrase structure rules). Highly frequent items prime less or are less sensitive to priming.

A joint model was built containing repetition data from both types of derivations: incremental and normal-form. We are only interested in cases where the two derivations differ: should we still find priming in both cases, then both types of derivations are viable analyses at least in some of the sentences. Should we find priming in only one of the derivation types, then we would conclude that our data should be described with only this variant of derivation. We therefore excluded all constituents where a string of words was analyzed as a constituent in both derivations. This produced a dataset where the two derivations could be contrasted.

A factor DERIVATION in the model indicates whether the repetition occurred in a normal-form (NF) or an incremental derivation (INC).

4.9.2 Results

The contrastive analysis for all factor combinations of PRIMETYPE (PP/CP) and DERIVATION (NF, Inc) shows significant and substantial priming for all conditions (for $\ln(\text{DIST})$, $\beta_{PP,Inc} = -2.181, \beta_{CP,Inc} = -1.380, \beta_{PP,NF} = -1.77, \beta_{CP,NF} = -0.423$, all $p < 0.0005$). The negative slopes indicate decay, hence priming in all factor combinations.

The logarithm of the normalized category frequency interacts with $\ln(\text{DIST})$ in each condition (for $\ln(\text{DIST})$:$\ln(\text{FREQ})$, $\beta_{PP,Inc} = 0.222, \beta_{CP,Inc} = 0.153, \beta_{PP,NF} = 0.175$, $\beta_{CP,NF} = 0.046$, all $p < 0.0005$). This indicates that priming weakens as frequency increases.

4.9.3 Discussion

If there was no priming of categories for incrementally formed constituents, we would expect to see a large effect of DERIVATION. On the contrary, we see no effect at a high p, where the regression method used is demonstrably powerful enough to detect even small changes in the priming effect. We conclude that there is no detectable difference in priming between the two derivation types.

The result is compatible with CCG's separation of derivation structure and the type of the result of derivation. It is not the derivation structure that primes, but rather the type of the result. It is also compatible with the possibility of a non-traditional constituent structure (such as the incremental analysis), even though it is clear that neither incremental nor normal-form derivations necessarily represent the ideal analysis.

An interesting further hypothesis arising from the CCG framework would possibly be that the incremental derivation of one sentence could prime the normal-form derivation of a later sentence, and vice verse. Unfortunately, the category sets occurring in the derivation variants had very few elements in common. Incremental and normal-form derivations produce different categories. This rendered the testing for actual repetition between different derivation types impossible.

4.10 Experiment 13: Priming between lexical and phrasal categories

CCG categories encode unsatisfied subcategorization constraints. Therefore, two constituents that would be very different from a traditional linguistic perspective can be assigned the same category under CCG. This is, perhaps, most evident in the categories of phrasal and lexical nodes (where, e.g., an intransitive verb is indistinguishable from a complete verb phrase).

Bock and Loebell's (1990) experiments suggest that priming effects are independent of the subcategorization frame. There, an active voice sentence primed a passive voice one with the same phrase structure, but a different subcategorization. If we find priming from lexical to phrasal categories, then our model demonstrates priming of subcategorization frames of the type that CCG assumes.

From a processing point of view, phrasal categories are distinct from lexical ones. Lexical nodes are bound to the lemma and thereby linked to the lexicon, while phrasal nodes are the result of a structural composition or decomposition process. The latter ones represent temporary states, encoding the syntactic *process*.

Here, we test whether lexical and phrasal categories can prime each other, and if so, contrast the strength of these priming effects.

4.10.1 Method

We built a model which allowed lexical and phrasal categories to prime each other.

Two factors PRIME LEVEL and TARGET LEVEL differentiate priming patterns, from and to lexical and phrasal levels. PRIMETYPE_{CP} distinguishes CP from PP priming. This design yields eight factor combinations.

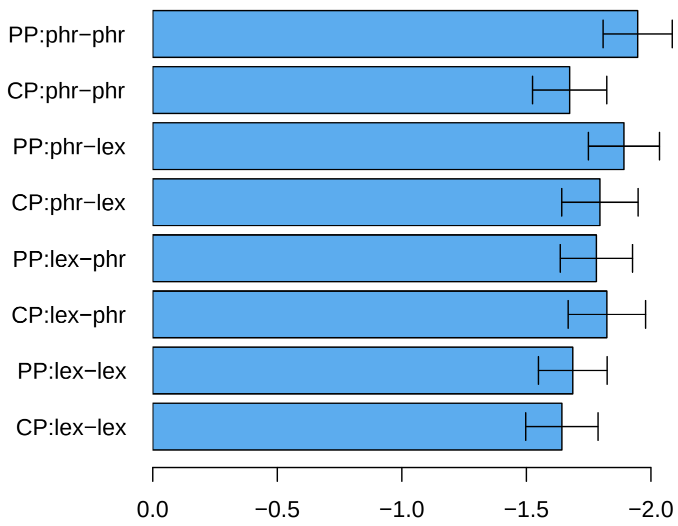


Figure 4.8: Coefficients for ln(DIST) in Experiment 13, for combinations of comprehension-production (CP) or production-production (PP) priming and lexical or phrasal primes and targets, e.g., the third bar denotes the decay in repetition probability of a phrasal category as prime and a lexical one as target, where prime and target occurred in utterances by the same speaker. Error bars show non-simultaneous 95% confidence intervals. Longer bars indicate more decay, hence more priming.

Recall that each data point encodes a possibility to repeat a CCG category, referring to a particular instance of a target category at time t and a time span of one second $[t-d-0.5, t-d+0.5]$, in which a priming instance of the same category could occur (*prime period*), at distance d seconds before the target. (The covariate for d in the model is ln(DIST)). If the prime category occurred at least once in the prime region, the data point was counted as a repetition (response variable: true), otherwise it was included as a case of non-repetition (response variable: false).

Thus, the target's type (lexical or phrasal) is simple to decide. For the category of the prime, the decision is more complex, because there could be both (or either) lexical or phrasal categories in the prime region. We therefore included two data points for each prime region, one for each type. The response, as usual, indicates whether a prime of the category of such a type occurred in the prime period.

We derived priming strengths for the different combinations of PRIME LEVEL (lex/phr) and TARGET LEVEL (lex/phr), determining whether priming differed between the levels. That is, PRIME LEVEL indicated whether the prime was on a lexical or phrasal level, and TARGET LEVEL indicated the same for the target.

4.10.2 Results

Figure 4.8 presents the resulting effect sizes for the eight conditions. Crucially, we see reliable syntactic priming in all eight cases.

4.10.3 Discussion

Categories of phrasal nodes prime other phrasal and lexical categories, and there also is a smaller priming effect emanating from lexical categories.

Albeit significant, we assume the effect of PRIME LEVEL is attributable to processing differences rather than the strong difference that would indicate that there is no priming of, e.g., lexical subcategorization frames. As the analysis of effect sizes shows, priming takes place both at the stage of lexical access and during the syntactic process.

Additionally, there is no evidence suggesting that, once frequency is taken into account, syntactic processes happening high up in derivation trees show more priming (see Scheepers, 2003).

Separate models for incremental and normal-form derivations were built. Both showed qualitatively the same results.

4.11 General discussion

We can confirm the syntactic priming effect for CCG categories. Priming occurs in incremental as well as in normal-form CCG derivations, and at different syntactic levels in those derivations: we demonstrated that priming effects persist across syntactic stages, from the lowest one (lexical categories) up to higher ones (phrasal categories). This is what CCG predicts if priming of categories is assumed.

Linguistic data is inherently noisy. Annotations contain errors, and conversions such as the one to CCG may add further error. However, since noise is distributed across the corpus, it is unlikely to affect priming effect strength or its interaction with the factors we used: priming, in this study, is defined as *decay* of repetition probability. We see the lack of control in the collection of a corpus like Switchboard not only as a challenge, but also as an advantage: it means that realistic data is present in the corpus, allowing us to conduct controlled experiments to validate a claim about a specific theory of competence grammar.

We find that priming effects occur not only between a canonical (normal-form) analysis whose constituents resemble those assigned by other grammar formalisms. Even the categories of constituents that result only from incremental derivations show priming. Such constituents are not assumed by other grammar formalisms.

The fact that incrementally produced categories prime normal-form ones is particularly interesting, since alternative viewpoints would propose unique parses at any point in time. Indeed, CCG suggests the existence of several analyses in parallel; a decision between incremental or normal-form analyses is not supported by our data. We see the two derivation types as extremes rather than as exhaustive or as configurational options.

Another observation we made is that there is a qualitative difference in the syntactic process during comprehension and production which accounts for marked differences between CP and PP priming. During production and comprehension, different types of derivations are preferably activated, or they are activated in different orders.

It seems like a puzzling dichotomy between the different results presented here that we find a relative lack of priming for distituent bigrams (as defined by normal form derivations), but priming in incremental CCG derivations for the same bigrams. Are the results from Experiments 9, 11 on the one hand, and from Experiments 12, 13 on the other hand antagonistic? It is important to keep in mind that distituency as defined very conservatively: it is likely that many structures considered distituent are part of a single constituent in a number of derivations. What the identification of distituent bigrams did was, in an informal sense, to maximize the number of structural boundaries crossed by each bigram. We created a set of bigrams that could be assumed to be at boundaries more often, and another set, that was more often within constituent boundaries. This holds true whether we analyze sentences using non-incremental phrase structure derivations, or possibly more incremental CCG derivations. Lesser priming for distituent bigrams indicates *that structure matters*—it does not indicate that the grammar that it is based on is necessarily correct.

The fact that CCG categories prime could be explained in a model that includes a basic form of subcategorization. All categories, if lexical or phrasal, contain a subcategorization frame, with only those categories present that have yet to be satisfied. Our CCG-based models make predictions for experimental studies, e.g., that specific heads with open subcategorization slots (such as transitive verbs) will be primed by phrases that require the same kinds of arguments (such as verbal phrases with a ditransitive verb and an argument).

The statistical models presented take the frequency of the syntactic category into account, reducing noise, especially in the conditions with lower numbers of (positive) repetition examples (e.g., CP and incremental derivations in Experiment 12). Whether there are significant qualitative and quantitative differences of PP and CP priming with respect to choice of derivation type—which would point out processing differences in comprehension vs. production priming—is a matter of future work.

We would caution against deriving very concrete claims about the architecture of the processor from a study done on a single corpus. Also, Chang et al.'s (2006) Dual Path model cannot be strongly rejected on the basis of our data: the Simple Recurrent Networks (SRN) used in their model are known to learn regularizations about derivational structures (Elman, 1990). It is unclear whether the SRN used in the Dual Path model has access to more than the abstract category of the previous word. Tests of distituent priming would sensibly be carried out using the model itself, which would require its extension to cover richer syntactic structure.

The general idea, however, is that different structural analyses with the same semantics can be kept and that a basic form of subcategorization frame as in a categorial grammar formalism exists and depends on priming. Comprehension and production make use of a shared linguistic representation. However, the fact that there is less priming between syntactic and lexical stages emanating from comprehending language is a hint that not all representations are shared.

So far, our data are compatible with the reality of a lexicalized, categorial grammar such as CCG as a component of the human sentence processor. CCG is unusual in allowing us to compare different types of derivational analyses within the same grammar framework. Focusing on CCG allowed us to contrast priming under different conditions, while still making a statistical and general statement about the priming effects for *all* syntactic phenomena covered by the grammar.

However, it is less clear from the CCG-based experiment alone, to what extent the syntactic structure assumed in priming studies can actually be arbitrary. If that was the case, we would be presented with priming of all sequences of lexical items or abstractions of lexical items. As we have seen before (Experiment 9), priming does not apply to all sequences equally. It is sensitive to general structural boundaries, even when we base the definition of those boundaries on a more traditional view of syntax rather than on CCG.

Which predictions would CCG make with respect to structural boundaries? If we assume that only one derivational structure is pursued during language production, then only this structure would cause and be sensitive to priming. Which derivation structure this is—recall that there is a spectrum from left- to right-branching derivations—cannot generally be determined in the corpus. An experimental design would have to force subjects to generate sentences incrementally, and then show that priming effects pertain to incremental combinations of structures rather than non-incremental ones, compared to a condition where subjects are allowed to plan their sentences non-incrementally. If we assume that all possible derivations are followed in parallel, then we could distinguish distituents from constituents in a corpus. However, almost any two adjacent words can combine in some way under CCG rules. Finding true distituents in the corpus is a futile endeavor unless we assume a certain derivation for each sentence. The priming results, however, are compatible with analyses that are sometimes more and sometimes less incremental, just like F. Ferreira and Swets (2002) concluded from their experiments.

4.12 Conclusions

The aim of this chapter was to shed light on the representations that underlie the human language production system by investigating the well-know structural priming effect that occurs when humans generate sentences. Structural priming, i.e., the repetition of previously used linguistic structures, can be explained using at least two alternative representational assumptions: either as the repetition of hierarchical representations generated by syntactic rules as proposed by Bock (1986b) and Branigan et al. (1999), or as the repetition of sequences of abstract lexical representations (e.g., parts of speech) as proposed by Chang et al. (2006).

We presented data from studies designed to distinguish the rule-based view from the sequencing view for priming. We investigated priming effects in a dialogue corpus for two types of part-of-speech pairs: Constituent POS pairs, which can occur within a syntactic constituent generated by a syntactic rule, and distituent

POS pairs, which cross constituent boundaries and can never occur solely within a constituent.

Experiment 9 dealt with short-term priming, i.e., with repetition effects that decay within a few seconds. We found a reliable priming effect for constituents bigrams, but less so for distituent bigrams. This finding is compatible with the structure-based view of priming, which would expect less priming of distituents, as these cannot be generated by syntactic rules. The results are at odds with the sequence priming view, which cannot distinguish between constituents and distituents, and would therefore predict priming for both.

Experiment 11 extended the study of syntactic priming to long-term adaptation effects. This repetition bias remains over long periods of time (hours and days). Its characteristics differ from those of short-term priming (e.g., no lexical boost). Our corpus study found a reliable long-term adaptation effect for low-frequency bigrams, which was similarly strong for distituents. This implies that the mechanisms underlying long-term adaptation and short-term priming differ.

Overall, our results are difficult to accommodate by simulations of sentence production such as the Dual-path Model, which assumes a sequence-based view of sentence production that does not involve a notion of constituency, and therefore cannot explain the lack of short-term priming for distituents. Also, the Chang et al. (2006) model assumes that a generalized implicit learning mechanism underlies both short-term and long-term priming. Again, this is at variance with our findings, which show clear differences between the two effects. Finally, we note that there are also experimental results, such as the priming of relative clause attachments (Scheepers, 2003) that are puzzling for the sequence-based view, as both high and low attachment involve the same POS sequence.

We conclude that an empirically adequate model of syntactic priming has to invoke a mechanism that operates on hierarchical syntactic representations to explain short-term priming, while a separate mechanism may be invoked to explain long-term priming. This is consistent with a structure-based view of priming. Priming operates in a time span where syntactic analysis in comprehension and syntactic realization in language production are affected. Adaptation is a memory effect, and simple sequences of linguistic representations may be implicitly learned.

Chapter 5

A Cognitive Model of Language Production

5.1 Introduction

In the work discussed so far, we have described two basic adaptation effects (short- and long-term) and their interactions with other parameters in the Switchboard and Map Task data. The interactions showed that adaptation levels differed between different types of dialogues, being greater in successful and task-oriented dialogues. Adaptation applies to syntactic structure as opposed to all sequences of lexical categories, but the effects are compatible with a more or less incremental production process.

Each of these interactions has its consequences for our understanding of human communication abilities or the architecture of the human language processor. Still, with this series of results, we have stopped short of actually specifying a model which could encode the algorithm that humans follow when they speak. Therefore, we will now seek to cast an idealized version of human language production as an instance of a general cognitive process. We will then show that structural priming follows from known properties of cognition.

In the preceding paragraph, we have, for the first time, used the term *model* in a different sense. All (cognitive) models are meant to simulate or explain limited aspects of cognition. However, they differ vastly in their general approach. *Statistical models* fitted in the studies described so far predict a very specific aspect of behavior (there: repetition) using a number of weighted, contributing measures. In the simple case of fixed-effects models, such a model is specified by a formula describing a set of measures and a vector of weights. Statistical models can support hypotheses derived from theoretical considerations, but they are unable to directly reduce such effects to more basic, cognitive principles, such as decision-making based on the activation of connected nodes.

Connectionist models take a step in this direction. They are able to simulate learning and decision-making using networks that approximate the basic, neurophysiological building blocks of human cognition. Neural networks, for instance, receive idealized input patterns, and their function is observable in their outputs. This function merely emerges as a whole, as Anderson (2007) points out, but the details of how the components function together are not part of a connectionist explanation. Once trained, a linear regression model can be seen as a specific form of a neural network. However, both statistical and connectionist models still lack a specification of the algorithm that is presumed to have generated the data in the first place.

This is what is provided by *cognitive models* such as the one we have developed. These models are designed to predict additional aspects of behavior: timing and difficulty or resource usage of the task at hand, as well as choice. Priming may influence all of these, but we concentrate on choice, given that is what our corpus methodology has been concerned with. Cognitive models, in our sense, detail the algorithm in step-by-step instructions as well as the data structures involved. They provide an end-to-end explanation that is comprehensive and believable. To make its predictions, a cognitive model defines an architecture that describes how separate components interact to achieve a specific behavior. Effectively, this gives an algorithm whose specification is flexible compared to statistical models. The model also defines data structures that represent pertinent information needed to carry out the task. Often, such a model comprises a broader task: while the statistical models describe distributions that relate to single, atomic syntactic choices made by speakers, the cognitive model describes the steps involved in composing a sentence.

A cognitive architecture, according to Anderson (2007), explains function on the basis of the modular system formed by the brain, keeping the representation of our physical *components* constant. Function does not emerge from a black box, but is traceable to the different components, and that is the stance taken in this chapter.

Founded on a general cognitive architecture, ACT-R (Anderson et al., 2004), we now introduce a model of human language production. We show how certain basic learning principles present in ACT-R can account for some of the priming effects discussed in this thesis.

Linguistic processing is regularly tied to particular components of the brain's architecture. So, it is plausible and likely that parts of the language facility will require specialized abilities. However, we assume that the general constraints of rule-based procedures, memory access and sub-symbolic processing still apply.

The task we set ourselves here is to frame results from the preceding chapters in the context of a more concrete language production model. We implement a simplified, yet plausible sentence generation model within ACT-R. The model produces grammatical English sentences given a semantic description, and it does so in line with selected empirical results pertaining to human language production. The model is not intended to cover all or even many aspects of syntax. We focus on a few syntactic constructions that have been used in priming experiments. However, the syntactic basis of the model (CCG) is formed by a syntactic theory that has been shown to cover a wide range of syntactic phenomena. CCG was previously shown (Chapter 4) to be adequate in a model of priming effects.

The motivation for choosing ACT-R lies in the comprehensive, plausible approach to cognitive modeling. ACT-R has been validated using (non-linguistic) experiments and is thus independently motivated. The choice of ACT-R also lies in our conviction that our empirical data resemble memory effects, for which ACT-R provides a detailed and well-established framework. We evaluate the model against a range of effects: these are evaluations carried out with either a crucial part of the model (Simulation 1) or with the full model (Simulations 2 and 3). Only then do we discuss further predictions arising from the model, for which experimental evidence had not been obtained at the time of model development. We evaluate some of these in Experiment 14.

The remainder of this chapter is structured as follows. Section 5.2 gives a summary of the underlying frameworks: ACT-R (cognition) and CCG (syntax). Section 5.3 introduces the language production model, explaining the components of ACT-R along the way. We discuss the emergence of priming and adaptation in Section 5.4. We motivate our evaluation methods in Section 5.5 and present three simulations in Sections 5.6–5.8. A further corpus experiment presented in Section 5.9

tests the prediction of a general lexical boost arising from the model. In Section 5.10, we compare the model to other language production and comprehension models and summarize the contributions in Section 5.11.

5.2 Background

We proceed to introduce ACT-R as a theory of cognition and its core principles. We then motivate CCG as the syntactic basis for the production model.

5.2.1 ACT-R

ACT-R is a general cognitive architecture developed by Anderson et al. (2004), whose constraints are intended to be cognitively realistic and motivated by empirical data. It has been widely used to model and match experimental data qualitatively and quantitatively. Like other architectures, ACT-R specifies how information is encoded in memory, how it is retrieved and combined. The core of ACT-R, with which the language production model has been realized, provides interfaces to sub-modules that concern vision, motor action and other functional modules that can be added. ACT-R, as it is known at present, was originally conceived as theory of memory alone (Anderson and Bower, 1973; Anderson, 1976). Working with the *Rational Analysis* assumption that cognition adapts optimally to the environment, Anderson developed the theory into one of the "Adaptive Character of Thought" (Anderson, 1990). ACT nowadays stands for *Adaptive Control of Thought*. The *-R* suffix points to the Rational Analysis character of the framework.

Architectures differ in which processes are serialized and which take place in parallel, and also in how these processes are specified. They also differ in the specific way information is encoded and can be accessed. ACT-R defines three core elements (see Figure 5.1). *Buffers* hold information temporarily. For instance, during language production, consider the case that we have uttered the phrase *I demand*. Then, the current syntactic state requires a sentential complement, i.e., a sentence usually beginning with *that*. This state is stored in a buffer, along with the semantics of the sentence. The second core element is procedural memory, consisting of *rules*. These are production rules with a condition (IF) and a consequence (THEN). The IF condition refers to the state of the buffers. For example, it may specify that

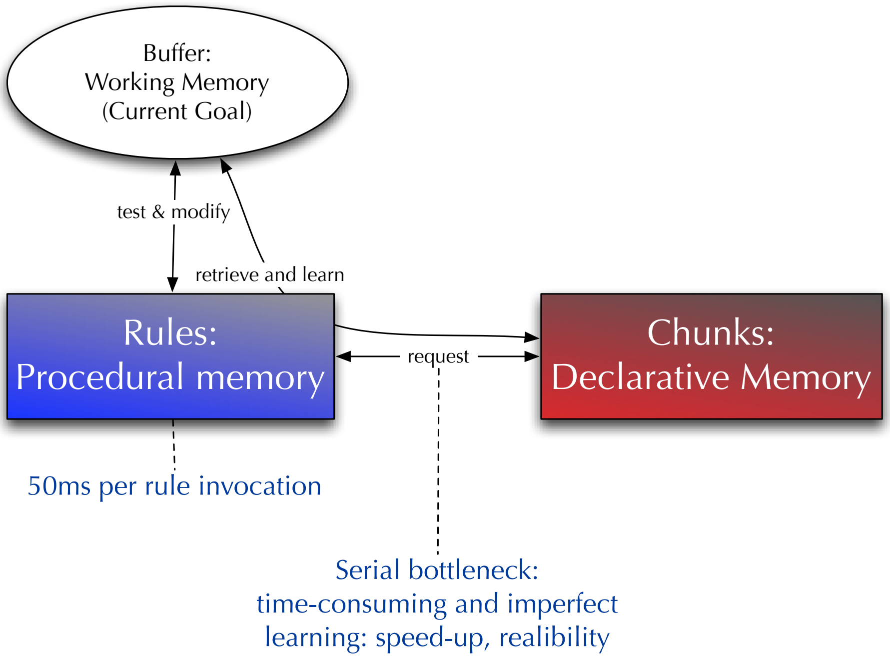


Figure 5.1: The three main components and their interaction in ACT-R.

a specific rule applies at the beginning of the sentence, where the syntactic state is empty. There, we would decide on a head verb of the clause. ACT-R defines a system that prioritizes rules and selects the correct one. In the first case, the THEN consequentially initiates the retrieval of a lexical item, such as the verb. This leads us to the third core element of ACT-R: declarative memory. Rules initiate memory retrievals, and they can also react to a retrieval once it has been completed. Memory is organized in *chunks*, which are bundles of information that are retrieved together. Chunks may compete for activation, and this is where lexical and syntactic decision-making takes place in our model.

While we have now summarized the *symbolic* part of ACT-R, there is also a *sub-symbolic* component. In many situations, we can have rules that compete for selection, or also several chunks in memory that each match a given retrieval request. The sub-symbolic system tells us which rule or chunk is selected, usually depending on how useful it has been in the past (in the case of rules), or how active it is at present (in the case of chunks).

Rule invocations and memory retrievals are not immediate. They require time, which largely depends on learning and activation effects. A model's predictions arise chiefly from these reaction times, and from the actual choices ACT-R makes when selecting a rule or a chunk in memory.

5.2.2 Syntactic basis

A design decision we need to make is whether structure-buildup during speech production follows only one structural analysis of the present partial utterance or tracks a multitude of syntactic structures. Similarly, we need to decide if large-scale syntactic decisions for the whole utterance are made before producing the first word (the one-structure view).

Incrementality in production questions the phrase-structure based accounts assumed in Chapter 3, as they stipulate planning processes that are computationally expensive and non-incremental. Linguistic evidence also supports syntactic frameworks that allow for a more flexible perspective on structural analysis (e.g., Phillips, 2003; Steedman, 2000). Common coordinate and also the less common gapping constructions (*Peter preferred to give, and John to receive money*) demonstrate that more than the traditional notion of a constituent (such as *give the money*) is needed. This is obviously relevant for both production and comprehension. Psycholinguistic experiments also point to incrementality in production and, to some extent, in comprehension (refer to Section 4.6.2, p. 99 for a review).

Incrementality is one of the most important properties motivating the use of Combinatory Categorial Grammar (CCG), a grammar formalism that has emerged from mainly computational considerations of human syntax. (CCG as a syntactic formalism has been introduced in Section 4.6.1.) Our model generates natural language incrementally. This way, speaking can start before the utterance production is completed. In incremental processing, we need to keep track of the syntax and semantics of the produced output. Semantically, we need to know *What else do I need to say in this utterance?* Syntactically, we need to know *How may a partial, grammatical utterance be continued?* This does not, however, entail retaining a complete syntactic representation of what has been said. Instead, our model often only stores a simple syntactic category describing the output. Thus, incremental CCG lends itself well to a cognitive view with strictly limited working memory, such as ACT-R.

CCG allows incremental and non-incremental realization at low computational cost (in particular regarding working memory). It assumes that multiple constituent structures are valid representations of a given realization and its semantics. These derivations may or may not exist in parallel. The parallel processing assumption does not, however, imply that multiple semantic analyses are maintained. Partial analyses can be produced for the words from the left edge of an utterance, with a minimum of representational burden: a single category is sufficient to describe the combinatorial properties of an incrementally generated phrase.

The second argument in favor of CCG is that we have demonstrated that syntactic priming effects can be explained as the preferred retrieval of categorial subcategorization frames. That is, syntactic CCG types are sensitive to priming.

Third, working memory is strictly constrained in ACT-R. CCG is compatible with that, since only a small amount of information about the current syntactic parse need be stored during sentence production. As long as the derivations used are incremental, the production algorithm makes do with a minimum of temporary storage to track its partial utterance.

(Linguistic) priming can be seen as a pre-activation of nodes representing lexical and syntactic information. We assume a categorical syntactic framework, with lemmas (lexicon) connected to subcategorization information. As shown in Chapter 4, priming can be modeled on a statistical level as a repetition bias that applies to CCG categories.

In the spirit of CCG, lemmata and syntactic categories are represented as declarative knowledge. We assume syntactic categories that encode information about the subcategorization frame of a given phrase and linearization information. This very much follows the idea of categorial grammars. Production rules are used to access such syntactic categories. The access of lemmas is biased by prior use, which implements priming effects. The access and syntactic combination of lexical material is controlled by a small set of rules which form the syntactic core. They encode the core of principles that can be seen as a universal, language-independent set of rules.

5.3 A language production model in ACT-R

In this section, we describe the model in the context of ACT-R's architecture, involving the core principles of working memory, procedural knowledge and declarative memory. We then explain, step by step, how the model produces a natural language sentence.

We adopt the following typographic conventions to distinguish the different concepts in ACT-R. The name of a chunk is indicated as 'offer-lexform'. A feature of the chunk is shown as AGENTSEM. The value of such a feature is normally the name of another chunk (set in plain type), or a textual representation, shown as "word". Chunks have types, for instance Lexical Form. A rule name is represented as *Select-Clause-Head*. Buffers are shown by their names, that is Goal or Retrieval.

5.3.1 Working memory

ACT-R's notion of working memory is very limited. Items from memory can be held in *buffers*. Each such buffer can only hold feature-value pairs. Each value stored in a buffer is atomic, and it is usually the name of a chunk in memory.

Initially, the language production model holds a semantic description of the utterance in the Goal buffer. The description consists of a flat predicate-argument structure. For a sentence such as *The policeman gave the speeding ticket to the driver* we describe the semantics with a predicate (*give*), an agent (*policeman*), a theme (*speeding ticket*) and a goal (*driver*).[1]

The Goal buffer also holds information about the state of the generation process. This way, rules can ensure that all the model's actions are carried out in the correct sequence.

Buffers represent a means for the different sub-modules of ACT-R to communicate with one another. The central control mechanism stores its state in the Goal buffer, but interacts with declarative memory. To retrieve a chunk from memory, a *Retrieval* buffer is filled, and in return, declarative memory augments the buffer with the information stored in the retrieved chunk. In Figure 5.2, the *Select-Clause-Head* rule requests a lexical form as the verb of the sentence using this buffer. It tracks its state by changing the STATE feature to 'retr-clause-head'.

[1]Note that the semantic do not reflect the actual choice of word form. The semantics of such a sentence will be more complex in models that cover more linguistic material, for instance, they will include tense. Such details could be represented as a combination of information stored in buffers and in declarative memory.

$$\begin{bmatrix} \text{Goal buffer} & \\ \textsc{AgentSem} & \text{'policeman'} \\ \textsc{PredSem} & \text{'give'} \\ \textsc{ThemeSem} & \text{'speeding fine'} \\ \textsc{GoalSem} & \text{'driver'} \\ \textsc{State} & \text{'initial'} \end{bmatrix} \begin{matrix} \rightarrow \\ \text{Rule:} \\ \textit{Select-} \\ \textit{Clause-Head} \\ \\ \\ \end{matrix} \begin{matrix} \begin{bmatrix} \text{Goal buffer} & \\ \textsc{AgentSem} & \text{'policeman'} \\ \textsc{PredSem} & \text{'give'} \\ \textsc{ThemeSem} & \text{'speeding fine'} \\ \textsc{GoalSem} & \text{'driver'} \\ \textsc{State} & \text{'retr-clause-head'} \end{bmatrix} \\ \begin{bmatrix} \text{Retrieval buffer} & \\ \textsc{Is-A} & \text{Lexical Form} \\ \textsc{Sem} & \text{'give'} \end{bmatrix} \end{matrix}$$

Figure 5.2: One step in the generation of a simple sentence, showing the states of Goal and Retrieval buffers before and after the invocation of a rule (*Select-Clause-Head*). The rule (shown in Figure 5.3) fills the Retrieval buffer to request a Lexical Form. The outcome of this request occurs later. Invoking the rule takes about 50ms.

5.3.2 Procedural knowledge

Having explained how temporary information is stored in the buffer, we now describe the control mechanism needed to change such information and retrieve it from memory.

Procedural knowledge is encoded in IF-THEN style production rules. A rule fires when its stated preconditions (IF) are met. Such preconditions check the contents of buffers: most commonly, they test whether a certain value is assigned to a feature in the feature-value pairs in a buffer. For instance, the *Select-Clause-Head* rule only applies in situations where the STATE feature is set to 'initial'.

Once a rule has been selected, it may change the contents of the buffers or interact with further modules, e.g., request the visual module to attend to an object seen at a specific location. Most commonly, though, the THEN part of the rule uses the Retrieval buffer to request a chunk from declarative memory. Figure 5.3 demonstrates such a case. The value of the PREDSEM feature contains the name of the semantics of the predicate. It is read from the Goal buffer and copied to the Retrieval buffer to request a lexical form whose semantics match the predicate. ($\boxed{1}$ indicates a variable local to this rule.)

The request to the memory module contains constraints similar to the precon-

Select-Clause-Head:

$$\begin{bmatrix} \text{Goal buffer} & \\ \textsc{PredSem} & \boxed{1} \\ \textsc{State} & \text{'initial'} \end{bmatrix} \Longrightarrow \begin{bmatrix} \text{Goal buffer} & \\ \textsc{State} & \text{'retr-clause-head'} \end{bmatrix} \begin{bmatrix} \text{Retrieval buffer} & \\ \textsc{Is-A} & \text{Lexical Form} \\ \textsc{Sem} & \boxed{1} \end{bmatrix}$$

Figure 5.3: The *Select-Clause-Head* rule of the form IF $\Longrightarrow$THEN requests a Lexical Form for the semantic predicate. Once the memory module has delivered the Lexical Form, another rule will deal with it. In the THEN side of the rule, we only show the changes that apply to the specific buffers. In the Goal buffer, all other information stays intact. By filling the Retrieval buffer, a new request is initiated.

dition in a rule (only positive tests are allowed): the name or some feature values of the chunk requested may be given. In this case, all lexical form chunks for specific semantics are eligible for retrieval.

Once the Retrieval buffer has been filled, the memory subsystem deals with retrieving the chunk. This takes time. Once the chunk has been retrieved, another rule will match and copy the retrieved chunk into the Goal buffer.

While the IF preconditions of all rules are evaluated in parallel, the actual invocation of a single rule takes time, by default $50ms$. Rule invocation and memory retrieval account for the total sentence production time (without phonological and phonetic processes). The model predicts about 4 seconds for the production of a simple sentence with a ditransitive verb, which is plausible given our corpus data.

5.3.3 Declarative memory

Figure 5.3 shows an example of a simple request for a chunk from declarative memory. We now detail the structure of such chunks in memory.

Chunks can be seen as feature bundles: each chunk is a set of attribute-value pairs. Values may be numbers, strings or references to (names of) other chunks. Unlike in often-used linguistic descriptions in the form of feature structures, values in ACT-R chunks cannot contain other chunks—they can only reference them. Thus, declarative chunks are flat. (The consequence of this is that the internal structure of a referenced chunk is not accessible.)

One of several lexical forms matching the request for the 'speeding fine' semantics contains the following pieces of information.

'speeding-fine-lexform'	
IS-A	Lexical Form
SEM	'speeding-fine'
LEX	"speeding ticket"

'speeding-fine-lexform' is the name of this chunk, IS-A, SEM and LEX are attribute names, Lexical Form, 'speeding fine' and "speeding ticket" are values, of which "speeding ticket" is a string, and Lexical Form gives the type the chunk. 'speeding-fine' is the name of another chunk, indicating the semantics the chunks realizes. In this model, the semantic chunk is atomic and does not carry further information.

This chunk is of type Lexical Form, it has a lexical realization ("speeding ticket") and realizes the semantics of 'speeding fine'. (In this model, the semantic chunk is atomic and does not carry further information.)

The decision to divide memorized information up into chunks is not merely one that allows us to operationalize an algorithm, demonstrating in this case that language production is possible with proven cognitive abilities and constraints. It is important for the functionality and predictions of the model, as each request for a chunk provides a separate opportunity for the processor to select a chunk out of a set of possible matching ones. Each memory retrieval provides a chance to model delays, errors and, possibly, priming biases. Features are not the only characteristics distinguishing chunks from one another. A chunk may also relate to other chunks, spreading activation to them whenever it is referenced in a buffer.

The lexical form structure above is still missing any specification of the syntactic properties of *speeding ticket*. Consider another lexical chunk that needs to be retrieved to realize the sentence: the lexical chunk for *gave*. Its syntactic properties are described in two separate *Syntax Chunks*. This is the Syntax Chunk for a ditransitive verb with an object in a prepositional phrase with *to*:

'ditrans-to'	
IS-A	Syntax Chunk
CLASS	'complex'
LEFT	'trans-to'
COMBINATOR	'forward'
RIGHT	'np'

The Syntax Chunk specifies a syntactic type in the CCG sense. The 'trans-to' type describes a transitive verb, so the ditransitive one is defined as a "transitive (to) verb with an additional NP". The 'trans-to' type, in turn, refers to a prepositional phrase type with *to*.

The alternative syntactic realization of *gave*, with a double object construction, looks very similar:

$$\begin{bmatrix} \text{'ditrans'} & \\ \textsc{Is-A} & \text{Syntax Chunk} \\ \textsc{Class} & \text{'complex'} \\ \textsc{Left} & \text{'trans'} \\ \textsc{Combinator} & \text{'forward'} \\ 90\ \textsc{Right} & \text{'np'} \end{bmatrix}$$

Empirical evidence for the fact that these syntax chunks are separate from one another and from the common lexical form comes from syntactic priming, which applies even when different verbs are used in prime and target. A preference for one syntactic form over another follows from one of ACT-R's sub-symbolic mechanisms: *spreading activation* (see Figure 5.4).

Spreading activation is one mechanism by which ACT-R may favor the retrieval of one chunk over others: the request for, e.g., a Syntax Chunk does not necessarily specify which syntactic realization is to be chosen. Speakers have some discretion here. The memory subsystem will retrieve the chunk that has the highest activation.

A chunk's activation consists of two main components: base-level activation and spreading activation. Base-level activation is learned over a long period of time. It increases with each *presentation* of the chunk, which can be thought of as retrieval.[2] Spreading activation depends on the chunks that are named in the current buffers and the links between them. In the above example, the lexical form is linked with two Syntax Chunks. Because the lexical form is present in the Goal buffer, it spreads activation to the Syntax Chunks.

Spreading activation makes it possible to retrieve a correct syntactic variant given the lexical form. Often lexical forms are connected to several Syntax Chunks,

[2]The precise definition of a presentation involves the clearing of a buffer after a chunk has been used; this detail is not relevant to the discussion of the model at hand.

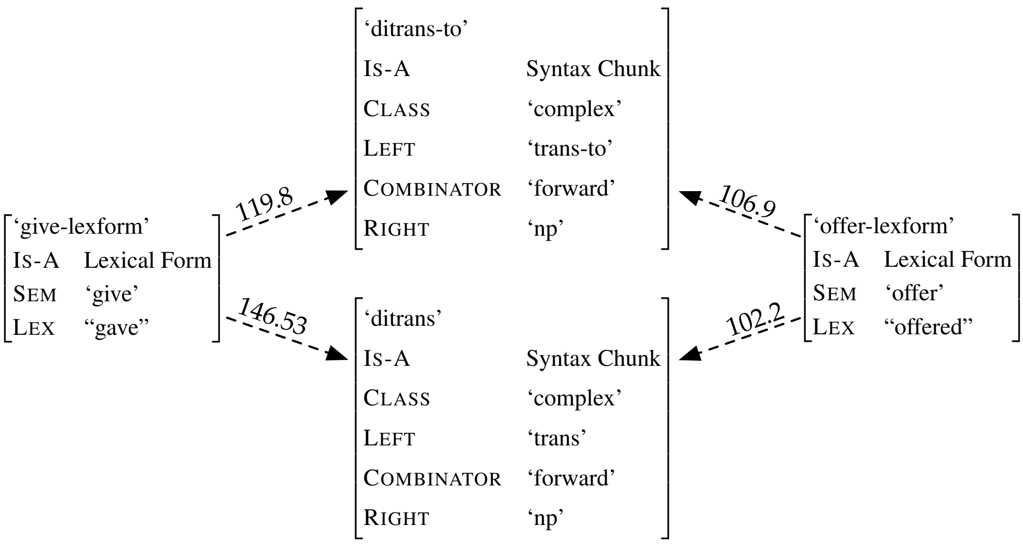


Figure 5.4: Lexical Forms (gave, offered) select syntactic categories (ditransitive with one "to" complement, top, and a ditransitive with NP-NP structure, bottom). The link strengths (unit-less) shown here were estimated from the Wall Street Journal (CCGBank) corpus, as explained in Section 5.3.8.

allowing the speaker to choose different variants. Some variants will be more and others less common: this distribution (the *frame selection bias*) is reflected in the different strengths of the links from lexical forms to Syntax Chunks (see Figure 5.4). For instance, the lexical form *gave* spreads activation to two Syntax Chunks, namely to ((S\NP)/PP[to])/NP (a ditransitive verb with a prepositional phrase complement; Figure 5.4, top) and to ((S\NP)/NP)/NP (a ditransitive verb with double object complements, Figure 5.4, bottom). Activation spread is not uniformly distributed across the different Syntax Chunks—verbs, for instance, will have more and less preferred (and more accessible) subcategorization frames (i.e., syntactic variants in the CCG sense). Spreading activation is thus stronger for the more common choice. However, speakers may make other choices, either due to random noise, or due to priming, which add to the overall activation of a Syntax Chunk.

The chunk's *base-level activation* is important in this model for the implementation of long-term adaptation. It changes as the syntactic type is used, and frequent retrieval will increase the base-level activation. The more recent a retrieval, the stronger is its impact: base-level activation decays over time. In the context of priming, base-level activation is the central mechanism to model preferential access of memorized material. ACT-R's base-level learning function causes an activation decay that appears similar to the priming effects observed. Consider Figure 5.5: here, the activation of a Syntax Chunk is shown over the course of $5,000$ seconds, with 14 presentations of the chunk at randomly chosen times. It was generated using the full model, i.e., a full sentence was generated for each presentation of the chunk. The chunk we activate is a syntactic form for a verb that subcategorizes a prepositional object (PO) complement with the preposition *to*, that is, the form 'ditrans-to'. The more highly this chunk is activated, the more likely the model is to choose the PO variant over the DO variant at the time.

Later in this chapter, we investigate whether this is an adequate explanation of the short-term and long-term priming effects and their interactions with chunk presentation frequency.

It is noteworthy that ACT-R has no explicit notion of short-term memory. Instead, the strong decay causes recently presented chunks to be much more accessible (for a few seconds). There is no place to store temporary structures apart from the buffer. Consequently, there is no cost associated with storage, and no "stack" with an associated storage cost is kept. This is relevant in light of theories of sentence complexity that depend on a stack of entities (e.g., Gibson 1998). In ACT-R, the storage cost is better modeled at retrieval time, where less recent retrievals are more costly (both in terms of time and accuracy). This has been demonstrated for language comprehension by Lewis and Vasishth (2005).

5.3.3.1 Types of chunks in memory

In the following, we describe different types of chunks as they are stored in memory. Recall that each chunk is a feature-value structure, with values often referring to another chunk (but not containing it). The *type* of a chunk implies a set of attributes that can be contained in chunks of said type. The type information is stored in an IS-A attribute (see the chunks in Section 5.3.3 for examples).

- Syntax Chunk: these chunks represent syntactic categories in the CCG sense. For instance, there is a chunk $\mathsf{S/(S\backslash NP)}$, containing the following feature-value structure:

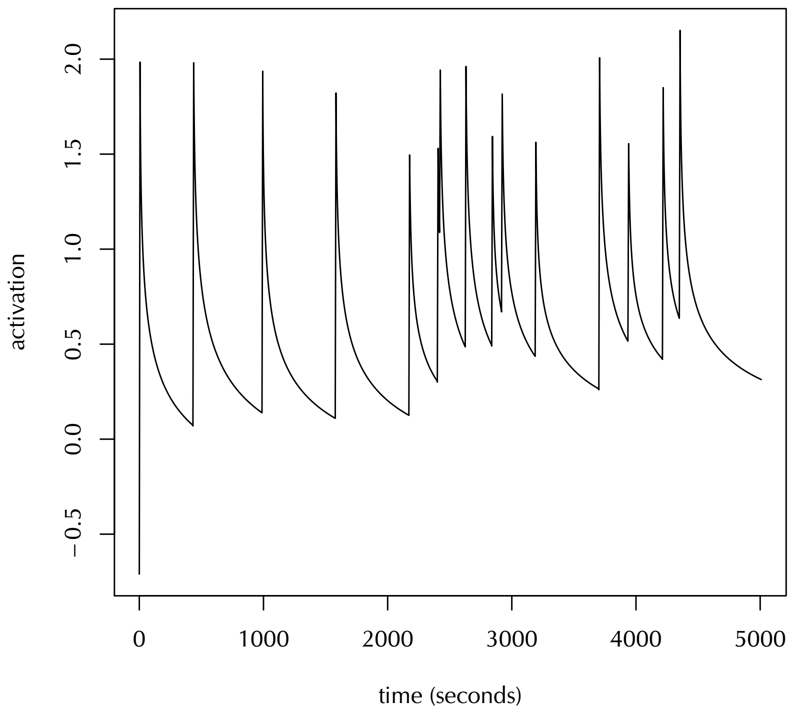


Figure 5.5: The activation level (unit-less) of the *ditransitive verb with PO(to) complement* Syntax Chunk during a series of presentations (retrieval cycles) of this chunk. The activation levels result from ACT-R's base-level learning function, which predicts a decay over time.

$$\begin{bmatrix} \text{`s-forward-s-backward-np'} & \\ \textsc{Is-A} & \text{Syntax Chunk} \\ \textsc{Left} & \text{`S'} \\ \textsc{Combinator} & \text{`forward-slash'} \\ \textsc{Right} & \text{`s-backward-np'} \\ \textsc{Attract} & \text{`null'} \end{bmatrix}$$

The Syntax Chunk contains its combinatorial components. (The *attract* feature is used for functional categories; in our model it specifies a preposition *to* in prepositional complements.)

- Lexical Forms contain core linguistic information about a full-form lexicon entry. A feature SEM refers to the concept, and LEX its linguistic realization. For instance, in the case of synonyms[3], we have two lexical forms containing the LEX values "the doctor" and "the physician", but the same SEM value, 'doc' (referring to some other chunk which is not specified further in our model). As a simplifying assumption, our model stores fully lexicalized noun phrases. Lexical Forms specify syntactic variants by means of spreading activation to Syntax Chunks (see Figure 5.4).

- Argument Order: these chunks provide argument ordering for a given combination of lexical form and syntactic variant. The order they specify controls the sequencing of thematic roles throughout the incremental generation process. Each such chunk refers to a lexical form (FOR-LEXFORM) and a Syntax Chunk (FOR-SYN).

$$\begin{bmatrix} \text{'gave-transto-order'} & \\ \text{IS-A} & \text{Argument Order} \\ \text{FOR-LEXFORM} & \text{'gave'} \\ \text{FOR-SYN} & \text{'ditrans-to'} \end{bmatrix}$$

  By means of spreading activation to thematic roles, the Argument Order chunks specify the order of complements, with most activation being spread to the first thematic role, for instance Agent.

- Thematic Role: these are atomic chunks named 'agent', 'theme', 'goal' and 'functor'. They receive spreading activation from the Argument Order chunks. Each Role chunk also spreads inhibitory activation to itself, preventing repeated retrieval. The 'functor' chunk identifies the semantic head of the clause, which is treated as an argument for the purpose of sequencing.

5.3.4 Procedural knowledge: Generation Algorithm

The basic algorithm is as follows. It assumes the semantic description of a single clause, i.e., earlier planning steps have already been carried out to the point at which syntactic realization can begin.

[3]The reader will note that we eschew a philosophical debate about the equivalence of meaning for the purposes of our model.

Initially, the Goal buffer holds the current semantics, consisting of a predicate and arguments associated with thematic roles (such as: AGENT, associated with 'policeman'). The names of argument chunks are stored in slots named AGENTSEM, THEMESEM, GOALSEM. During processing, the Goal buffer holds values in the following further slots:

- CONTEXT TYPE, a slot to describe the syntactic (CCG) type (a chunk name) of the currently generated phrase and is initially set to a special value 'beginning-of-clause'.
- NEW TYPE, a slot to store the syntactic (CCG) type of the portion of text currently generated, which is to be adjoined to the CONTEXT TYPE. It is empty initially. (Further slots are used for administrative purposes, which are omitted here for the sake of simplicity. N.B.: We use the slot names in lieu of the values they hold.)

The algorithm proceeds as follows:

1. Retrieve a Lexical Form of the semantic head for the semantics in the Goal buffer. (This will be the verb if a clause is to be realized.)

   Repeat:

   (a) Request and retrieve the next (most active) Thematic Role from memory. Stop if there is no argument in the current semantics for the role, or if no further role can be retrieved.

   (b) Identify the argument associated with the retrieved Thematic Role, and request and retrieve from memory a Lexical Form for the semantics of this argument.

   (c) Request and retrieve a Syntax Chunk from memory for the retrieved Lexical Form and store the 'left', 'comb', 'right' values of that node in the Goal buffer as the NEW TYPE.

   (d) Adjoin: Combine the NEW TYPE with the CONTEXT TYPE according to one of the combinatorial rules.

This algorithm would be sufficient if generation could take place in a fully incremental fashion. However, the notion of *flexible incrementality* as suggested by the results in Chapter 4 requires the syntactic realization algorithm to plan ahead. We therefore introduce a step of (limited) recursion. The current state (represented by the semantics and by CONTEXT TYPE) needs to be stored. A sub-phrase is begun, starting with an empty CONTEXT TYPE, with new material forming a separate constituent until the current CONTEXT TYPE may be adjoined to the saved type on the stack. Here, we implement a limited version with stack size 1:

1. Retrieve a Lexical Form of the semantic head for the given semantics. (This will be the verb if a clause is to be realized.)

   Repeat:

   (a) Request and retrieve the next (most active) Thematic Role from memory. Stop if there is no argument in the current semantics for the role, or if no further role can be retrieved.

   (b) Identify the argument associated with the retrieved Thematic Role, and request and retrieve from memory a Lexical Form for the semantics of this argument.

   (c) Retrieve a Syntax Chunk for the lexical form and store the 'left', 'comb', 'right' values of that node in the Goal buffer as the NEW TYPE.

   (d) Adjoin: Combine the NEW TYPE with the CONTEXT TYPE either according to one of the combinatorial rules or by retrieving a learned combination from memory[4], updating CONTEXT TYPE with the resulting combination and clearing NEW TYPE.

      i. *If unsuccessful (not combinable) and* STACKED TYPE *is empty*, copy the CONTEXT TYPE into STACKED TYPE and move NEW TYPE into CONTEXT TYPE.

      ii. *If successful and* STACKED TYPE *is filled,* attempt to adjoin the (new) CONTEXT TYPE to STACKED TYPE, updat-

[4]As a simplification, we use a number of stored combination patterns to specify valid combinations of categories, as not all combinations of deeply hierarchical categories may be tested in the precondition of an ACT-R production rule. This refers to cases where A/(B/C) forward-combines with (C/(D/E)) to A/B/D/E, a situation where the internal structure of B/C is not accessible directly in the ACT-R rule.

ing CONTEXT TYPE with the resulting combination and clearing STACKED TYPE.

This implementation of n-fold recursion is possible for finite n. In the ACT-R context it predicts that the STACKED TYPE spreads activation.

An alternative model would store the stacked type in declarative memory. This would implement a stack, because the most recently acquired goal would be the most accessible one and could be retrieved first. Deletion of stacked goals and also repeated shelving of the same syntactic type would pose serious challenges in the ACT-R context. This variant predicts more processing difficulty in cases where production cannot (or does not) proceed incrementally, including retrieval errors of stacked items, and possibly even language evolving towards incrementally generatable structures. The (implemented) Stacked Type variant predicts a hard limit for stacking in the context of ACT-R's production rules and "small" buffers. The differentiation of the two methods is beyond the scope for this thesis.

Note that incremental derivations in CCG often require the use of type-raising. Rather than exploring each type-raised version (potentially in parallel with the non-type-raised one), we store type-raised variants in the lexicon. The correct analysis is retrieved because the preceding syntactic context spreads activation to it. For instance, to generate *Stella saw Amit*, the subject noun retrieved from the lexicon is of form $\mathsf{S}/(\mathsf{S}\backslash\mathsf{NP})$ (type-raised) rather than NP, as it is at the beginning of the sentence. This way it can combine with *saw* ($(\mathsf{S}\backslash\mathsf{NP})/\mathsf{NP}$) to yield *Stella saw* (S/NP).

5.3.5 Argument order

Recall that we started the generation with the Agent role. How do we decide about the order in which arguments are realized? This question is relevant not just from an algorithmic point of view. The decision also leads to predictions about whether arguments and argument order can be primed.

There is little distinction between grammatical functions (subject, objects) and the order of thematic roles. The order of arguments is retrieved as a chunk after the head (in our case: the verb) and its syntactic realization is chosen (see Section 5.3.3.1 for a description of these chunks). Thus, we bind a sequence of grammatical functions (as defined by the syntactic nodes) to a sequence of thematic roles.[5]

[5] This is a simplifying implementation.

Arguments are ordered according to the activation of thematic role chunks. This activation pattern results from spreading activation, coming from the Argument Order node that is retrieved immediately after the head is chosen.

The most highly activated thematic role is retrieved first. As a default, Agents are retrieved first, while Goals are retrieved later. Additionally, this accessibility is specific to the verb, such that other lexical form-syntax combinations may require a different order. For instance, the verb *give* has syntactic variants for the double object realization (*He gave Mary the flowers*) and for the prepositional-object variant (*He gave the flowers to Mary*). These variants differ in their argument order.

Argument order is constrained by several factors. A combination of a lemma and a Syntax Chunk preselects a number of possible argument orders. Such argument orders are stored as separate chunks which would, in principle, be sensitive to priming upon retrieval. However, since the argument order is decided only after a lexical form and its syntactic variant have been chosen, priming of argument order cannot influence decisions about lexical forms and syntactic variants. Thus, no long-term adaptation of argument order is predicted unless lexical form and syntactic variant are repeated between prime and target. However, argument order decisions may still influence syntactic choice, as they frequently do in order to satisfy information structure related conventions, e.g., theme-rheme ordering specifying a preference to present known information early in the sentence, and new information late (Halliday, 1967; Grosz et al., 1995). Such a bias can be modeled as spreading activation in ACT-R. In that case, activation spreads from chunks still present in a buffer from processing the previous utterance. The clear prediction arising from this is that any argument order priming effects in production must be short-lived. In our model, such priming would resemble lexical boost effects.

With each realized referent, the activation of the chunks representing the arguments is reinforced, using the standard ACT-R theory of learning.

A top-down realization algorithm would start from a semantic description, making choices about the realization of constituents as they arise, with the lexical realization coming at the bottom of the syntactic tree. The current algorithm always begins with the semantics and generates in a left-corner fashion, even though it retrieves the lexical entry for the semantic head of the clause (the verb) first.

5.3.6 Lexical forms govern syntactic choice

Lexical Forms such as *gave* or *offered* spread activation to syntactic variants. Variants are retrieved after lexical forms, which implies that the lexical form and its semantic contribution governs the production process.

In ACT-R, once a production rule has been selected, its effects will not be undone: there is no backtracking.[6] This means that an earlier choice is not influenced by the accessibility of choices that follow later. Concretely this means that the choice of a lexical form does not depend on the preferred syntactic variant. Once a lexical form is chosen, however, the syntactic variant is subject to any bias that is introduced by priming (or other) effects.

Throughout the statistical analysis, we have implied no such role for any decisions taken prior to the selection of a syntactic form. There, we looked at syntactic rules independent of the semantic or lexical content. For priming, it did not matter whether there was actually a syntactic choice, as would be the case in different syntactic variants of the same semantic content.[7] The ACT-R model, on the other hand, now adds this constraint. The constraint naturally follows from the fact that ACT-R does not prioritize the retrieval of a lexical form depending on the accessibility of other chunks referenced by that lexical form. In other words, accessibility is not determined compositionally. So, while we can statistically model choice independently of the concept of *syntactic alternation*, a serial processing model is faced with making choices based on defined semantics and, as in this case, even a lexical form.

5.3.7 Incrementality in the model

Our algorithm begins to generate an utterance with the first words, rather than to plan it in detail before beginning to speak. It bears some resemblance to Purver and Otsuka's (2003) incremental generation algorithm, which chooses one word at a time and tries to use normal incremental parsing techniques to integrate the word into the representation of the partial sentence. This representation must be subsumed by the semantic representation. Not all possible lexical items need to be tried. Instead, the initial semantics activate the right words. (Their algorithm

[6]While repair in spoken language is beyond the scope of this model, they may be a result of lexical and syntactic decisions to which a speaker has committed.

[7]Examples of syntactic choice given constant semantics come from the experimental priming paradigm: Double-Object vs. Prepositional Object, verb particle placement, active/passive constructions, that deletion, and others.

builds on the Dynamic Syntax framework, Kempson et al., 2000). However, our algorithm is directed by thematic roles rather than lexical items. Only after the role is chosen does the processor pick a lexical form.

The incremental algorithm proposed here differs greatly from a CCG generation approach such as White and Baldridge's (2003), where a number of realization variants are generated via a chart from the given semantics, and an n-gram language model is used to pick the most natural variant. Our generation algorithm is not purely head-driven—we do not always begin with heads, even though we do decide on the head (verb) of the finite clause initially. It is incremental (see also Hoffman, 1994) and not purely bottom-up: lexical material is adjoined to the context in a left-to-right fashion, but the order thematic roles determines the piece of information that is verbalized next.

Having decided to realize the subject early in the sentence, we employ incremental CCG constituents to keep track of the sentence that has been realized so far. For example, *the policeman* ($\mathsf{S/(S\backslash NP)}$) combines via forward combination with *gave* ($\mathsf{(S\backslash NP)/NP/NP}$) to $\mathsf{S/NP/NP}$. The latter category represents the syntactic state having realized *The policeman gave*.

This incremental analysis presents a problem in the way we choose lexical categories: when *the policeman* is initially realized, we need to commit to either the type-raised variant $\mathsf{S/(S\backslash NP)}$, as above, or the non-type-raised variant NP. Type-raising is necessary in the subject position, while as an object, in CCG we would choose not to type-raise the noun phrase, in which case its lexical category is NP. When making this decision, the processor knows that the current syntactic state is empty: we are at the beginning of a sentence. This configuration spreads activation to the type-raised variants of syntactic categories, based on a beginning-of-sentence marker chunk in the Goal buffer. Thus, $\mathsf{S/(S\backslash NP)}$ is preferred over an NP.

5.3.8 Initializing model parameters

General ACT-R parameters were left at or near their defaults, with base-level learning decay set to 0.5 (default), the activation noise set to 0.4 (a common choice) and the maximum associative strength to 50.0. Activation is unit-less.

Some of the effects may depend on the general frequency of structures and lexical forms. ACT-R models such general exposure as the *base-level activation* of chunks, which strongly influences which chunks are retrieved from memory, and how quickly this can happen. To vary the frequency of syntactic types and lexical forms stored in the lexicon, we initialized their base-levels with data acquired from a corpus. There are two options for a choice of corpus. The *Switchboard* corpus has been converted to CCG as described in Section 4.8.3 ($684,000$ words without disfluencies). The *CCGBank* is a reliably annotated, sufficiently large dataset (1 $148,000$ words). It is the only CCG corpus available publicly. CCGBank consists of articles that appeared in the *Wall Street Journal* (WSJ) and is based on a conversion process from the Penn Treebank annotations like the one employed to produce the Switchboard CCG data. Unlike the spontaneous language production simulated here, this is edited, written language, from a different genre. The set of syntactic constructions and lexical forms in any corpus will be heavily biased. For instance, the frequency ratio of *gave* to *offered* is $107 : 247$ in CCGBank, and $160 : 23$ in (our subset of) the Switchboard corpus. To do so, incremental derivations were produced for the corpora, which is relevant given that the algorithm discussed here is incremental, and that the category set differs significantly. Certain categories common to incremental derivations are very rare in non-incremental ones.

The model can be initialized with CCGBank and with Switchboard data, and we have verified that it generates the simulated range of utterances reliably with both cases, giving similar results regarding priming. To maintain coherence with the corpus studies, we initialized the syntactic types and lexical forms in the model using the Switchboard data.

The core categories in our simulation, namely those for ditransitive verbs with a prepositional complement, and ditransitive verbs with two noun phrase complements, are the same in the production model and the corpora. For instance, the verb *give* is annotated with category $\mathsf{S[dcl]\backslash NP)/PP[to])/NP}$ (ditransitive with prepositional phrase complement) in the sentence *Mr. Pilson scribbled a frighteningly large figure on a slip of paper, sealed it in an envelope and gave it to sports negotiators* (WSJ). Thus, the verb categories lend themselves to a simple mapping from the corpus frequencies and simulated exposure.

From the corpora, we can derive the relative distributions of lexical entries. However, what is less clear is the amount of overall language exposure that we need to assume in order to create a realistic picture. In an extensive study of linguistic development and socioeconomic status, Hart and Risley (1995) assessed the linguistic exposure that children aged 12–36 months experienced, i.e., the number of words addressed to the child was estimated. They find a range of 10 million to about 35 million words, depending on the social class of the family, increasing linearly with age. Extrapolated, this translates to 50 million to 180 million words comprehended by the age of 15 (when first-language acquisition is expected to be completed). We assume a total of 225 million words (comprehended and produced), scaled over 15 years. Given the high between-person variability and the low decay at such long learning periods, a more precise estimate would only be sensible to capture differences between subjects.

The distribution of syntactic choices given a lexical form is also estimated from the corpora. The link strengths from lexical forms to syntactic chunks (as shown in Figure 5.4) result from the frequencies of syntactic forms for given lexical choices. Each strength is estimated as $0.5\phi(1+\hat{p}(syn|lex))$ for a given lexical form *lex* and a syntactic category *syn* in the corpus. Thus association is derived from the conditional probability of a particular syntactic realization given the lexical form ($\phi = 75$ is a norming parameter applicable to all link strengths across the model).

5.4 Priming mechanism

In the following, we outline two mechanisms of priming: a *learning-based account*, which explains structural priming as the modulation of accessibility of syntactic rules stored along with the lexical forms in memory, and, secondly, a *spreading activation explanation*, in which activation emanates from lexical forms retained in buffers.

5.4.1 Priming as learning

The previous chapters have pointed out two kinds of repetition biases: short-term priming and long-term adaptation. The empirical analysis (as well as the original methodology) have suggested that these effects have separate cognitive bases. The question addressed in the model is whether we can unify the repetition biases under the simple and elegant learning framework provided by ACT-R. The interaction of priming with frequency plays an important role in this investigation.

So far, we have treated short-term priming and long-term adaptation as effects with two separate cognitive bases. Qualitatively, the effects seem to differ in their *decay* (short-term priming decays quickly). The two effects differ in their interactions with dialogue genre, with task success and with structural properties of what is repeated.

A commonly asked question is whether the two kinds of repetition effects can be unified. In this context, one may ask: can short-term priming and long-term adaptation both be based on a single *learning* effect as defined by ACT-R's *base-level-learning* function?

What would be learned is a relative preference for syntactic categories in the CCG sense, which are tied to entries in the lexicon. For example, a ditransitive verb expecting a prepositional-object realization would be such a type. This lexicon-oriented view of syntactic memory is supported by some priming studies, for instance Melinger and Dobel (2005). In their study, subjects could be primed to use either DO/PO realizations (in German and Dutch) in a picture description task. Primes consisted of just a semantically unrelated ditransitive verb, which allowed only one of the two argument patterns.

Section 5.6 describes a simulation of just this. There, we show that the fact that low-frequency syntactic decisions show more priming can be modeled using the learning function. It should be noted that this effect is not unique to ACT-R. In general, conditioning depends on the discrepancy between the expected and the observed (Rescorla and Wagner, 1972). Consequently, the expectation of low-frequency events is less precise. Thus, it is less likely to match the observation, and leads to more surprisal.

However, two other effects are unexplained by the priming-as-base-level-learning analysis.

- Long-term adaptation is correlated with task success, but short-term adaptation is not (see Experiments 6 and 7, Chapter 3.7). Under a unified view of priming, we would expect that whenever we see strong long-term adaptation, we should find short-term priming. Due to the strong decay, interaction effects with covariates such as task success should be stronger for short-term priming.

- Lexical Boost: Repeating open-class words in prime and target boosts the priming effect. It has been shown that short-term priming effects are stronger, when lexical material (usually the head verb of a clause) is repeated. This has been demonstrated in many experiments, for instance, by Pickering and

Branigan (1998); Branigan et al. (1999) for prepositional object vs. double object dative constructions in written sentence completion. Branigan et al. (2003) finds it in a confederate scripting experiment (spoken dialogue), Gries (2005) in a corpus-based study, Cleland and Pickering (2003) for noun phrases (repeating the head noun) and for second-language speakers of English (Schoonbaert et al., 2007). It is unclear how a learning effect could explain this boost. Either lexical and syntactic information is learned separately, in which case we would expect only a small learning effect of the relatively highly frequent verbs, or lexical and syntactic information is stored jointly in one chunk, in which case we would expect no syntactic priming at all in different lexical contexts.

- Lexical boost is short-lived: the strength of priming is unaffected by head verb repetition when there is intervening linguistic material, i.e., when the prime-target distance is not minimal. Hartsuiker et al. (2008) elicited prime-target pairs at varying distances, manipulating whether verbs in the prime and target sentences were repeated. They found a lexical boost only in sentences that were adjacent, but not when two or six sentences intervened. In a series of studies, Kaschak and colleagues examined long-term priming effects and found no lexical boost, i.e., no enhanced structural repetition if the verb was repeated (Kaschak et al., 2006; Kaschak, 2007; Kaschak and Borregine, 2008). Under a unified account of short-term priming and long-term adaptation, we would expect that the two effects are equally sensitive to lexical repetition. In other words, we would expect a lexical boost for long-term adaptation as well, and not just for short-term priming. Kaschak et al.'s empirical evidence does not support that.

5.4.2 Priming as spreading activation

The second account sees priming as an effect that follows from activation spreading from working memory (buffers) to longer-term memory, thus making retrieval more likely and also faster. The account suggests that lexical forms used during production are held in buffers for a short while after they have been processed, often beyond the immediate utterance at hand. Holding the lexical forms in buffers is sensible, given that consecutive utterances tend to be linked via some of their referents if the discourse is coherent.

As we have shown in Experiments 3 and 9 (and virtually all other linear models), the short-term priming effect interacts with the frequency of the syntactic structure type: rare constructions show stronger short-term priming and adaptation. For long-term adaptation, we explain this interaction through ACT-R's base-level learning function. For short-term priming, the *fan effect* provides a potential explanation. As described by Anderson (1993), the fan effect means that chunks associated with a given cue are retrieved more slowly when other chunks are also associated with that cue. This effect scales with the number of chunks associated with the one to be retrieved. The fan effect was discovered using a recognition task. In Anderson (1974), participants studied 26 facts about people, whereas the number of facts per person varied (1, 2 or 3 facts). Then, participants judged sentences giving information about the fictional people as true or false—some of these sentences reflected the facts studied earlier (true), others did not (false). The more facts were associated with a person the slower participants were to respond.

In our model, lexical forms may persist in a buffer in order to process their semantic contribution, usually for the duration of a sentential unit, until they are replaced by other lexical forms. Similarly, semantic information may persist even beyond the utterance. By virtue of being in a buffer, lexical forms and semantic information spread activation from the buffer, most importantly to their own equivalents in memory and also to the chunks representing syntactic categories. So, while the lexical and semantic material is in the buffer, it is acting as a cue to retrieve a syntactic category (or indeed another lexical form) in the next processing step. The more frequent the syntactic category is, the greater is the *fan*: other lexical and semantic material will also be potential cues for the (same) category. The fan effect decreases the effect the lexical/semantic material has on retrieval. Thus, a highly frequent category will see less priming.

To summarize: both forms of priming contribute to the overall priming effects seen, but only spreading activation causes lexical boost effects.

5.4.3 Short-term priming and lexical boost

The fact that short-term priming decays can still be seen as the result of a base-level learning effect: decay of base-level learning is initially strong. We show that frequency effects typical for short-term priming hold for base-level learning (Simulation 1).

In ACT (and the subsequent ACT-R), general priming is commonly explained as a spreading activation effect. For the case of semantic correlates, for instance, *dog* would be retrieved from memory more quickly when *cat* has been retrieved before and is now available in a (semantic) buffer. Because of the relatedness between the two, activation spreads from the buffer to the chunk *dog* in memory (Anderson, 1990). In the context of language production, this account predicts a facilitatory effect of related linguistic material. For instance, in a head-initial language, the head would facilitate the recognition or production of its complements. In a head-final language, the roles are reversed: a distinctive complement could facilitate the recognition or production of the head (a prediction not tested here).

Semantic priming effects such as the one mentioned above are, empirically, short-lived. Lexical boost effects are similar: they are extremely short-lived and do not commonly survive more than one sentence (see Section 5.4.1). In the model, the repetition of lexical material boosts syntactic priming because lexical form and Syntax Chunk are associated with one another during the prime phase. Shortly afterwards, this association is still strong.

ACT-R version 6.0 does not provide a form of association learning. Clearly, the links that enable activation to spread between chunks must be acquired, *learned* somehow. In ACT-R (5.0), this mechanism is *association learning*. It occurs whenever a chunk i is requested (needed, event: N_i), while another chunk j is *in the context* (event: C_j), that is, it is in a buffer. The *empirical ratio*

$$E_{ji} = \frac{P_e(N_i|C_j)}{P_e(N_i)}$$

determines the positive adjustment of the association between i and j that results from the request of a chunk i when j is in a buffer (Anderson, 1993). The empirical ratio can be transformed to

$$E_{ji} = \frac{P_e(N_i \& C_j)}{P_e(C_j)P_e(N_i)}$$

which makes it obvious that it is the degree of dependence between events N_i and C_j that creates the link between the chunks.[8]

[8] Anderson has since removed the learning mechanism again from ACT-R (in version 6.0), presumably because the resulting actions were difficult to manage. This pragmatic move does not prevent us from arguing that association strengths leading to spreading activation must be acquired in some way.

That means the strength of learning is moderated by cognitive activity (e.g., productions matched). It predicts a decay of short-term priming over cognitive activity, or, if a fixed time span is associated with matching a production, time. Similarly a time-based decay can be assumed, also predicting decay of short-term priming over time. The exact nature of the association function is subject to future work.

The association function translates not just to the fan effect described earlier. It also explains associative priming (*cat* primes *dog*). In the context of the model described in the present chapter, association learning associates lexical material (*lexical forms*) with syntactic choices (*Syntax Chunks*).

There is an alternative explanation for the short-lived lexical boost effects. If semantic and syntactic material is retained in the buffer across utterances, it would spread activation, making repeated syntactic choices more likely. This implies strong lexicalization, as lexicalization means that syntactic and semantic material are retained together. Support for such a retainment view comes from coherence phenomena. Without discussing the details of coherence models, their essence is that sentences aim to continue the topic of a preceding sentence, placing referents presented late in the previous sentence early in the current one. Centering, a prominent theory of discourse coherence, posits: "Sequences of continuation are preferred over sequences of retaining; and sequences of retaining are to be preferred over sequences of shifting." (Grosz et al., 1995, page 214, rule 2). In that case, we would say that the short-lived enhancement causing strong short-term priming and lexical boost effects are based on the same semantic retainment effect that causes coherence. A testable prediction would be a correlation of the effects: sentences between which a topic is continued would be more likely to show short-term priming and lexical boost effects.

5.5 Evaluation

Commonly, ACT-R models are evaluated against the direct experimental data, which were collected under controlled conditions. For instance, a production priming experiment may be designed as follows: the participant would listen to a prime sentence (prime), whose syntactic construction is manipulated. Then, they describe a picture (target). For syntactic priming to be shown, the syntactic construction chosen in the target description would have to correlate with the choice of prime construction.

To benchmark an ACT-R model to such an experiment, a modeler would have to implement comprehension and production for all the experiment's materials, observe the reactions of the model and compare the sample of reactions to the sample obtained in the experiment with human subjects.

To compare a model to a corpus study, the methods used to evaluate a model must be adapted in several ways. A priming experiment controls the semantics of the utterances obtained, eliminating utterances where the subject did not produce a semantically correct target sentence. In contrast, a corpus study has little semantic control, and there is no semantic coding of sufficient detail available to allow us to simulate the production of all the data found in a corpus. Even with a large-coverage grammar in place, semantic specifications are typically not constrained enough to reproduce a substantial number of the original utterances, particularly in speech (as opposed to written language). Quite generally, realization systems are designed to make their jobs computationally tractable, as opposed to cognitively plausible. As a consequence, our computational model does not attempt to match the raw data. Instead, the model aims to produce the linguistic output known from syntactic priming experiments in order to reflect a set of priming effects. Its architecture complies with further results concerning incrementality. The underlying syntactic framework has been shown to be able to cope with a broad range of syntactic phenomena found in corpus data (Hockenmaier and Steedman, 2007).

We begin by examining whether the basic premise of memory access in ACT-R provides an explanation for adaptation effects. We apply the same methodology as was used for corpus analyses and experiments with human subjects, the benchmark being that the same effects emerge. In particular, we use a large set of (artificial) verbs to show that ACT-R's base-level learning mechanism produces long-term adaptation of syntactic structure and the inverse frequency interaction.

In a second study, we look at the actual ACT-R model, that is, include the language production algorithm. We simulate the production of a number of sentences, alternating double object (DO) and prepositional object (PO) realizations of the same semantics. We simulate a priming experiment, in which a subject produces either a DO or a PO variant and is then asked to choose a variant freely in a later target elicitation. The evaluation is designed to show that the simulated subject adapts to the chosen syntactic variant, and that there is little noticeable decay of this adaptation (after the first few seconds).

Examining lexical boost effects is more intricate. To explain the lexical boost, we rely on *association learning*. This learning defines the spreading-activation links between chunks. However, a function describing association learning is still a matter of research. We therefore do not assume a particular function, but we expect it to share a strong decay with base-level learning.

The realization model does not implement semantic activity. Any reasoning or semantic contextualization processes would likely be tied to the particular task of the dialogue, such as giving each other directions (Map Task) or chatting (Switchboard). To simulate short-term priming and the lexical boost, we could easily retain semantic (or other) chunks in a buffer and stipulate a manually set spreading-activation link between those buffers and syntactic material. However, evaluating the resulting effect would not yield any insights. With the mechanism being ad-hoc, we do not show psycholinguistic behavior as it emerges from the basic properties of ACT-R. The argument we make about short-term priming and lexical boost effects is, consequently, a qualitative one.

In the psycholinguistic literature, *models* are often viewed as *theories*. For instance, models of syntactic or other linguistic processes are qualitatively analyzed in terms of their potential to explain commonly known effects. It is in this sense that we make qualitative arguments in our evaluation. Theories also result in predictions, and we turn once again to our corpus data to test one of them in our third evaluation study. There, we examine the behavior of lexical boost with respect to general lexical material.

A central goal of cognitive modeling is to reduce effects to their cognitive bases. Consequently, we would hope that priming effects similar to the ones found empirically emerge from the lower-level cognitive principles defined by ACT-R. It is often possible to coerce the outcome of simulations to closely match experimentally obtained data: there are numerous parameters to adjust, but also many non-architectural choices to make in the implementation of a model. Thus, we do not only choose parameters using empirical data, but also concentrate on a qualitative evaluation against the known phenomena. That is, the question is not whether the model can replicate the experimental data precisely; the goal is for a well-motivated model to replicate the empirically observed effects.

5.6 Simulation 1: Learning and short-term priming

5.6.1 Method

In a pilot study using simulated language data, we estimated whether base-level learning can reproduce short-term priming effects. We simulated base-level learning on a set of artificial syntactic categories with varying frequencies. (In the model, such categories translate to syntactic choices typically used in priming studies, such as whether to realize a sentence with a ditransitive with PO (to) completion, or with a DO (NP-NP) completion. The categories represent the individual syntactic decisions as discussed in Chapter 4.)

A dataset was constructed from a set of category frequencies. Syntactic categories occurred randomly over a period of $50,000$ seconds, but the probability of their occurrence was defined according to their frequency, which was sampled from a Zipfian (power law) distribution. (Epochs are usually taken to be seconds.) For this period of time, we simulated a system encountering the rules at the assigned points in time. Each rule presentation increased the rule's base-level activation, and this activation decayed over time, both according to ACT-R's base-level learning function

$$B_i = \log \sum_{j=1}^{n} t_j^{-d} + \beta_i$$

where n identifies the number of presentations for chunk (syntactic rule) i and t_j the time since the j-th presentation. d is a decay parameter (set to 0.5, a value typical in ACT-R models) and β_i a parameter kept constant across all chunks.

This step provided the base-level activation of each syntactic rule. We would expect such activation as a result of the normal learning that occurs as the language processor is exposed to the rules.

We simulate the exposure to the rule that will be interpreted as the *prime*, as it would occur in any utterance in an actual corpus. Then, lag is simulated, ranging from 0 to 15 seconds, before activation levels are sampled. This applies the sampling methodology applied throughout the previous chapters to estimate short-term priming levels. Then, the same statistical methodology is applied, including the two-way interaction of the priming level ($\ln(\text{DIST})$ parameter) with rule frequency ($\ln(\text{FREQ})$). Activation level is the response variable. We use linear mixed-effects regression.

5.6.2 Results

We obtained a decay effect of ln(DIST) ($\beta = -0.25$, $p < 0.0001$) and also an interaction with rule frequency, ln(DIST):ln(FREQ) ($\beta = 0.0026$, $p < 0.0001$). (Intercept is $\beta = 10.33$, simple effect of ln(FREQ) $\beta = 0.0165$, $p < 0.0001$.) Thus, the simulation results in syntactic priming, which is stronger for low-frequency rules.

5.6.3 Discussion

We see that the basic priming effect can be explained by the learning function, as one would expect from a function that prescribed logarithmic decay. What is more interesting is that base-level learning explains the frequency interaction as well: low-frequency items consistently show more priming, as seen in the corpus studies.

This result, taken on its own, appears to be compatible with the "priming as implicit learning" hypothesis. According to this model, priming is the result of improved accessibility of syntactic constructions after they have been learned. The difference in short-term priming and long-term adaptation is due to the initially strong decay of the learning effect. Activation is high shortly after presentation (use) of a syntactic construction, but decays strongly within a few seconds to converge to a plateau that is higher than before the presentation. In the context of ACT-R, we would speak of enhanced memory access, which leads to greater reliability and also faster access. In other words, the ACT-R model would be more likely to choose the prime constructions, but it would also be quicker to do so. The nature of the syntactic structure accessed is not relevant for this argument. Learning could work with lexicalized, combinatorial items retrieved from memory, but it could also apply to sequences of more general, abstract categories (which we have refuted on other grounds).

The priming-as-learning hypothesis would, obviously, also predict an effect of frequency on long-term learning. However, this effect can be expected to be very small, which explains why such an effect was not found empirically in Experiment 7.

5.7 Simulation 2: DO/PO production adaptation

In this section, we describe a simulation in which the ACT-R model simulates the generation of sentences. We elicit DO and PO primes by forcing the model to choose a particular syntactic structure for given semantics. Then, other semantics are given, and the model is free to choose any syntactic variant. The simulation is set up to be similar to an actual priming experiment.

We aim to show that the model exhibits long-term adaptation similar to what we have seen in the corpus data.

5.7.1 Method

The model was to generate sentences with semantics equivalent to *The doctor gave a flower to the policeman*. Two conditions were used: a prime condition (PRIMED$_1$), and a control condition (PRIMED$_0$). In both conditions, the model was first given semantics to generate from. We alternated a constraint that forced the model to either choose prepositional object (primed) or double object constructions (control condition) in the prime sentence. The model was free to choose different lexical realizations (a number of synonymous noun phrases were given for the arguments).

We then simulated a random lag (60–$1,000$ seconds, uniformly distributed) with no activity in order to give any short-term effects a chance to decay.[9] In each condition, 100 repeated trials were sampled. Then, a target sentence was elicited (semantics equivalent to those of *The cheerleader offered the seat to a friend*). This time, the model was free to chose a syntactic variant, i.e., *The cheerleader offered the seat to a friend* or *The cheerleader offered a friend the seat*.

We did not alter the items. Experimental designs would use a number of different stimuli, but given the model implemented in ACT-R, the only source of variation is the general noise added to the system and the preexisting, corpus-acquired activation of the lexical material and their links to Syntax Chunks. Therefore, using different stimuli would not yield any more sensible results.

We report the results of a χ^2 test. Note that these results are intended to generalize beyond the present activation noise and the model's choices, but not beyond prime/target semantics and verbs.

[9]Simulating the production of intervening sentences as opposed to 'quiet' lag would not affect the activation of those particular syntactic types as long as the sentences do not make use of them. That is, the presentation of other syntactic material does not influence the activation of the PO type.

5.7.2 Results

In the control condition, the model produced prepositional object (PO) constructions in 23% of the trials. In the PO priming condition, the model produced PO constructions in 38% of all trials. For the given semantics and verbs, we see priming of prepositional object constructions ($\chi^2 = 4.6, p < 0.05$). Prime-target distance had no reliable effect on repetition probability (by GLM, $\beta = 0.0005$, $p = 0.65$).

5.7.3 Discussion

The model showed long-term adaptation, where we define *long-term adaptation* as increased repetition of argument structure at least 60 seconds after the prime.

The effect of prime-target lag is present in theory, given the decay in underlying base-level activation, which follows from ACT-R's base-level learning function. The decay, however, is too small to be detected assuming realistic numbers of trials and standard noise levels, as the statistical analysis shows. This is compatible with empirical corpus studies that involve large prime-target lags, e.g., Jaeger and Snider (2007), in which no effect of distance could be shown.

5.8 Simulation 3: Cumulative priming

Jaeger and Snider (2007) show data on complementizer omission that suggest that priming is cumulative. They find that the more clauses with a full *that* complementizer speakers use, the more likely they become to choose an optional *that* complementizer at a later point. Consequently, the more reduced clauses one speaker uses, the less likely their interlocutor is to use a full *that* construction. A similar analysis applies to the number of passive voice constructions predicting future passives across speakers. In this simulation, we attempt to examine whether such cumulative priming is replicated by the ACT-R model, albeit in the different syntactic context of PO priming.

5.8.1 Method

As in Simulation 2, we elicited target sentences for given semantics. This time, we ran two simulations, one using DO primes, and another one using PO primes. We manipulated the number of prime sentences, which ranged from 0–25 (coded as NUMPRIMES). 14 trials were carried out for each number of prime sentences. A randomized pause was introduced between the prime sentences (5-30 seconds, uniformly distributed), and a random lag between primes and targets of 60-$1,000$ seconds (uniformly distributed), as before. A total of $1,300$ trials were produced in each simulation.

A generalized linear mixed effects model was fitted, with a response variable coding repetition (1) or no repetition (0) of the chosen prime structure in the target, with the number of primes as predictor. As before, a random covariate grouped by items was entered to account for repeated measurements (repeated for each number of primes).

5.8.2 Results

Figure 5.6 shows the repetition probabilities resulting from the PO and DO simulations. The statistical analysis included all trials with numbers of prime sentences above 1, i.e., we only look at cases of priming, which is conservative, as it excludes the relatively strong contribution of the no-priming control case (leftmost data point in the Figure).

For PO primes, we find a steady increase of repetition probability with increasing number of primes ($\beta = 0.022$, $p < 0.01$). For DO primes, we fail to find evidence for such a correlation ($\beta = 0.005$, $p = 0.59$).

5.8.3 Discussion

Long-term adaptation, according to our model, is cumulative. For preposition-object primes, the effect appears to be weaker than that of *that* complementizer deletion or passive constructions found empirically in Jaeger and Snider's (2007) study.

The failure to find reliable cumulativity of DO adaptation can be attributed largely to its higher relative probability (either relative to PO constructions, or overall), which is the only relevant difference between PO and DO in our model.

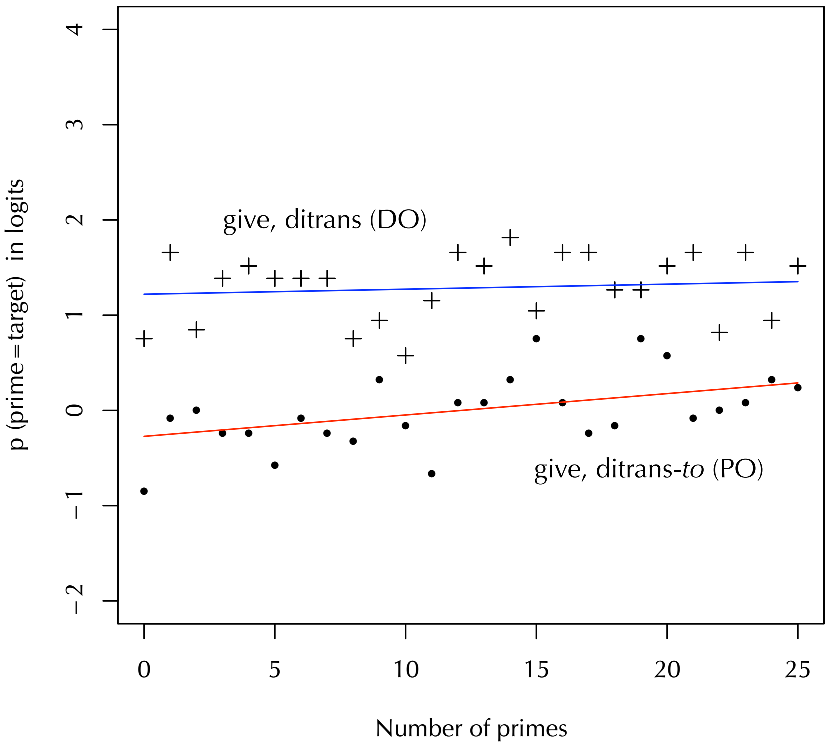


Figure 5.6: Cumulative priming: Proportion of matched targets for 0–25 primes (x-axis). Dots represent PO primes (and proportion of PO targets), crosses represent DO primes (and proportion of DO targets). The slopes of the two fits indicate cumulative priming, they exclude the no-priming condition (0 primes). Probability scale (y) in logits.

5.9 Experiment 14: General lexical boost as the resulting prediction

The previous evaluation steps are based on a form of simulation. Here, we turn once again to an empirical experiment with corpus data in order to test a prediction that the model makes. This prediction arises from the spread of activation from lexical forms to syntax chunks: The repetition of lexical forms will boost the priming effect as associations between the items in the buffer are learned. We know that the lexical boost interaction increases priming of the structure of a verbal phrase if the verb is repeated, and likewise for noun phrases, if the noun is repeated: the lexical boost is incurred by head repetition. The model's prediction exceeds this boost: it predicts a boost whenever lexical material is repeated, whether it is the head or not. Commonly, heads introduce a distribution of the possible syntactic variants of the structure they govern: For verbs such as *give* or *offer*, the DO realization is more likely than the alternative (PO) (see Figure 5.4, p. 128). This distribution is called *frame selection bias*.

We cannot explicitly simulate this kind of boost with the model, as the current ACT-R framework (version 6.0) does not specify association learning, which modulates the spreading activation from lexical to syntactic chunks. Any direct simulation of spreading activation would be ad-hoc: it would yield neither surprising nor particularly convincing results, especially given that a newly defined dynamic association learning function would not be motivated by independent empirical data on non-linguistic forms of cognition. Nevertheless, we can test the prediction arising from the lexical influence on priming directly with our dialogue data.

Such a test also has its value independently of the evaluation of the model. In structural priming, the tendency to repeat the structure of a verbal phrase is enhanced if the head verb is lexically repeated, and similarly for nouns and noun phrase structure (Cleland and Pickering, 2003). To test the prediction of our model, we now turn to corpus data again. We ask two questions: Is it the frame selection bias tied to the head that causes this boost? And: Can only heads lead to a short-term lexical boost? Language production models that select the head and then plan the structure of the constituent (e.g., Pickering and Branigan, 1998) only predict the head to boost structural priming, while models of incremental production also allow other lexical material to boost priming (e.g., de Smedt and Kempen, 1991; Hoffman, 1995; F. Ferreira and Swets, 2002).

Our model predicts not only the empirically known lexical boost, but also that the repetition of material other than heads can boost priming. This boost effect emerges from the spread from general lexical and semantic material present in the Goal buffer to syntactic material as it is being retrieved. Such lexical-syntactic associations are acquired independently of whether lexical material happens to serve as head of a phrase.

5.9.1 Method

Examining the role of heads in a variety of structures necessitates large datasets. We annotated the Switchboard corpus with lexical heads using Collins's (1999) head-finding rules.

The method we use to detect short-term priming was described in Chapter 2. As before, priming is reflected in the (negative) effect of $\ln(\textsc{Dist})$, i.e., we see a decay of repetition probability over time. Examining repetition decay controls for any frame selection bias, which occurs when head lexical forms prefer certain structures (the repetition of head words is indeed positively correlated with the repetition of the full syntactic structure in our data).

We are interested in whether heads or lexical material in general can boost priming. We only include data from cases where at least one word, but not all words were repeated between prime and target. We included in our model head repetition as binary factor and also a measure of word repetition: the proportion of repeated words between prime region and target constituent. (If all the words in the target region also occur somewhere in the prime region, this proportion would be 1.)

5.9.2 Results

As in previous experiments on Switchboard in this thesis, we obtain an effect of $\ln(\textsc{Dist})$ (in seconds, $\beta = -0.31$, $p < 0.001$), indicating priming, which also shows stronger priming for less frequent rules. A lexical boost for heads would show up as a negative interaction of head repetition with $\ln(\textsc{Dist})$, i.e., head repetition would strengthen the decay. However, we see no such interaction ($\beta = 0.045$, $p = 0.29$). Instead, we observe an interaction of word repetition with $\ln(\textsc{Dist})$ ($\beta = -0.158$, $p < 0.001$). This suggests that it is any lexical repetition that boosts priming rather than specifically head repetition.

5.9.3 Discussion

While consistent with the literature that finds a lexical boost for head repetition, the results generalize the lexical boost effect to other lexical material. Once the (constant) frame selection bias is accounted for, heads play no special role compared to other lexical material. This supports models that analyze short-term priming as a lexical and/or semantic effect.

The exact mechanism of the effect may be related to association learning. Association learning increases the links between lexical material (in a buffer) and a syntactic construction. This increase in link strength occurs with any lexical material present in the buffer (not just heads) and any syntactic nodes that are retrieved while this material is still in the buffer. Stronger links lead to stronger lexical and syntactic associations. A reasonable underlying assumption is that learned association strengths decay in a way similar to base-level activation. The greater the distance between prime and target, the smaller is the effect of lexical repetition, because at greater distances, the learned association will have decayed. Thus, this mechanism explains how spreading activation patterns are acquired and why lexical boost occurs over a short period of time after the prime.

5.10 Comparison with other models

A number of contributions have been made towards the development of a comprehensive model of human language production. Levelt's (1989) model of "Speaking" assumes several autonomous processing components, which do their well-delineated work autonomously. Levelt's model provides a comprehensive architecture (rather than a model of priming in the syntactic realization process as in the one presented here). The model produces speech incrementally (like ours). It distinguishes lexical and syntactic encoding more than lexicalized models. The distinction between lemma (the meaning of a lexeme) and form (full form and syntactic properties) holds for the model presented here as well. Information can be stored on the lexical level, i.e., in the Lexical Form chunks, or it can be represented in Syntax Chunks. However, our model focuses closely on a description of how syntactic information directs the combination of words to phrases and sentences, rather than specifying the overarching architecture.

Kempen and colleagues have developed *Performance Grammar* (de Smedt and Kempen, 1991; Vosse and Kempen, 2000; Bond, 2005), modeling lexical-syntactic processes in comprehension and production. In their model, the retrieval of information from memory shares properties with retrieval in ACT-R, as does the merging of information, for instance during lexical retrieval and syntactic attachment. That is, such *unification* processes are similarly non-recursive. A crucial difference, however, is how syntactic composition takes place. Kempen's model presupposes a network of interconnected lexical nodes, i.e., it goes beyond our assumption of a very limited working memory during incremental generation (made possible by CCG's combinatory properties). Essentially, this resembles models that assume temporary storage in memory rather than in data structures equivalent to ACT-R's buffers. The same applies to Lewis and Vasishth's (2005) model discussed below.

Memory access is also the focus of a model being developed (unpublished, but see Badecker and Lewis, 2007). While based on ACT-R, it does not employ ACT-R's production rules, but concentrates on explaining speech errors in production, which are, in Badecker and Lewis's simulations, the result of cue-based memory retrievals.

Roelofs; Roelofs's (1992; 1993) network model of language production specifies an encoding of syntactic preferences for verb forms that is similar to our spreading activation account within ACT-R. This model has been extended by Pickering and Branigan (1998) to form a theory of syntactic sentence production. There, features such as tense, aspect and number are encoded separately. As follows from syntactic priming effects (including those shown in their experiments with written language), syntactic representations are separate from the word form. Syntactic variants are encoded as combinatorial categories such as NP, NP (forcing a DO construction) and NP, PP (forcing a PO construction). However, their model does not store syntactic knowledge in lexicalized categories, but keeps a separate representation of a word category (such as *Verb*). Priming follows from a pre-activation of the combinatorial categories. The network model is seen as a theory and motivated qualitatively through priming experiments. Pickering and Branigan (1998) can be credited as an early use of syntactic priming to create a cognitively plausible account of the syntactic production process.

Lewis and Vasishth (2005, L&V) present an ACT-R model of language comprehension, in which temporary analyses of the partial sentence are stored in and retrieved from memory as it is being analyzed. Comprehension difficulties are explained through the decay of accessibility of stored information, as opposed to a general cost associated with temporary storage. Their model is interesting in this context given that comprehension and production systems can be assumed to share information stored in memory, i.e., lexical and probably syntactic knowledge. L&V's model differs from the model presented in this chapter. L&V store syntactic knowledge as production rules, as they make clear in their article: "The model ...

assumes that much grammatical knowledge is encoded procedurally in a large set of quite specific production rules that embody the skill of parsing. The model thus posits a structural distinction between the representation of lexical knowledge and the representation of abstract grammatical knowledge."

This view has much conceptual appeal. In the context of ACT-R, however, it remains to be shown how syntactic knowledge in such a model can transfer from comprehension to production, given that the rules themselves are likely to encode such an algorithm.

ACT-R defines forms of rule-learning: on the sub-symbolic level, this involves learning a rule's *utility* and thus learning to choose the best rules. ACT-R's current framework does not predict a decay of such rule preferences. Thus, L&V would not be able to account for decay effects in syntactic priming using ACT-R's procedural memory. Furthermore, lexical boost effects require links from lexical to syntactic knowledge. Such links are symbolic in L&V's model and do not explain the probabilistic nature of priming and lexical boost effects.

Chang et al. (2006) present a connectionist model called the *Dual Path Model* that is primarily concerned with language acquisition (see also Section 4.2, p. 80). It is trained using artificial language data (our model only sees "training" in the form of rule frequencies). Similar to part of our model, the Dual Path Model likens structural priming to learning. However, it learns transitions of abstractions of words, similar to part-of-speech categories (or perhaps higher-level syntactic structure). Our model adapts the base-level activation of lexicalized, combinatorial syntactic information.

Connectionist models are difficult to compare to models that combine symbolic and sub-symbolic explanations such as those within the ACT-R framework. Such a comparison is best attempted using explanatory power with respect to empirically known effects. The Dual Path Model explains priming phenomena including the inverse frequency interaction. However, it is unclear whether such a model can explain the sensitivity of sequence priming to syntactic structure, which was shown empirically (Experiments 9 and 11), while priming affects syntactic structure directly in our model. This remains a theoretical claim. Syntactic variation in the simulation data in both the Dual Path Model and our model is too limited to estimate the priming of constituents and distituents.

More importantly however, Chang et al.'s (2006) "syntactic route" alone, or the actual implementation of this model do not explain the qualitative differences between short-term priming and long-term adaptation. However, we see the Dual Path Model as an explanation of syntax acquisition and long-term adaptation (learning) effects. Chang et al. (2006) are early proponents of a multi-tiered explanation of the production process that involves a *meaning system*, influencing the *sequencing system* as it produces the output.

5.11 Conclusion

This chapter combined a number of results presented in this thesis with those obtained by others. The intention was to take a step towards the ultimate reason why psycholinguistic studies are carried out: to specify an accurate model of the human language faculty.

What a model generally cannot demonstrate is that it is the only viable account. We can safely say that no theory of a syntactic process can accurately reflect the structures and processes involved in the human language processing mechanism. Models are a concrete instantiation of theoretical conclusions, usually based on empirical data. They necessarily simplify and omit, but as any work, they always aim to be a contribution that merely constitutes *the next step*.

We have implemented a model that generates simple, English sentences. Naturally, its linguistic abilities in terms of language generation are limited and focus on an alternation of sentence structure that has traditionally been used to show priming effects. Its underlying syntactic framework, however, is flexible enough to describe a wide range of syntactic phenomena as they occur in natural text and dialogue, as in the two datasets that have featured in this thesis. The remarkable feature of the syntactic formalism and the algorithm is that it supports incremental language production without the need to store large amounts of information during the process. This is compatible with results from the psycholinguistic literature that point to an optionally incremental language production process. It sits particularly well with a view that casts language production as a process that underlies general cognitive principles, such as the ones postulated by the ACT-R framework.

The cognitive framework defines a number of validated properties of rule-governed, serial control and cue-based, parallel memory access. It specifies learning and contextualization. These independently motivated constraints lead to the modeled priming effects through our specific language production algorithm and our specific syntactic framework. The emergence of the priming phenomena does not depend on every precise aspect of the model: for instance, the phenomena may be replicable with different syntactic assumptions. The key idea, however, is that syntactic priming arises from lower-level properties of cognition that are not specific to language processing.

In simulations, the model showed syntactic adaptation in the long term, as well as short-term priming and its inverse frequency interaction. The frequency effect we observe empirically emerges in the model from a specific property of ACT-R's learning mechanisms: rarely accessed information with a low initial base-level activation is boosted more strongly through presentation than is common information with a high initial activation. Recall that short-term priming results from both base-level and association learning. We show how base-level learning is affected by initial activation, and we assume association learning to behave in the same way (the "fan" effect can be seen as a result of this property Anderson, 2007). This explanation of frequency effects as a result of learning mechanisms refers to *surprisal* a cause of greater adaptation. Surprisal describes the violation of expected structure as the cause of learning (see also Rescorla and Wagner, 1972). Cue-based memory retrieval in our framework limits the build-up of expectations to the ones that can be derived directly from material held in a buffer when the syntactic decision is made. There is no higher-order reasoning (beyond associations) that would lead to a greater surprisal and greater p riming.

As a theory, the model explains the short-lived lexical boost that is associated with priming. It explains the lexical boost of general lexical or semantic material, a prediction which we tested using a corpus. The same mechanism can also provide a post-hoc explanation of the short-term priming boost found in task-oriented dialogue, as opposed to spontaneous conversation.

By design, priming in the model applies to syntactic structure, in particular to combinatorial categories as syntactic descriptions of subcategorization. The empirical rationale for this was discussed in Chapter 4. We argue that this model gives a concrete explanation for such effects, an explanation that is missing in prior models of language production.

The model implements *flexible incrementality*. It composes syntactic structure incrementally by default, even though planning is possible. The empirical motivation for this was presented in Experiment 12. However, the exact extent of incrementality has not been investigated using structural priming. The model's ability to plan is limited by cognitive resources, with incremental production being the most efficient way to construct the syntax of sentences. Conversely, the model predicts that sentences which require non-incremental production, will take longer to produce and yield more errors. A possible example of such non-incremental constructions (in CCG) would be object relative clauses.

The central argument of this chapter is to demonstrate that syntactic priming can be explained as a two-level learning effect: the learning of individual syntactic representations, and the acquisition of links between the same syntactic representations and lexical/semantic material. Syntactic priming is neither due to a specialized pre-activation property of individually memorized information, nor is it due to a single implicit learning effect. Syntactic priming emerges from two learning effects, which are, to a large degree, understood as general cognitive principles.

Chapter 6

Conclusions

This thesis presents a range of psycholinguistic results derived from the analysis of language corpora in order to arrive at a model of language production that explains structural priming effects.

6.1 Contributions

The conclusion we have arrived at is that syntactic priming is the result of basic, cognitive learning principles. Syntactic priming arises through the modulation of memory retrieval. The basic mechanism that allows speakers to learn to produce linguistic structure is what causes syntactic priming. However, we identified a second effect, occurring early after a prime. This short-term priming is primarily caused by a second type of learning: association learning. We propose that associations are acquired between semantic, lexical and syntactic choices. These associations bias syntactic choices. The decay of such semantic-lexical-syntactic associations causes short-term priming.

This conclusion is based on a range of results, both from the literature and from corpus experiments guided by the search for the most plausible cognitive explanation of syntactic priming. We introduce a method to measure short-term priming in syntactically annotated corpus data. It leverages the rapid decay and is thus affected by neither lexical-syntactic bias nor chance repetition. Even though the decay has been a factor in other work (e.g., Szmrecsanyi, 2005), most other studies were concerned primarily with long-term adaptation. To our knowledge, our work is the first to extensively use decay to contrast priming in different conditions. We show that short-term (decay-based) syntactic adaptation exists in naturalistic language, that is, in corpora. We also demonstrate long-term syntactic adaptation in the same data sets using a different method. This method quantifies the repetition of syntactic decisions between the first and second halves of each dialogue, contrasting it with repetition between different dialogues in the same corpus.

The two methods play a crucial role throughout the investigation of priming in terms of its function in dialogue and its locus in syntactic language production. The first set of priming results concerns properties of human-human dialogue. Priming has been hypothesized to be instrumental in speakers' mutual understanding of how situated discourse refers to the environment. According to the Interactive Alignment Model (Pickering and Garrod, 2004), speakers arrive at a common situation model through a cascade of alignment at lexical, syntactic and semantic levels. Using task-oriented dialogue in the HCRC Map Task corpus (Anderson et al., 1991), we show that those speakers who align better also perform better at a given task. To our knowledge, this represents the first large-scale empirical verification of a prediction arising from the Interactive Alignment Model. The to-date unknown correlation between syntactic adaptation and task success is also exploited in a machine-learning based algorithm. To evaluate this, we define two tasks, which involve the prediction of task success from either the initial portion of dialogues, or from the whole dialogues. These tasks can also be addressed using different methods.

Crucial for the question of cognitive provenance of the priming effects, the priming–task success correlation is only found for long-term adaptation, but not for short-term priming. Thus, the two effects are qualitatively different and cannot result from the same cognitive basis.

Comparing syntactic priming between the two types of dialogue (task-oriented in Map Task, and spontaneous conversation in the Switchboard corpus, Marcus et al., 1994), we find that short-term priming is stronger in task-oriented dialogue than in spontaneous conversation. We argue that short-term priming is primarily related to the strengthening of semantic-syntactic associations. Thus, strong semantic activity and the persistence of discourse objects throughout the dialogue (as in Map Task as opposed to Switchboard) leads to stronger syntactic adaptation. We argue against the possibility that short-term priming is a strategically modulated effect and argue that it is largely mechanistic.

We argue that psycholinguistic models need to incorporate linguistic accounts of syntax in order to plausibly explain the structural variety present in natural language. With the two methods to measure short-term and long-term adaptation, we combine grammatical accounts developed in computational linguistics with psycholinguistic processing hypotheses. We generalize the priming effect from selected syntactic constructions to a broad variety of syntactic micro-decisions. This general model applies priming to single phrase-structure rules. We examine this assumption and extend the statistical models to cover arbitrary sequences of lexical categories as well as complex lexical and phrasal categories from Combinatory Categorial Grammar (CCG, Steedman, 2000).

We find support for the CCG-based model as well as for predictions arising from CCG. Specifically, we devise and support the *flexible incrementality hypothesis*, which postulates that speakers can compose the syntactic structure of their utterances in a more or less incremental fashion. In CCG, the degree of incrementality affects the structure of each derivation. Structural priming occurs not only for the case of *planning ahead*, where derivations are created before speaking begins, but it also occurs when we assume maximally *incremental* derivations, where sentence structure is planned after the first words have been spoken. An accurate language production model will incorporate an incremental process and may allow for differences in transient structures depending on the level of incrementality. We also find corpus evidence of the equivalence of two types of structures that these formalisms predict to be equivalent: Structures of the first kind are transient and built during the syntactic production or parsing process. Structures of the second kind are lexical, that is, they are retrieved from memory.

Based on the notion of *distituency*, we find evidence for a structural basis of syntactic priming: a non-structural account cannot explain why priming of word category sequences is weaker at structural boundaries, as we have found in an experiment.

With these results we demonstrate how syntactic priming can provide further evidence of a syntactic model. Priming reveals properties of syntactic processes.

We propose a computational model of language production based on these results. The basic premise of this model is that language production is an instance of a general cognitive process. While there may be a dedicated language processing mechanism, the central point is that the algorithmic devices at hand are the same.

Anderson et al.'s (2004) ACT-R model is, to our knowledge, the best-tested and most stable modeling framework.

The model demonstrates that structural adaptation effects can be replicated by two omnipresent effects that form ACT-R's foundations: base-level learning and spreading activation. Base-level learning explains strongly decaying short-term priming as well as a cumulative, long-term adaptation effect. Spreading activation is a mechanism of contextualization, which explains lexical boost effects present in short-term priming as the result of temporarily present lexical and syntactic information, which facilitates the retrieval of related syntactic material.

The model we present is well-motivated by the empirical results, both our own and those of other researchers. The model is justified in terms of its linguistic, grammatical basis, by its algorithm providing flexible incrementality, and also by way of its independently motivated cognitive architecture. The technique used to measure short-term priming in corpora yields equivalent results for the model, specifically for priming, its interaction with rule frequency, cumulativity of priming. A prediction borne out of the model's short-term priming mechanism (association learning) was that the repetition of general lexical material boosts short-term priming. The hypothesis held true for corpus data.

6.2 Future work

The ACT-R model of language production is relatively close to an end-to-end explanation of the syntactic production process. It neither picks out a particular sub-problem of language production, nor does its architecture pick a sub-problem of syntactic priming. We know of no impediment *in principle* that would prevent us from adding further grammatical coverage.

Even a more extended form of such a model will not be able to generate natural language sentences found in a corpus given the limited form of semantics available through parsing. Such a form of end-to-end coverage isn't necessary in order to give a detailed account of production. Even if the syntactic process is constrained to follow exactly the structural analyses found in a corpus, the model will still define a subset of the memory accesses, the structural descriptions and procedural steps necessary to produce natural language. Thus, the model allows us to operationalize the language production task. Its predictions are then testable. There are several areas in which further investigation appears worthwhile:

Flexible incrementality allows speakers to vary the degree to which sentence structure is planned ahead. The model predicts optimized syntactic construction processes for the incremental variant, while other constraints may prompt a speaker to opt for non-incremental sentence planning. The correlation between cognitive load during production and incrementality of planning could be tested experimentally.

Associative links between different chunks cause the lexical boost. The exact formulation of a learning function for these is currently less clear; most importantly, the current probabilistic definition lacks an explanation for the dynamic adaptation of such links. Assuming the model is right, the lexical boost allows us to estimate association learning from corpora.

The role of information structure, coherence and syntactic priming: As pointed out earlier, association learning is only one explanation of lexical boost in syntactic priming. Coherence may be required to maintain lexical boost effects in syntactic priming. It is generally unclear whether short-term priming shows decay over time, syntactic or semantic activity. A semantic account, contrasting with one based on association learning, would predict that less coherent pairs of sentences show less priming than more coherent ones: for instance, priming would be weaker when the topic shifts between the sentences.

Priming and alignment on other levels: The memory-retrieval based model we have arrived at in this thesis may be a viable explanation for priming at other levels, such as lexical and phonological priming. A more detailed language processing account would incorporate comprehension as well as production. With such a model, we may be able to explain mutual adaptation on prosodic levels and the development of alignment between interlocutors over the course of dialogues. Such alignment is likely to be modulated by social factors such as affect. Some interlocutors may serve as a stronger source of contextualization, while others, with whom a speaker does not endeavor to associate, may be deliberately kept further away. An alternative hypothesis would state that all alignment effects are based on mechanistic priming and that social relationships between interlocutors are not afforded any influence on memory retrieval.

Priming as a paradigm to detect syntactic structure: Because priming is sensitive to syntactic structure, we can employ the short-term priming measure of decay to motivate a set of constituent boundaries, or lend support to particular variants of analysis over others. Syntactic priming may well be a tool allowing us to investigate more specific syntactic claims than those that have formed the basis of Chapter 4.

In general, the cognitive model, but also the statistical regression models presented in this thesis employ approaches from computational language modeling and cognitive science to examine psycholinguistic hypotheses. We draw from linguistic research that endeavours to achieve broad coverage and cognitive plausibility and evaluate its psycholinguistic predictions. We warn against too much enthusiasm about large data sets. Corpus studies are appropriate whenever potentially confounding factors can be explicitly modeled, or when the main effect under discussion is not masked. We have argued that this is the case for the present studies. The use of corpus data has proved beneficial to this inquiry: it allowed us to generalize priming effects, and it gave us a chance to test a number of hypotheses (we report all tests conducted within this research program, including those that did not yield significant results). It is hoped that our corpus work points out a novel avenue for computational psycholinguistics.

Bibliography

A. Agresti. 2002. *Categorical data analysis*. Wiley, New York, NY.

A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC Map Task corpus. *Language and Speech*, 34(4):351–366.

J. R. Anderson. 1974. Retrieval of prepositional information from long-term memory. *Cognitive Psychology*, 6:451–474.

J. R. Anderson. 1976. *Language, Memory and Thought*. Erlbaum, Hillsdale, NJ.

J. R. Anderson. 1990. *The adaptive Character of Thought*. Erlbaum, Hillsdale, NJ.

J. R. Anderson. 1993. *Rules of the Mind*. Erlbaum, Hillsdale, NJ.

J. R. Anderson. 2007. *How can the human mind occur in the physical universe?* Oxford University Press, Oxford, UK.

J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Quin. 2004. An integrated theory of mind. *Psychological Review*, 111:1036–1060.

J. R. Anderson and G. H. Bower. 1973. *Human Associative Memory*. Winston and Sons, Washington, D.C.

R. H. Baayen, D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *To appear in Journal of Memory and Language, special issue on Emerging Data Analysis Techniques*.

W. Badecker and R. Lewis. 2007. A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue- based retrieval. Abstract. Abstract, 20th Annual CUNY Conference on Human Sentence Processing.

D. M. Bates. 2008. Linear mixed model implementation in lme4. http://cran.r-project.org/web/packages/lme4/index.html.

J. K. Bock. 1986a. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12:575–586.

J. K. Bock. 1986b. Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.

J. K. Bock and Z. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129:177–192.

J. K. Bock and H. Loebell. 1990. Framing sentences. *Cognition*, 35:1–39.

K. Bock, G. S. Dell, F. Chang, and K. H. Onishi. 2007. Persistent structural priming from language comprehension to language production. *Cognition*, 104(3):437–458.

A. H. Bond. 2005. A psycholinguistically and neurolinguistically plausible system-level model of natural-language syntax processing. *Neurocomputing*, 65-66:833–841.

H. P. Branigan, M. J. Pickering, and A. A. Cleland. 1999. Syntactic priming in language production: Evidence for rapid decay. *Psychonomic Bulletin and Review*, 6(4):635–640.

H. P. Branigan, M. J. Pickering, and A. A. Cleland. 2000a. Syntactic co-ordination in dialogue. *Cognition*, 75:B13–25.

H. P. Branigan, M. J. Pickering, J. Pearson, J. F. McLean, and C. Nass. 2003. Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society (CogSci)*.

H. P. Branigan, M. J. Pickering, A. Steweart, and J. F. McLean. 2000b. Syntactic priming in spoken production: Linguistic and temporal interference. *Memory and Cognition*, 28(8).

C. Brockmann, A. Isard, J. Oberlander, and M. White. 2005. Modelling alignment for affective dialogue. In *Workshop on Adapting the Interaction Style to Affective*

Factors at the 10th International Conference on User Modeling. Edinburgh, UK.

P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479.

M. Brysbaert, W. Fias, and M.-P. Noel. 1998. The Whorfian hypothesis and numerical cognition: Is "twenty-four" processed in the same way as "four-and-twenty"? *Cognition*, 66(1):51–77.

J. Carletta, S. Dingare, M. Nissim, and T. Nikitina. 2004. Using the NITE XML toolkit on the Switchboard corpus to study syntactic choice: A case study. In *Proceedings of 4th Language Resources and Evaluation Conference (LREC)*. Lisbon, Portugal.

F. Chang, G. Dell, and K. Bock. 2006. Becoming syntactic. *Psychological Review*, 113(2):234–272.

N. Chomsky. 1957. *Syntactic Structures*. Mouton, The Hague, Netherlands.

K. W. Church. 2000. Empirial estimates of adaptation: The chance of two noriegas is closer to $p/2$ than p^2. In *Proceedings of the 18th conference on Computational linguistics (COLING)*, pages 180–186. Saarbrücken, Germany.

H. H. Clark and T. Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology*, 37:201–242.

S. Clark and J. R. Curran. 2004. Parsing the WSJ using CCG and log-linear models. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. Barcelona, Spain.

A. A. Cleland and M. J. Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49:214–230.

M. Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

M. J. Crawley. 2005. *Statistics. An Introduction using R*. Wiley, New York, NY.

M. F. Damian and R. C. Martin. 1999. Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory and*

Cognition, 25(2):345–361.

K. de Smedt and G. Kempen. 1991. Segment grammar: A formalism for incremental sentence generation. In C. L. Paris, W. R. Swartout, and W. C. Mann, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pages 329–349. Kluwer, Boston, MA.

A. Dubey, F. Keller, and P. Sturt. 2005. Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, pages 827–834. Vancouver, Canada.

J. Eisner. 1996. Efficient normal-form parsing for combinatory categorial grammar. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 79–86. Santa Cruz,CA.

J. L. Elman. 1990. Finding structure in time. *Cognitive Science*, 14(2):179–211.

D. Estival. 1985. Syntactic priming of the passive in english. *Text*, 5:7–21.

F. Ferreira and B. Swets. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46:57–84.

V. Ferreira and K. Bock. 2006. The functions of structural priming. *Language and Cognitive Processes*, 21(7-8):1011–1029.

V. S. Ferreira. 1996. Is it better to give than to donate? syntactic flexibility in language production. *Journal of Memory and Language*, 35:724–755.

V. S. Ferreira. 2003. The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, 48(2):379–398.

M. Ford. 1982. Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 797–827. MIT Press, Cambridge, MA.

S. Garrod and A. Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, 27:181–218.

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

S. T. Gries. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4):365–399.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 2(21):203–225.

M. Halliday. 1967. *Intonation and Grammar in British English*. Mouton, The Hague, Netherlands.

B. Hart and T. R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Paul H Brookes, Baltimore, MD.

R. J. Hartsuiker, S. Bernolet, S. Schoonbaert, S. Speybroeck, and D. Vanderelst. 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, 58(2):214–238.

R. J. Hartsuiker, H. Kolk, and P. Huiskamp. 1999. Priming word order in sentence production. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 52A(1):129–147.

R. J. Hartsuiker and H. H. Kolk. 1998. Syntactic persistence in dutch. *Language and Speech*, 41:143–184.

R. J. Hartsuiker and C. Westenberg. 2000. Persistence of word order in written and spoken sentence production. *Cognition*, 75B:27–39.

J. Hockenmaier and M. Steedman. 2002. Generative models for statistical parsing with Combinatory Categorial Grammar. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 335–342. Philadelphia, PA.

J. Hockenmaier and M. Steedman. 2007. CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn treebank. *Computational Linguistics*, 33(3):355–396.

B. Hoffman. 1994. Generating context-appropriate word orders in turkish. In *Proceedings of the 7th International Workshop on Natural Language Generation*. Kennebunkport, ME.

B. Hoffman. 1995. Integrating free word order, syntax, and information structure. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, Dublin*, pages 245–252. Morgan Kaufmann, San Francisco, CA.

V. M. Holmes. 1988. Hesitations and sentence planning. *Language and Cognitive Processes*, 3:323–361.

T. F. Jaeger. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University.

T. F. Jaeger and N. Snider. 2007. Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences*, 3(1):26–44.

M. Johnson and E. Charniak. 2004. A tag-based noisy channel model of speech repairs. In *Proceedings of 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 33–39. Barcelona, Spain.

M. P. Kaschak. 2007. Long-term structural priming affects subsequent patterns of language production. *Memory and Cognition*, 35:925–937.

M. P. Kaschak and K. L. Borregine. 2008. Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, 58(3):862–878.

M. P. Kaschak, R. A. Loney, and K. L. Borregine. 2006. Recent experience affects the strength of structural priming. *Cognition*, 99:B73–B82.

R. Kempson, W. Meyer-Viol, and D. Gabbay. 2000. *Dynamic Syntax: The Flow of Language Understanding*. Blackwell, Oxford, UK.

A. Kilgarriff. 2005. Language is never ever ever random. *Corpus Linguistics and Linguistic Theory*, 1.

A. Konopka and J. K. Bock. 2005. Helping syntax out: What do words do? Abstract, 18th Annual CUNY Conference on Human Sentence Processing.

J. Kuhn. 2004. Experiments in parallel-text based grammar induction. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 470–477. Barcelona, Spain.

R. Kuhn and R. D. Mori. 1990. A cache-based natural language model for speech recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(6):570–583.

W. J. M. Levelt. 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

W. J. M. Levelt and S. Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology*, 14:78–106.

R. L. Lewis and S. Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29:1–45.

V. Lombardo and P. Sturt. 1999. Incrementality and lexicalism: A treebank study. In S. Stevenson and P. Merlo, editors, *Lexical Representations in Sentence Processing*. John Benjamins.

D. M. Magerman and M. P. Marcus. 1990. Parsing a natural language using mutual information statistics. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 984–989. AAAI Press, Cambridge, MA.

M. Marcus, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn treebank: Annotating predicate argument structure. In *Proceedings of ARPA Human Language Technology Workshop*. Plainsboro, NJ.

A. Melinger and C. Dobel. 2005. Lexically-driven syntactic priming. *Cognition*, 98:B11–B20.

A. S. Meyer. 1996. Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, 35:477–496.

M. Niv. 1994. A psycholinguistically motivated parser for CCG. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 125–132.

J. Pearson, M. Pickering, H. Branigan, J. McLean, C. Nass, and J. Hu. 2004. The influence of beliefs about an interlocutor on lexical and syntactic alignment: Evidence from human-computer dialogues. Abstract, 10th Annual Conference on Architectures and Mechanisms of Language Processing (AMLaP).

C. Phillips. 2003. Linear order and constituency. *Linguistic Inquiry*, 34:37–90.

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

M. J. Pickering, H. P. Branigan, and J. F. McLean. 2002. Constituent structure is formulated in one stage. *Journal of Memory and Language*, 46:586–605.

M. J. Pickering and S. Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225.

M. C. Potter and L. Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38:265–282.

M. Purver and M. Otsuka. 2003. Incremental generation by incremental parsing: Tactical generation in dynamic syntax. In *Proceedings of the 9th European Workshop on Natural Language Generation*.

D. Reitter. 2008. The repetition of general lexical material boosts structural priming in language production. Poster at 21st Annual CUNY Conference on Human Sentence Processing.

D. Reitter, J. Hockenmaier, and F. Keller. 2006a. Priming effects in Combinatory Categorial Grammar. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 308–316. Sydney, Australia.

D. Reitter and F. Keller. 2007. Against sequence priming: Evidence from constituents and distituents in corpus data. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society (CogSci)*, pages 1421–1426. Nashville, TN.

D. Reitter and F. Keller. 2008. Structural priming in language production as the result of learning and spreading activation in an act-r model. Talk at 21st Annual CUNY Conference on Human Sentence Processing.

D. Reitter, F. Keller, and J. Hockenmaier. 2007. Corpus-based evidence against sequence priming. Poster at 20th Annual CUNY Conference on Human Sentence Processing.

D. Reitter, F. Keller, and J. D. Moore. 2006b. Computational modeling of structural priming in dialogue. In *Proceedings of Human Language Technology Con-*

ference/North American Chapter of the Association for Computational Linguistics (HLT/NAACL), pages 121–124. New York, NY.

D. Reitter and J. D. Moore. 2007a. Predicting success in dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 808–815. Prague, Czech Republic.

D. Reitter and J. D. Moore. 2007b. Successful dialogue requires syntactic alignment. Poster at 20th Annual CUNY Conference on Human Sentence Processing.

D. Reitter, J. D. Moore, and F. Keller. 2006c. Corpus-based evidence for syntactic priming as rule repetition. Talk at the 12th Annual Conference on Architectures and Mechanisms for Language Processing.

D. Reitter, J. D. Moore, and F. Keller. 2006d. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)*, pages 685–690. Vancouver, Canada.

R. A. Rescorla and A. R. Wagner. 1972. theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy, editors, *Classical Conditioning II*, pages 64–99. Appleton-Century-Crofts, New York, NY.

A. Roelofs. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42:107–142.

A. Roelofs. 1993. Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, 47:59–87.

D. Roland and D. Jurafsky. 2002. Verb sense and verb subcategorization probabilities. In S. Stevenson and P. Merlo, editors, *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, pages 325–346. John Benjamins, Amsterdam, Netherlands.

A. Salamoura and J. N. Williams. 2007. Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics*, 28:627–660.

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

C. Scheepers and M. Corley. 2000. Syntactic priming in German sentence production. In L. R. Gleitman and A. K. Joshi, editors, *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pages 435–440. Erlbaum, Mahwah, NJ.

J. Schenkein. 1980. A taxonomy for repeating action sequences in natural conversation. In *Language production*, volume 1. Academic Press, London, UK.

S. Schoonbaert, R. J. Hartsuiker, and M. J. Pickering. 2007. The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, 56:153–171.

M. Smith and L. Wheeldon. 2001. Syntactic priming in spoken sentence production - an online study. *Cognition*, 78:123–164.

L. M. Stallings, M. C. MacDonald, and P. G. O'Seaghdha. 1998. Phrasal order constraints in sentence production: Phrase length and verb disposition in heavy-np shift. *Journal of Memory and Language*, 39:392–417.

M. Steedman. 1999. Connectionist sentence processing in perspective. *Cognitive Science*, 23:615–634.

M. Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA.

D. Swinney. 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*, 18(6):645–659.

B. Szmrecsanyi. 2005. Creatures of habit: A corpus-linguistic analysis of persistence in spoken english. *Corpus Linguistics and Linguistic Theory*, 1(1):113–149.

B. Szmrecsanyi. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Mouton de Gruyter, Berlin, Germany.

D. Tannen. 1989. *Talking voices: Repetition, dialogue and imagery in conversational discourse*. Cambridge University Press.

W. N. Venables and B. D. Ripley. 2002. *Modern Applied Statistics with S. Fourth Edition.* Springer, New York, NY.

T. Vosse and G. Kempen. 2000. Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75:105–143.

M. Walker, I. Langkilde, J. Wright, A. Gorin, and D. Litman. 2000. Learning to predict problematic situations in a spoken dialogue system: Experiments with How may I help you? In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 210–217. San Francisco, CA.

E. J. Weiner and W. Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics*, 19:29–58.

L. Wheeldon and A. Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.

M. White and J. Baldridge. 2003. Adapting chart realization to CCG. In *Proceedings of 9th European Workshop on Natural Language Generation*. Budapest, Hungary.