ABSTRACT


Title of Dissertation: PARSING, GENERATION, AND GRAMMAR

Shota Momma, Doctor of Philosophy, 2016

Dissertation directed by: Professor Colin Phillips, Linguistics



Humans use their grammatical knowledge in more than one way. On one hand, they use it to understand what others say. On the other hand, they use it to say what they want to convey to others (or to themselves). In either case, they need to assemble the structure of sentences in a systematic fashion, in accordance with the grammar of their language. Despite the fact that the structures that comprehenders and speakers assemble are systematic in an identical fashion (i.e., obey the same grammatical constraints), the two 'modes' of assembling sentence structures might or might not be performed by the same cognitive mechanisms. Currently, the field of psycholinguistics implicitly adopts the position that they are supported by different cognitive mechanisms, as evident from the fact that most psycholinguistic models seek to explain either comprehension or production phenomena. The potential existence of two independent cognitive systems underlying linguistic performance doubles the problem of linking the theory of linguistic knowledge and the theory of linguistic performance, making the integration of linguistics and psycholinguistic harder. This thesis thus aims to unify the structure building system

in comprehension, i.e., parser, and the structure building system in production, i.e., generator, into one, so that the linking theory between knowledge and performance can also be unified into one. I will discuss and unify both existing and new data pertaining to how structures are assembled in understanding and speaking, and attempt to show that the unification between parsing and generation is at least a plausible research enterprise.

In Chapter 1, I will discuss the previous and current views on how parsing and generation are related to each other. I will outline the challenges for the current view that the parser and the generator are the same cognitive mechanism. This single system view is discussed and evaluated in the rest of the chapters.

In Chapter 2, I will present new experimental evidence suggesting that the grain size of the pre-compiled structural units (henceforth simply structural units) is rather small, contrary to some models of sentence production. In particular, I will show that the internal structure of the verb phrase in a ditransitive sentence (e.g., *The chef is donating the book to the monk*) is not specified at the onset of speech, but is specified before the first internal argument (*the book*) needs to be uttered. I will also show that this timing of structural processes with respect to the verb phrase structure is earlier than the lexical processes of verb internal arguments. These two results in concert show that the size of structure building units in sentence production is rather small, contrary to some models of sentence production, yet structural processes still precede lexical processes. I argue that this view of generation resembles the widely accepted model of parsing that utilizes both top-down and bottom-up structure building procedures.

In Chapter 3, I will present new experimental evidence suggesting that the structural representation strongly constrains the subsequent lexical processes. In particular, I will show that conceptually similar lexical items interfere with each other only when they share the same syntactic category in sentence production. The mechanism that I call *syntactic gating*, will be proposed, and this mechanism characterizes how the structural and lexical processes interact in generation. I will present two Event Related Potential (ERP) experiments that show that the lexical retrieval in (predictive) comprehension is also constrained by syntactic categories. I will argue that the syntactic gating mechanism is operative both in parsing and generation, and that the interaction between structural and lexical processes in both parsing and generation can be characterized in the same fashion.

In Chapter 4, I will present a series of experiments examining the timing at which verbs' lexical representations are planned in sentence production. It will be shown that verbs are planned before the articulation of their internal arguments, regardless of the target language (Japanese or English) and regardless of the sentence type (active object-initial sentence in Japanese, passive sentences in English, and unaccusative sentences in English). I will discuss how this result sheds light on the notion of incrementality in generation.

In Chapter 5, I will synthesize the experimental findings presented in this thesis and in previous research to address the challenges to the single system view I outlined in Chapter 1. I will then conclude by presenting a preliminary single system model that can potentially capture both the key sentence comprehension and sentence production data without assuming distinct mechanisms for each.

# PARSING, GENERATION, AND GRAMMAR

by

Shota Momma

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

2016

Advisory Committee:

Professor Colin Phillips, Chair

Professor Bill Idsardi

Professor Ellen Lau

Professor L. Robert Slevc

Professor Rochelle Newman

Acknowledgements

I am finally in the stage of writing this part of the dissertation. If I described in detail all the great support that I received from my official advisor, semi-official advisor, and nonofficial advisor, collaborators, colleagues, friends and family, this section would be longer than the rest of this thesis. So I'll try to be relatively brief. However, the reader should note that the length of this section does not correlate with how thankful I am to all the people I met while my time at Maryland.

First and foremost, I would like to thank my advisor and committee chair, Colin Phillips. He has patiently listened to my half-baked ideas every week and has given me countless pieces of useful feedback. I sometimes did not immediately understand some of his feedback, but without exception I had some moments later when I realized how to-the-point his feedback was. I have encountered such moments numerous times throughout my graduate career, and that was a slightly scary experience. He also responded immediately to my numerous last minute requests for help, no matter what time it was. That was also a slightly scary experience. Given that humans must sleep, I almost came to a conclusion that there are more than one Colins in the department. I even asked Andrea Zukowski, his wife and my colleague, if he actually has a twin brother (he doesn't). What those slightly scary stories indicate is that he is such a dedicated and effective advisor. I want to be like him when I advise my future students. On the side note, he, as a British man, advised me on how to dress properly in a conference and in a Skype interview.

I would like to thank my semi-official advisor and committee member, Bob Slevc. He is one of the most humorous and approachable people I have met, and he is the one who made it possible for me to do research on sentence production. More than half of the experiments in this thesis would have been impossible without his advising and his generous support. I first thought that designing production experiments was so hard and that it is practically impossible to study the subtle linguistic processes that I was interested in. He has convinced me otherwise - with a lot of humor.

I would like to thank my unofficial advisor and committee member, Ellen Lau. She has given many constructive criticisms that greatly improved the arguments and experiments presented in this thesis. She was available whenever I needed some help. I admire her for being an extremely productive scholar and effective advisor while having her own family at the same time. She is the one who demonstrated most clearly that having a satisfying academic and non-academic life at the same time is possible.

I would like to thank my remaining committee members, Bill Idsardi and Rochelle Newman. They both kindly agreed to serve as committee members, and have provided constructive feedback on this thesis from their own perspective, which is immensely valuable because most of the experiments and arguments in this thesis have mostly been presented only to a community specialized in sentence processing.

I would like to thank my generous collaborator, Hiromu Sakai. He generously let me use his equipment and lab space in Japan, and also helped me improve my experimental designs. Without him, all the Japanese experiments that I presented in this thesis would not have been conducted.

I would like to thank my research assistants, Shweta Roy, John Mathena and Becca Kraut. I am so lucky to work with such talented young students. Shweta was a high school student back when she worked with me, and I cannot even imagine myself as a high school student doing what she accomplished. John and Becca were my research assistants, but at the same time my research collaborators. I am very much looking forward to the outcome of the studies that I will continue to work on with them.

I would like to thank the lab managers and research assistants of the CNL lab, Julia Buffinton, Jon Burnsky, and Neomi Rao. They made my life (and all others' lives) a whole lot easier. I have not seen any person who is as organized as Julia. She is so organized that her leaving the lab created chaos in our lab. Jon and Neomi generously edited this thesis for typos and English mistakes. Without their help, this thesis would be full of agreement errors, extra determiners, and missing determiners, to the extent that the errors impair the readers' understanding.

I would like to thank my undergraduate advisor, Lee Osterhout. Lee was the first psycholinguist I met in the U.S., and he is the primary reason that I chose to do psycholinguistics. He generously gave me an opportunity to work in his lab, even though I clearly did not know what psycholinguistics was and spoke pretty bad English. Without this opportunity, the word *psycholinguistics* probably would not be listed in my mental lexicon.

I would like to thank my undergraduate mentor, Darren Tanner. He is the one who taught me how to do electroencephalography studies on language, and he is also the first linguist I met in a psychology department. He demonstrated to me that the disciplinary boundary is an illusion. Without him, I would not have chosen to go to graduate school in linguistics after getting an undergraduate degree in psychology.

I would like to thank all of my colleagues and friends in the Maryland language science community, and all the members in Hiromu Sakai's lab at Hiroshima University and Waseda University in Japan. I certainly want to individually thank all of them, but I do not want to take the risk of missing some of the people's names. The number of amazing people that I met at Maryland far exceeds my working memory capacity. As I am obviously writing this section at the very last minute, I fear that I will miss some of the names. So, I dare not do that.

Finally I would like to acknowledge that this research was generously funded by an NSF dissertation improvement grant (NSF-DDRI: #BCS-1530332).

This thesis is an interim conclusion of my 9-year journey in the United States, to be continued. I was a naive high school graduate who spoke very little English when I first arrived in the U.S. I often think about how it is possible for such a naive high school graduate with poor English skills to get to this stage, and the only possible conclusion I can draw is that I have been receiving unusually good support from all the people I met throughout my academic career in the United States. So, thank you all.

Table of Contents

List of Tables

List of Figures

# 1 Introduction

Humans use their linguistic knowledge in more than one way. On one hand, they use their linguistic knowledge to comprehend; on the other hand, they use their linguistic knowledge to speak. The goal of this thesis is to explore the relationship between these two modes of using linguistic knowledge, specifically focusing on the structural aspects of language. I will make the following argument: the parser and the generator are an identical cognitive mechanism executing the same processes in understanding and speaking. At the end of this thesis, I will sketch out a general architecture for a single unified model that captures structure building in both sentence comprehension, i.e., parsing, and in sentence production, i.e., generation.

It is worth clarifying at the outset that the goal of the current thesis is not to argue that comprehension and production as a whole are identical. The claim that comprehension and production are identical is trivially false. After all, I do not use our ears to speak nor do I use our mouth to comprehend. I also will not argue for the slightly more restrictive claim that the cognitive processes that are involved in comprehension and production are the same. Such a claim requires that cognitive processes like motor planning are centrally involved in comprehension or that auditory processing (or visual processing, in the case of sign language) is centrally involved in production. Although sometimes such a claim has been advanced under the theme of embodied cognition (e.g., Pickering & Garrod, 2007; 2013), I remain agnostic about it and the current arguments have nothing to do with such a view. Throughout this dissertation I specifically argue that *structure building* in comprehension and production are performed by an identical, single mechanism.

Of course, sentence comprehension and sentence production involve more than just structure building. I focus on structure building because it is the central component necessary for systematic interpretation and systematic utterance generation, and it is the component that instantiates human grammatical knowledge. As discussed below, the main benefit of the current thesis, if successful, is that it paves the way to develop a transparent linking hypothesis between grammar and processing. Thus, I will focus on building an unified model of structure building in sentence comprehension and sentence production (along the line of Kempen, 2000; Kempen et al., 2012; Phillips & Lewis , 2013; Lewis & Phillips, 2015).

The current attempt to build a unified model of parsing and generation is not the same as some of the previous attempts to relate comprehension and production. For instance, Pickering & Garrod, (2007; 2012) argued that the comprehension and production systems are heavily utilized in both in the single task of understanding and the single task of speaking, proposing a type of embodied model of language use. MacDonald (2012) separately proposed that comprehension and production are related to each other in the sense that the processing biases of the production system shape the distributional properties of the corpus, which in turn shapes how comprehenders process language. These two views constitute important claims about how comprehension and production relate to each other. However, the current attempt is entirely about the structure building mechanism, which is a sub-system of the comprehension and production systems. The claim that a sub-system is shared between two systems is different from the claim that  the comprehension system and production system, as a whole, interact closely. Also, the current attempt to unify the models of parsing and generation is strictly about how parsing and generation relate in individuals' mind. Thus, McDonald's view that comprehension and production interact via the external world is orthogonal to the current attempt. See section 1.1 below for a more detailed discussion on how the current attempt is different from the previous proposals on comprehension and production relations.

Why should I care to unify the models of parsing and generation? A major consequence of the current proposal for a unified model of structure building is that one-to-one alignment between grammatical and processing theories becomes possible. If there are two distinct theories of structure building, one for comprehension and the other for production, there need to be two distinct linking hypotheses between grammar and processing. Under this *dual-system view* (Figure 1-1a), two distinct linking theories are required to connect the processing and grammatical theories. This is a rather theoretically unattractive position, and can be used as an excuse to not think about the relationship between grammar and processing. In questioning this view, I attempt to reduce the two theoretical objects, the parser and the generator, into one - the structure builder - so that the linking theory can also be reduced to one. This *single system view* (Figure 1-1b) is the prerequisite for maintaining the stronger hypothesis that grammar is *implementation-dependent* (Figure 1-1c; Phillips & Lewis, 2013). The implementation-dependence hypothesis states that there is only one way to realize the grammar algorithmically. That is, there is only one way to compute a specific grammatical object. This poses a rather strong constraint on the linking theory between grammar and processing, because logically speaking, the same grammatical object can be reached by various algorithms, just as arithmetical multiplication can be computed in more than one way. This strong position is theoretically appealing, because it not only reduces two theoretical objects into one but also reduces two linking hypotheses between grammatical theories and processing theories into one. In its strongest form, it allows the grammar and the structure builder to be identified, in the form of an integrated grammar-structure builder (Phillips, 1996; Figure 1-1c). Under this strong view, there is no need for a linking theory, because the theory of structure building is the integrated part of grammatical theories. However, for the implementation dependance hypothesis to be true, there must be a single mechanism that is performing a single process that subserves parsing and generation. Otherwise, grammatical theories remain to be an abstraction over two different ways of building structure. This thesis thus aims to reduce the model of parsing and the model of generation into one in order to meet the prerequisite for considering the implementation dependence view.

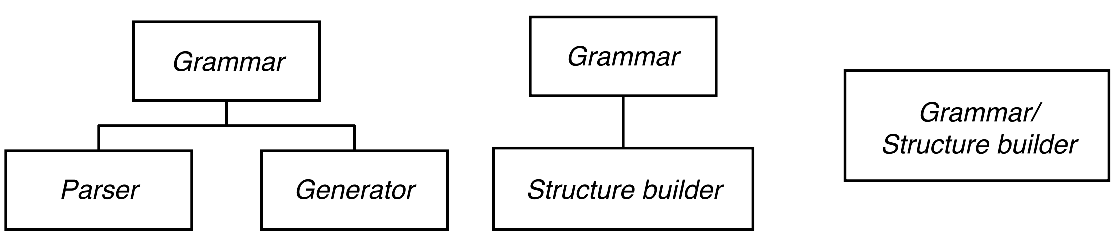



Figure 1-1abc: Three views on the relationship between parsing, generation, and grammar. Left: dual system view; Middle: one system view; Right: Implementation dependence view. Each box represents a separate cognitive theory, and the line represents the linking theories between different cognitive theories.

### *1.1 Relationship between parsing and generation*

Since the early days of the neuropsychology of language, two separate neurological substrates have been thought to subserve language comprehension and language production, perhaps most famously in the Wernicke-Geschwind model (Geschwind, 1970). This separation stems mainly from the evidence for a neuropsychological double dissociation: Broca's aphasia has been thought to reflect a selective deficit of sentence production, while Wernicke's aphasia has been thought to reflect a selective deficit of sentence comprehension. This apparent neuropsychological double-dissociation is a strong sign of two separate cognitive mechanisms.

However, this double dissociation has been questioned as more psycho-linguistically rigorous investigations were carried out. Perhaps most prominently, Caramazza & Zurif (1976) demonstrated that the comprehension of Broca's aphasics (and conduction aphasics) is far from normal. Specifically, they showed that Broca's aphasics's can arrive at the correct interpretation when world-knowledge is sufficient to constrain the interpretation (when arguments are not reversible without violating world knowledge; e.g., *the apple that the boy is eating is brown*), but their performance dropped to chance level when syntactic information is required to arrive at the correct interpretation (when arguments are reversible without violating world knowledge; e.g., *the lion that the tiger is chasing is fat*). This pattern of deficit shows that Broca's patients rely on some heuristic strategy to arrive at an interpretation using their world knowledge. In other words, they interpreted Broca's aphasia as a deficit of algorithmic computation involving (morpho-)syntax, a modality neutral computational mechanism. Broca's aphasia does not selectively impair production, and the classical double-dissociation between comprehension and production falls apart. Of course, Broca's and Wernicke's aphasias still affect comprehension and production differently, and such differences need an explanation. However, the differences between Broca's and Wernicke's aphasias can be attributed to more modality-independent deficits of syntactic and semantic knowledge, respectively.

More recently, researchers started to seriously explore the relationship between comprehension and production, under the general statement that sentence comprehension and sentence production are more 'tightly interwoven' than previously thought (Pickering & Garrod, 2012). This view on the tight interweaving between comprehension and production is shared between multiple researchers (Dell & Chang, 2014; Kempen, 2012; MacDonald, 2012; Federmeier, 2007; Garrett, 2000). However, there are at least five distinct classes of claims under this general statement. Despite their surface similarity, these claims are distinct in their essence. Here I summarize those claims, in order to compare and contrast these views, along with the view that I advocate for in this dissertation. Those five views can be captured roughly by the following three categories: the interactionist, externalist, and shared mechanism views. The interactionist view claims that the comprehension and production systems are distinct, but that they are deployed together interactively in an individual act of comprehending or producing an utterance. The externalist view focuses on the relationship between the comprehension and production systems as they are mediated by the external world, specifically by the corpus of utterances that a speaker experiences. Finally, the shared mechanism view claims that some or all the components of the comprehension and the production systems are identical and thus shared. These three categories of views are not necessarily mutually exclusive, but they form theoretically distinct claims about the relationship between comprehension and production.

### 1.1.1 Interactionist view

This view claims that comprehension and production mechanisms interact closely during individual acts of speaking or understanding. There are (at least) three variants of this view, one proposed by Pickering and Garrod (2007; 2012), another proposed by Federmeier (2007), and another by Garrett (2000) and partly by Levelt (1989). Also, the model described in Dell & Chang (2014) can also be classified as an interactionist model.

**Embodiment**: Pickering and Garrod (2007; 2013) argued that comprehension and production are ‘tightly interwoven,’ suggesting that the production mechanism is recruited during comprehension and the comprehension mechanism is recruited in production. Under their framework, comprehension is a form of action perception and production is a form of action[1]. Embodied cognition asserts that action perception occurs via action. Thus, under the theme of embodied cognition, comprehension proceeds via production. Pickering and Garrod emphasize that this framework can explain the predictive nature of language processing, a property of language processing that has attracted great deal of attention in past 15 years or so (e.g., Altman & Kamide, 1999; DeLong, Ubach & Kutas, 2005). They explained the predictive nature of language processing in terms of *covert imitation*, which critically is a function of the production system, and argued that their embodied view of language naturally explains the predictive nature of language processing. They also invoked the notion of *forward model* from the motor planning literature to justify their reverse claim that production proceeds via comprehension. In particular, in the motor planning literature, the perceptual consequences of one's own motor plan are predicted (creating what is called a forward model) and are compared to the actual perceptual consequences of the movement (e.g., Wolpert, 1997). Importing this notion to language, Pickering & Garrod argue that speakers generate an internal representation of the utterance percept using the production mechanism, and predict the perceptual consequences of it using the comprehension mechanism. In other words, comprehension operates via a kind of analysis-by-synthesis mechanism in which the production system has a central role.

---

[1] This need not be the case, as comprehension and production can occur in absence of immediate speech act. For instance, comprehension can occur without simultaneous speech (as in writing), and production can occur without immediate articulation (as in internal speech). However, under the view of Pickering & Garrod (2013) and the embodied cognition theory more generally, such forms of comprehension and production are still claimed to be action perception and action respectively, i.e., they critically involve action mechanisms like motor control.

Importantly for the current purpose, however, Pickering and Garrod's claim does not imply that parsing and generation are performed by a single mechanism. The comprehension mechanism and production mechanism are maintained in a distinct fashion, though they interact heavily (as in, e.g., MacKay, 1982). For instance, in their model (as illustrated in p. 12, p. 16), a comprehenders's perception of an interlocutors' utterance (which includes representations of phonology, syntax and semantics) is generated by the comprehension implementer. This comprehension implementer exists independently from any of the components of their model, including the *inverse model generator*, *forward model generator,* and *production implementer*. That is, there exists a syntactic parser that builds (at least rudimentary) structures of utterances independently of the production mechanisms. Likewise, there exists a syntactic generator that builds structure independently of the comprehension mechanisms. In this sense, the parser and the generator are at least partially separate components in their model. It is noteworthy that their model, just like other analysis-by-synthesis models (discussed below, see also Townsend & Bever, 2001), must be able to build initial percepts without relying on production mechanisms. In this sense, Pickering and Garrod's (2007; 2013) model can be characterized as interactionist, and it does not assume that structure building for comprehension and production is identical.

**Production Affects Reception in the Left Only (PARLO)**: Another version of the interactionist view has been proposed by Federmeier (2007). In her PARLO framework, she argues that top-down, predictive processes in comprehension are more robust and fine-grained in the left-hemisphere of the human cortex, because predictive processes are indeed the function of the production mechanism, which arguably resides in the left hemisphere. This view is interactionist in the sense that the anatomical-functional loci of comprehension and production are assumed to be distinct, and prediction is the product of interaction between the comprehension and production systems.

**Analysis-by-synthesis**: Finally, there are a number of claims that fall under the view that comprehension and production involve an 'analysis-by-synthesis' mechanism. In essence, the analysis-by-synthesis view proposes that the analysis of perceptual inputs in comprehension is done via a matching process between the input and an internally generated representation. Pickering & Garrod's (2013) model discussed above is also a type of analysis-by-synthesis view, though I treated their model separately because some of their claims do not necessarily follow from the analysis-by-synthesis view, and also because their model discusses the relationship between comprehension and production more explicitly than other analysis-by-synthesis views.

Different analysis-by-synthesis accounts have different means to generate the internal representations that are compared against the input. For instance, Townsend & Bever (2001) assume a heuristic parser, which adopts strategies like assuming a canonical transitive structure (i.e., Noun-Verb-Noun), creating a 'rough-and-ready' representation of the input. This initial representation is used to create more detailed, grammatically accurate internal representations. Under this type of analysis-by-synthesis account, both the parser and the generator exist at least partially independently, as the initial analysis of the input is performed by the heuristic parser that is different from the generator. Similarly, Garrett (2000) argued that analyses proposed by the parsing mechanism are evaluated and filtered by the production mechanism. In this view, the parser proposes an analysis or analyses of the input stings, and the proposed analyses of the input is compared against the structure proposed by the generator.

Although the assumed quality of such representations might differ from Townsend & Bever (2001), the model proposed by Pickering & Garrod (2013) also generates an initial representation of the (partial) input independently from the generator. The involvement of the generator is limited to the prediction of the upcoming input, and the analysis of the current input must be possible without invoking the generator. In this sense, the analysis-by-synthesis claims advanced by Townsend & Bever (2001), Garrett (2000), and Pickering & Garrod (2013) do not entail that the parser and the generator are the same mechanism. Indeed, to the extent that the initial analysis of the input must be done via the parser, the parser and the generator exist independently. If the parser and the generator exist independently, they can follow different principles in how they build structures. Of course, the analysis-by-synthesis view does not need to be in conflict with the view that the parser and the generator are identical *per se*. Phillips (2013) for example does not assume generator-independent creation of an initial perceptual analysis, unlike other analysis-by-synthesis models discussed above. As long as the syntactic representation does not have to be created based on a generator-independent parser, the analysis-by-synthesis view can be compatible with the view that the parser and the generator are identical.

### 1.1.2 Externalist view

This view, proposed by MacDonald and colleagues (MacDonald, 2012; MacDonald & Thornton, 2009; see also Dell & Chang, 2011), states that processing biases in production influence the processing biases of the comprehension system. This is dubbed as the Production-Distribution-Comprehension (PDC) view. Undoubtedly, production biases shape the distributional properties of the corpus that individuals are exposed to. This production bias effect on the corpus distribution affects how comprehenders learn to process the target language efficiently. For example, this view explains the relative comprehension difficulty of animate-first object-relative clauses (e.g., *the director that the movie pleased*) as compared to inanimate-first object-relative clauses (e.g., *the movie that the director watched…*), as reported in Gennari & MacDonald (2009), in the following steps. First, the production mechanism has a bias, among other biases, for producing animate nouns first (e.g., McDonald, Bock & Kelly, 1993). This shapes the distribution of the corpus such that animate nouns are more likely to be used for passive-relative clauses while inanimate nouns are more likely to be used for active relative clauses. Comprehenders learn to expect an object relative clause when they encounter an inanimate noun and a sign of a relative clause. This kind of effect is assumed to happen both on developmental and historical time scales, affecting both moment-by-moment processing mechanisms and language typology. This view is mainly concerned with the developmental and diachronic relationship between comprehension and production, and therefore it makes only an indirect commitment to how the parser and the producer as cognitive systems are related within an individual's mind at a single point in time. For this reason, I call this view an externalist view, in the sense that the interaction between the production and comprehension systems is mediated by the external world, i.e., the corpus.

### 1.1.3 Shared representation view

Another claim is that the syntactic representations constructed during comprehension and production are shared. In other words, the final representational products of parsing and generation processes are the same. This is a relatively uncontroversial view, in the sense that comprehended and produced utterances obey the same grammatical constraints. The fact that people can communicate via language is sufficient evidence that the grammar used in comprehension and production is at least vastly overlapping. This view, however, is incompatible with the logically possible (though not seriously proposed) dual-grammar view (Figure 1-2) where neither the representations nor the processors for syntactic structure are shared. However, rejecting the dual-grammar view does not necessarily entail the single-mechanism view. This is because a single grammar can be used by two separate mechanisms as described in Figure 1-1a. Even when there is only one grammar, the systems that deploy it might be different between parsing and generation.

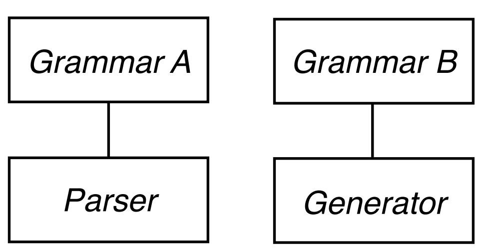



Figure 1-2: the dual grammar view.

### 1.1.4 Shared representation and processor view

A stronger view than the shared representation view is that not only the final product, but also the mechanism that computes the final representation is shared. That is, syntactic representations are computed by the very same mechanism in the very same way in comprehension and production. Though the details are different, this view is shared by Kempen (2000); Kempen et al., (2012), and possibly by Dell & Chang (2012).

### 1.1.5 Shared representation, processor, and order view

Finally, the shared representation and processor view can further be augmented to include the claim that the order in which structures are built is the same between parsing and generation. In its strong form, this view claims that the same structural representation is built in the same order by the same processor. This view is reflected by Phillips & Lewis (2013). This view is not the same as the shared representation and processor view,. This is because the shared representation and processor view does not demand that structures are built in the same order between parsing and generation. Specifically, under the shared representation and processor view, the order in which the parser and the generator build structures can be different. For example, the parser builds structure mostly from left to right, because the sound input unfolds sequentially from left to right. On the other hand, it is in principle possible (though probably not true) that the generator builds structure from right to left, because the message input might unfold differently. If this is the case, even if the processor used in parsing and generation is the same, the order in which the structures of the different sub-parts of a sentence are built might be distinct.

Thus, the shared representation, processor and order view has an extra condition to be satisfied compared to the shared representation and processor view. The shared representation, processor and order view not only requires that the underlying processor is the same, but also that the order of steps of structure building is the same.

### 1.1.6 Current view

Having summarized five different claims about the comprehension-production relations, now I discuss the view that I propose in this thesis in some more detail in comparison to those five views. This exercise will make it clear that the closest view is the fifth view that the processing mechanism for structure building in comprehension (parser) and in production (generator) is subserved by a single processor executing the same algorithms. However, I will make comparisons one by one to make the current proposal clearer.

First, the current proposal is not the same as the interactionist views. This is mainly because the interactions between the parser and the generator, no matter how close they are, do not entail that any components of those systems are identical. Regarding the embodiment view, I specifically focus on the structure building component of the comprehension and production mechanisms, which is only a sub-component of action and perception. The claim that comprehension and production are both deployed in an individual act of understanding and speaking does not entail that the the parser and the generator are identical. For instance, it is possible that there are two structure builders operating on different principles in comprehension and production, and both of these structure builders operate in an individual act of understanding and speaking, in a way that is often assumed by the analysis-by-synthesis view. The PARLO framework also is not obviously relevant to the current proposal. Indeed, the PARLO framework and the current proposal can be construed as not easily compatible. The PARLO framework argues that the structure building system for production resides only in the left hemisphere of the brain while the structure building system for comprehension resides in both. No parts of this model require that the structure building components of comprehension and production are identical. Indeed, such a neuroanatomical claim can be construed as assuming distinct systems for parsing and generation with distinct neuroanatomical substrates.

Also, the current view does not follow from either Townsend & Bever's (2001) or Garrett's (2000) views. Those analysis-by-synthesis views assume that the parser, independently from the generator, creates a preliminary analysis of the input stimuli. Such a view presupposes the independent existence of the parser and the generator. Thus, the views that advocate for heavy interaction between comprehension and production do not equate with the claim that the parser and the generator are identical.

Second, the current proposal is not directly related to the externalist view as described by MacDonald (2012). This is because MacDonald's view connects comprehension behavior and production behavior via the external world, namely via the statistical distribution of language expressions in a corpus. The current proposal is strictly about the relationship between the parser and the generator within an individual's mind, and hence it is not the same as a claim about how the external manifestation of the production mechanism influences comprehension. MacDonald's view and the current view are not incompatible with each other, but the focus of the two views is simply different.

Third, the current proposal goes beyond the view that syntactic representations are shared, i.e., modality neutral. The existence of modality neutral representations is not very controversial, as speakers and comprehenders of the same language use a vastly overlapping grammar and vastly overlapping lexicon. Successful communication is otherwise not possible. Experimentally, structural priming effects can be obtained across comprehension and production (Bock et al., 2007). This cross-modal priming suggests that something must be shared between comprehension and production. Emerging neurophysiological data (Segart et al., 2012) also suggests that structural and lexical resources are subserved by the same neural populations (based on the functional Magnetic Resonance Imaging (fMRI) repetition priming paradigm; Grill-Spector & Malach, 2001). Other studies also found that the neural systems supporting comprehension and production are largely shared in more naturalistic speech (Silbert et al., 2014). Given that the default is to believe that the representations that speakers and comprehenders build is the same, these provide sufficient evidence to believe that *at least* the structural representations constructed in comprehension and production are shared between comprehension and production, and possibly more. Thus, in this dissertation this view is assumed as given: at some point in parsing and generation, the same syntactic knowledge is consulted and the same cognitive object is represented.[2] This view, however, is still compatible with the dual-system view as in Figure 1-2. For instance, syntactic priming might be due to the priming of knowledge representations, rather than priming of processes that deploy such knowledge representation. If that is the case, the processes to reach a specific syntactic representation might still differ between comprehension and production. Adding fMRI data does not resolve this problem, partly because it is practically impossible to prove the non-existence of unshared neural resources, and partly because fMRI has a poor temporal resolution that prevents researchers from seeing the intermediate stages of structure building. In other words, the extant evidence for cross-modal syntactic priming is not decisive evidence with respect to whether a single cognitive system subserves parsing and generation, and the evidence for shared neural populations is evidence for some shared knowledge or processes, but it is not clear which.

Finally, the central claims of the current proposal are that (i) the processor for parsing and generation is identical and that (ii) the order in which that processor builds structure during parsing and generation is identical. (i) essentially captures the shared representation and processor view, and the combination of (i) and (ii) reflects the shared representation, processor and order view above. In this thesis, I will argue for (i), and evaluate if (ii) is tenable.

[2] This claim is actually not entirely agreed upon, especially in comprehension. For instance, in the *good-enough* approach to sentence comprehension (e.g., Ferreira & Putnam, 2007), comprehenders do not draw upon detailed grammatical knowledge and parse sentences heuristically. For more discussion, see Stroud & Phillips (2012). An analogous view is not usually assumed in sentence production.

There is some interesting previous evidence for the shared representation and processor view. Namely, Kempen et al. (2012) showed that speakers override comprehended sentence structures with produced sentence structures when paraphrasing. They compared the response onset latency of anaphoric pronouns in a paraphrasing task in which participants were asked paraphrased the sentence word-by-word from direct speech as in (1-1a) to indirect speech, as in (1-1b). The paraphrased version of (1-1a) is (1-1c).

(1-1a) *The angry headmaster complained: "I have seen a nasty cartoon of myself in the hall."*

(1-1b) **The angry headmaster complained: "I have seen a nasty cartoon of himself in the hall."*

*(1-1c) The angry headmaster complained that he has seen a nasty cartoon of himself in the hall.*

Kempen and colleagues found that the response time to the anaphoric pronoun in the ungrammatical sentence (1-1b) was just as fast as, or if anything even faster than, the grammatical counterpart (1-1a). Importantly, a slow-down effect was obtained in a proofreading task, where participants simply corrected the ungrammaticality of the sentences, and also in a canonical Self-Paced Reading (SPR) task using the same materials. This pattern can be interpreted as evidence that the parsed structure, specifically the gender feature of the embedded subject noun, is overridden by the generated structure due to the structure generated during the paraphrasing task. Kempen et al., (2012) took this to be the evidence that the parser and the generator really construct a single, token-identical, structural object.

The evidence from Kempen et al. (2012) is important and it does provide some support for the shared representation and processor view, but is not sufficient to support the shared representation, processor, and order view. Hence, in this thesis, I have two tasks. The first is to elaborate and provide more arguments for the shared representation and processor view, and elaborate on what it takes to show the shared representation, processor and order view. Before going into the details of the argument, however, I will try to clarify what it means for the structure building mechanisms in comprehension and production to be identical.

#### 1.1.7 What does it mean to be identical?

To argue for the single-system view is to argue for an identity relationship between the parser and the generator. But what does it means for them to be identical? I argue that there are at least three different levels on which the identity claim can be stated. First, the parser and the generator can be ***representationally identical***. They are representationally identical if the representational products of the parser and the generator are the same. This identity claim is rather uncontroversial, because, after all, comprehenders and speakers obey the same grammatical constraints, i.e., represent structures in the same way. Second, the parser and the generator can be ***processor identical***. They are processor identical if they share the same units of structure building (e.g., a precompiled memory representation of a structural configuration of a certain size), the same set of operations (e.g., combinatorial operations and retrieval operations), and the same cognitive resources to perform these operations (e.g.,general purpose working memory). Unlike the representational identity claim, the processor identity claim is not obviously true. Apparent dissociations between parsing and generation, both developmental and neuropsychological, suggest that the parser and the generator might not be not processor identical. For instance, comprehension is usually assumed to precede production, as noted by Keenan & MacWhinney (1987). The reverse also seems to exist for certain domains of language. 4-6-year old children have been reported to misinterpret sentences like *Ernie washed him* to mean *Ernie washed himself* (Chien & Wexler, 1990), but they hardly ever misproduce incorrect pronoun types (Bloom, Barss, Nicol, and Conway, 1994). In neuropsychology, Broca's aphasia disproportionately affects production, and Wernicke's aphasia disproportionately affects comprehension, although neither type of aphasia is strictly selective to production or comprehension. The challenge that such developmental and neuropsychological double dissociation poses is whether they can be attributed to either the difference in input in the task of speaking and comprehending, or a factor other than structure building, like message formulation in production or auditory analysis in comprehension. Third, the parser and the generator could be ***order identical***. They are order identical if the order in which a certain structure is built is identical. The term 'order' here is not just about the linear order of words in a sentence, but about the order in which each node of the hierarchical structure is elaborated. Indeed, if one allows structures to be built from the roots rather than from the leaves (i.e., if one allows structures to be built top-down), the order identity view can be further classified into two versions: **vertical order identity** (the ordering between the creation of parent node and child nodes), and **horizontal order identity** (the order in which the each terminal node is bound to a specific lexical representation). Neither version of the order identity claim is obviously true, because sound and message - the two different starting points for parsing and generation respectively - unfold differently over time. If the starting points for the parser and the generator have a different temporal unfolding, then there can easily be differences in which part of a sentence is elaborated at what point in time in parsing and generation.

In the rest of this thesis, I will argue that the parser and the generator are processor and order identical. This is what I mean by the single-system view as stated in the shared representation, processor, and order view above. Throughout this thesis, I will make both conceptual and empirical arguments for processor and order identity, but both forms of arguments are essentially based on an argument from simplicity. Namely, I will aim to eliminate both existing and possible discrepancies between parsing and generation in terms of mechanisms and order of structure building, without sacrificing the empirical success of the resulting single model.

Before diving into the details of the argument, it must be emphasized again that comprehension and production have rather different starting points. In comprehension, the starting point is sounds (or written characters, or visual signs in the case of signed languages). In production, the starting point is the message. Sounds and signs clearly have an inherent linear structure, as they unfold sequentially over time. In comparison, it is far from clear that messages have an inherent linear structure. How does this difference in the starting points square with the claim that the parser and the generator are mechanistically and algorithmically identical?

In order to answer such a question, the basic operations that the parser and the generator perform must first be defined. In any parsing and generation models, (at least) two kinds of operations are usually involved: retrieval and combination. The retrieval operations retrieve pre-compiled structural units from long-term memory. The grain size of such structural units varies across different models of parsing and generation: they can be as large as an entire clause (e.g., like an elementary tree in Tree Adjoining Grammar (TAG)), or as small as the syntactic category of a single word. The combinatorial operations combine structural units into a single coherent structure. This operation is needed to capture linguistic creativity: the fact that humans can comprehend and produce indefinitely many sentences (Chomsky, 1966; Humboldt, 1836). It is implausible to assume that humans understand and produce sentences merely by retrieving structural units in memory, and thus some means to combine the retrieved units to create a new linguistic object is necessary. Models of parsing and generation could differ in the relative contributions from retrieval and combination in building sentence structures. Some models attribute most of the work to combinatorial operations (e.g., Frazier & Fodor, 1978), meaning that the size of the pre-stored structural unit is minimal. Other models attribute most of the work to retrieval operations (e.g., MacDonald, Pearlmutter &

Seidenberg, 1994), assuming larger structural units[3]. Nevertheless, these two basic kinds of operations - retrieval operations and combinatorial operations - must be present to some extent in any adequate model of human parsing and generation.

Having characterized these two basic kinds of operation in parsing and generation, now I can properly discuss whether the parser and the generator can plausibly be identical at each level, despite clearly different starting points. This discussion will expose several extant and potential misalignments between parsing and generation models.

## 1.2 *Processor identity*

The processor identity view states that the processor for parsing and generation, defined as the collection of basic operations and resources needed to perform those operations, is identical. However, there are several possible challenges for this view, as outlined below. I specifically present four challenges for the view that the parser and the generator are processor identical, which I will attempt to resolve in this thesis.

### 1.2.1 Grain size of structural units

The first potential challenge is that the parser and the generator might use distinct sets of pre-compiled structures in memory. For instance, it is conceivable that the parser

[3] They do not explicitly claim that the size of structural units is large. However, they do assume that the resolution of the structural choice between active and passive voice (in a reduced relative clause) is done via a retrieval process of a structural memory. This entails that the clausal structure of a sentence is a unit in memory. This is the sense in which I take their model to adopt the view that basic structural units can be larger than in many other models of parsing.

might routinely use only minimal pre-compiled structural memories that contain only the category of a single word while the generator might routinely use extensive precompiled structural memories that correspond to a whole clause. In such an extreme case, parsing would involve frequent interleaving between structure building and post-structural processes such as interpretation, while generation would involve much less frequent interleaving between structure building and post-structural processes such as articulation. If this (or the reverse) is the case, then the parser and the generator would use two distinct sets of structural items in memory containing differently sized items. This difference would undermine the processor identity claim, as it amounts to the existence of parsing- or generation-specific long-term memory storage. Existing models indeed disagree with each other in terms of the grain size of structural processing and hence the size of structural unit in memory, especially between models of parsing and models generation. Thus, a challenge for the processor identity view is to align the grain size of the structural processing unit in parsing and generation models without losing the empirical coverage of the resulting single system model.

### 1.2.2 Input differences

As mentioned above, it is an undeniable fact that comprehension and production are based on different kinds of input: sounds and messages. This is the primary motivation for thinking that parsing and generation are different. Under the above-described view that parsing and generation can be decomposed into retrieval of structural units and combinatorial operations between structural units, the real question is whether the combinatorial operations are triggered by two different kinds of input directly. Retrieval of structural units is simply a memory access operation, and I know that retrieval operations in general can be triggered by diverse cue types. For example, people can retrieve a memory of real world event based on olfactory, visual, auditory, tactile or virtually any kind of cue that is associated with a specific memory. Thus, even if the retrieval operations are triggered by different kinds of cues in parsing and generation, that does not constitute evidence for the existence of two separate systems. For instance, one wouldn't argue that there are two lexical systems for comprehension and production simply based on the fact that lexical retrieval can be initiated by conceptual features and by sound features (cf. TRACE model; McClelland & Elman, 1985). Thus, the real question is whether the use of distinct kinds of input in comprehension and production directly affects the combinatorial operations of structure building, in a way that is not attributable to the retrieval operations. If the answer to this question is affirmative, it poses a problem for the single system view.

### 1.2.3 Seriality and parallelism

Regardless of the grain size of the structural units and how they are retrieved, the parser and the generator could differ in terms of whether these operations are performed serially or in parallel. In terms of retrieval operations, the parser could search for syntactic representations in memory in parallel, simultaneously evaluating all that match the features of a given retrieval cue. In contrast, it could also search memory serially, considering only one item at a time until it becomes incompatible with the input. In terms of combinatorial operations, the parser could keep track of all the structural possibilities that are compatible with the input in parallel (e.g., Hickok, 1993). In contrast, it could combine two sub-structures so that only one resulting structure is ever represented (e.g., Frazier, 1989). Likewise, the generator might retrieve/combine pairs of retrieved structures serially or in parallel as in parsing. If the parser and the generator are mechanistically non-identical, it is possible that this serial/parallel (also called depth-first/ breath-first) parameter is set differently in retrieval operations, combinatorial operations, or both. A challenge for the processor identity view is to align this parameter between the two tasks.

### 1.2.4 Incrementality and cost metric

It is generally assumed that the parser and the generator are shaped by severe limitations of short-term/working memory. This means that models of human parsing and generation commonly include a metric for determining memory cost. If parsing and generation are processor identical, then it is expected that the memory costs incurred by the parser and the generator should be evaluated by the same metric.

In the memory literature, it is generally assumed that there is a positive correlation between the number of items held in memory and memory costs. This positive correlation might be due to capacity limits (e.g., Baddeley & Hitch, 1974 ; Cowan, 2005; Miller, 1956) and/or due to similarity-based interference (e.g., Gillund & Shiffrin, 1984; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006), depending on the theory of memory one adopts. Regardless, in order to assess a cost in parsing and generation, it is necessary to define what counts as a single item over which the number of items that occupy the capacity limited memory slots and over which similarity across items are defined. In other words, it is necessary to define what forms a chunk.

Despite the positive correlation between the number of items held in memory and memory cost, the difficulty of parsing or generation does not seem to increase as a linear function of the number of words in a sentence. One explanation of this is that the words that are encountered are integrated, or 'chunked,' before the incoming input overwhelms memory capacity. This chunking strategy is closely related to the notion of incrementality in parsing. Incremental parsing models analyze incoming input as soon as it arrives. To the extent that the analysis of the incoming input involves establishing syntactic relations with the preceding input, the parser establishes coherent relationships between current and prior input. This feature of the parser is largely motivated by considerations of memory cost - it is too costly for humans to store multiple unintegrated subparts of sentences until the end of a sentence. In psycholinguistics, this is sometimes instantiated as the notion of connectedness (Sturt & Lombardo, 2005). If all sub-structures are 'connected' into one coherent structure, then memory cost does not increase as a linear function of the number of words/morphemes in a sentence. Under the view that the parser and the generator are processor identical, the same function should work in reducing the memory cost in generation. However, in incremental generation models (Kempen & Hoenkamp, 1986; De Smedt, 1990; 1996; Levelt, 1989), the primary function for reducing memory cost is not chunking-like function, but the strategy that send the information in a particular level of representation to the next level of representation. This strategy is sometimes called *cascading*. Consequently, they evaluate the memory cost in a different way than in parsing, and the strategy of reducing cost is also different than in parsing. This discrepancy in how the memory cost is defined and how the memory cost is reduced challenges the view that the parser and the generator are processor identical.

### *1.3 Order identity*

The order identity view states that the order in which different parts of structures are built is identical between parsing and generation. However, this view faces a challenge, because the difference in starting points between parsing and generation suggests that the order in which information corresponding to a particular part of a sentence becomes available can easily be different. If a particular part of a sentence structure becomes available in a different order, the order in which structures are built can reasonably be different.

The order in which structures are built is intimately related to the size of the pre-stored structural units. Generally, the challenge for the order identity view grows bigger as the pre-stored structural units get smaller. As discussed above, it is in principle possible that the parser and the generator retrieve a large structural unit that covers a whole sentence structure, even down to the lexical level. If this is the case, there is no ‘order’ in building structure, except in the process of lexical insertion. This means that, if the structural units in memory are extremely large and detailed, the order identity view could be vacuously true. The parser and the generator just retrieve a sentence at some point, without really ‘building’ structures. However, such a view is problematic given the creative nature of human language. As soon as multiple structural parts need to be retrieved, the parser and the generator must determine the order in which structural units combine. Such combinatorial operations could build structures from top to bottom (i.e., top-down structure building), or from bottom to top (i.e., bottom-up structure building), or a mixture of both (e.g., left-corner parsing). This possible variation in how the structure building procedure interacts with the order in which the particular part of structures are retrieved from memory might differ between parsing and generation. The challenge for the order identity view is thus to figure out whether the parser and the generator can plausibly have the same order in which the pieces of structures are retrieved and the same manner of combining the retrieved units, given the two very different inputs.

## *1.4 Outline of the thesis*

Once the processor identity and order identity views are adopted, there is no motivation to keep the model of parsing and the model of generation separate. In the rest of this thesis, I aim to show that the challenges for the processor and order identity views outlined above can be overcome. The end result of this exercise is a single model of parsing and generation. This is the basic goal of this thesis.

The rest of this thesis is structured as follows. In Chapter 2, I will present new experimental evidence suggesting that the grain size of the structural units is rather small, contrary to some models of sentence production. In particular, I will show that the internal structure of the verb phrase in a ditransitive sentence (e.g., *The chef is donating* *the book to the monk*) is not specified at the onset of speech, but is specified before the first internal argument (*the book*) needs to be uttered. I will also show that this timing of structural processes still is earlier than the lexical processes of verb internal arguments. These two results in concert show that the grain size of structure building in sentence production is rather small, contrary to some models of sentence production, yet structural processes still precede lexical processes. I argue that this view of generation resembles the widely accepted model of parsing that utilizes both top-down and bottom-up structure building procedures.

In Chapter 3, I will present new experimental evidence suggesting that the structural representation constrains the subsequent lexical processes. In particular, I will show that conceptually similar lexical items interfere with each other only when they share the same syntactic category in sentence production. The mechanism that I call *syntactic gating*, will be proposed, and this mechanism characterizes how the structural and lexical processes interact in parsing and generation. I will present two Event Related Potential (ERP) experiments that show that the lexical retrieval in (predictive) comprehension is also constrained by syntactic categories. I will argue that the syntactic gating mechanism is operative both in parsing and generation, and that the interaction between structural and lexical processes in both parsing and generation can be characterized in the same fashion.

In Chapter 4, I will present a series of experiments examining the timing at which verbs' lexical representations are planned in sentence production. It will be shown that verbs are planned before the articulation of their internal arguments, regardless of the target language (Japanese or English) and regardless of the sentence type (active object-initial sentence in Japanese, passive sentences in English, and unaccusative sentences in English). I will discuss how this result sheds light on the notion of incrementality in generation.

In Chapter 5, I will synthesize the experimental findings presented in this thesis and in previous research to address the five challenges to the single system view I outlined in Chapter 1. I will then conclude by presenting a preliminary single system model that can potentially capture both the key sentence comprehension and sentence production data without assuming distinct mechanisms for each.

# 2 Structure building in sentence production

Despite the fact that relatively comprehensive models of sentence production are available (e.g., Garrett, 1980; 1988; Levelt, 1989; Bock & Levelt, 1994; V. Ferreira & Slevc, 2007; Bock & Ferreira, 2014), the details of how structural representations develop in generation remain less well-understood than in parsing. This is not surprising; having precise control over what speakers say is much harder than having precise control over what comprehenders hear or read. This methodological challenge creates a discrepancy between parsing and generation in terms of how detailed models can describe the step-by-step processes of structure building as sentences unfold over time. Such a discrepancy makes it hard to compare structure building models in parsing and generation. This chapter aims to minimize this discrepancy, and to promote the single system view as discussed in Chapter 1. To do so, I will discuss structure building procedures in parsing and generation in parallel terms, identify potential disagreements in terms of how structures are built in parsing and generation, and eliminate such disagreements via an experiment.

## *2.1 Manner of structure building in generation*

Sentence production involves the problem of serial ordering (Lashley, 1951), and hence hierarchical planning. Some early models of sentence production accordingly claimed that sentence production is controlled by a hierarchical plan that spans an entire clause or possibly even two (Garrett, 1975; 1980; 1988; Ford, 1982; Ford & Holmes, 1978; Griffin & Weinstein-Tull, 2003; F. Ferreira & Swets, 2005; Bock & Cutting, 1992; Ferreira, 1991). Important bodies of evidence for such hierarchal planning in sentence production include speech errors (Garrett, 1975; Fromkin, 1973), and a large body of psycholinguistic experiments investigating syntactic priming (see Ferreira & Pickering, 2008 and reference therein). Studies using reaction times and/or hesitation measurements have also shown increases in processing time as a function of the complexity of yet-to-be-spoken parts of a sentence (e.g., F. Ferreira, 1991; Lindsley, 1975; Kempen & Huijbers, 1983).

Hierarchical planning in speech production is often assumed to be the planning of syntactic structure devoid of lexical content. Such a view is generally shared in various versions of 'slot-and-filler' models of sentence production (e.g., Garrett, 1975; 1980; 1988; Bock, Irwin, & Davidson, 2004; Bock, Irwin, & Levelt, 2003; Griffin & Bock, 2000). For instance, Garrett (1988) proposed a model of sentence production in which lexical processing is temporally preceded and guided by the creation of structural representations, encompassing a clause or even two. This claim is often motivated by speech error evidence. For instance, word exchange errors (e.g., *the cat is out of the house* -> *the house is out of the cat*) involves distant words, and is constrained by syntactic categories such that exchanges between words from different categories (e.g., noun vs. verb) rarely occurs (see also Fromkin, 1973). One way to interpret this is to assume the existence of structural representations that span across the exchanged words, which span a long-distance.

If hierarchical planning is planning of syntactic structure without lexical content, then it resembles the intermediate stages in top-down structure building processes. In top-down structure building, category nodes are created before the lexical nodes are, and there exist some points in time where a syntactic 'slot' is yet to be filled with lexical items, just like in hierarchical planning. This is the sense in which I identify hierarchical planning in generation and top-down structure building in parsing.

Under the view that the hierarchal planning is identified with top-down structure building, models of sentence production can be recaptured relatively straightforwardly in the theoretical terms of parsing models. Namely, structure-driven production models build structure in a radically top-down fashion, whereas word-driven models build structure in a radically bottom-up fashion. This contrast is schematically illustrated by Figure 2-1ab, in which the steps of structure building procedures in production are roughly specified. This does not mean that all models of sentence production can be characterized as involving either purely top-down or bottom-up structure building processes. For instance, some recent models of production allow some flexibility in terms of the two 'modes' of sentence production (Kuchinsky, et al., 2011; van de Velde et al., 2014; Bock & Ferreira, 2015). However, such flexibility involves flexibility across different utterances, and hence these models still build the structure of a single sentence either in a radically top-down or bottom-up fashion.

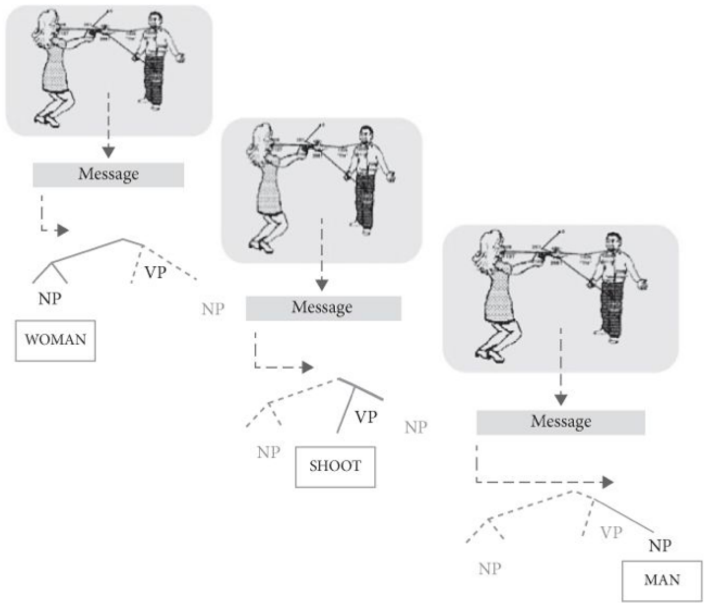



Figure 2-1a: A schematic illustration of radically top-down approach (sometimes called hierarchical incrementality approach) in sentence production (adopted from Bock & Ferreira, 2014).

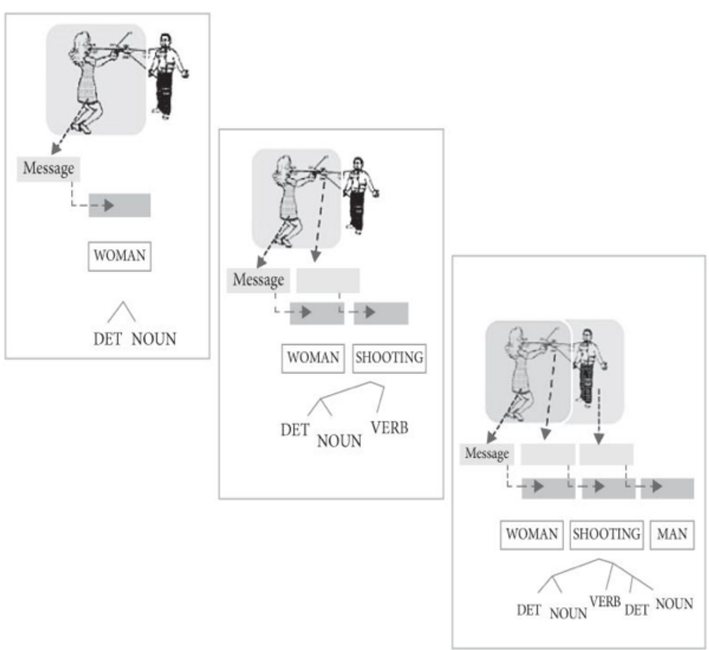


Figure 2-1b: A schematic illustration of radically bottom-up approach (sometimes called linearly incrementality approach) in sentence production (adopted from Bock & Ferreira, 2014).

There is a good reason why the existing models of sentence production diverge in whether structures are built bottom-up or top-down; one body of evidence suggests the involvement of bottom-up structure building while another body of evidence suggests the involvement of top-down structure building. First, the evidence for bottom-up structure building comes from availability effects on word order. For instance, Bock (1987) showed that semantically primed words tend to appear early in a sentence. This kind of study suggests that ease-of-retrieval, i.e., availability, of a word influences its syntactic position in a sentence (see Slevc, 2011 for a review). If the availability of a word influences syntactic structure, the logic goes, the word must come before the syntactic representation. This thus counts as evidence for bottom-up structure building, where structure grows from the leaves, i.e., words.

Second, the involvement of top-down structure building has been motivated by a constraint on speech errors known as the syntactic category constraint. When speakers exchange or substitute lexical items in a sentence, it is overwhelmingly the case that the misproduced lexical item belongs to the same syntactic category (Garrett, 1975; Nooteboom, 1969; Dell et al., 2008; Fromkin, 1984). This suggests that, before the process of lexical selection completes, there must be a syntactic representation that specifies at least the syntactic category. The existence of such a constraint on speech errors has also been experimentally demonstrated, and has also been shown to apply to errors that on the surface seem to violate such a constraint (V. Ferreira & Humphreys, 2002). This strongly supports the involvement of top-down structure building in generation: the existence of syntactic position before lexical retrieval. Some studies using what is called picture-word interference paradigm might also have demonstrated the analogous effects in experimental settings (Pechmann & Zerbst, 2002; Pechmann, Garrett & Zerbst, 2004; Vigliocco, Vinson & Siri 2005; cf. Janssen et al., 2010). Since these two bodies of evidence point toward seemingly conflicting positions with respect to the manner of structure building, it is not surprising that existing models differ in how they build structures.

### *2.2 Grain size of structural processing in generation*

The notion of hierarchical planning is also tightly related to the grain size of structural processing in generation, and hence the grain size of pre-compiled structural memories that the generator uses. The grain size of structural processing can be defined roughly as follows: it is the unit of structural processes within which there is no interleaving between structure building processes and post-structural processes, including lexical insertion[4] and articulation. Such a unit determines how extensive hierarchical planning needs to happen before a structure can start to be filled with lexical items, and hence before the sentence can start to be articulated. The grain size of structural processing directly reflects the units of pre-compiled structural memory; the reason why there should be no interleaving between structural and post-structural processes within a

[4] assuming a non-lexicalized grammar

single structural unit is because the structural processing of that unit is a single process of retrieving a pre-compiled structural memory.

Models of sentence production disagree in the grain size of structural processing in generation. One one hand, models that emphasize hierarchical planning often assume that structural representations are planned extensively at the outset (Garrett, 1975; Ford, 1982; Ford & Holmes, 1978; Griffin & Weinstein-Tull, 2003). That is, in such models syntactic processes are not interleaved with post-syntactic processes within a single clause.[5] This extensive structural planning is often characterized as a characteristic feature of 'non-lexicalist,' 'slot-and-filler,' or *structure-driven* models (e.g., Bock & Ferreira, 2015).[6] On the other hand, other models of sentence production assume much less extensive, possibly negligible, hierarchical planning and hence small structural units. Notably, Kempen & Hoenkamp (1983) proposed a model of sentence production in which structures are built as lexical items become available (see also, Levelt, 1989; De Smedt, 1990; 1996). In such a model structural processes and articulation interleave frequently. Models of this kind are often described as lexicalist models, or *word-driven*

---

[5] This does not mean that the complete representation of a sentence is planned before the onset of articulation in such models. For instance, articulation can start before all the structural positions are filled with lexical items.

[6] However, structure-driven models can be lexicalist in sentence production. For instance, F. Ferreira's (2000) lTAG-based model of production is strictly lexical in the sense that it adopts a lexicalized (≠ lexicalist) grammar as its formal basis, but involves extensive planning of empty structural slots to be filled at the beginning stage of sentence planning. Thus, some models in both the non-lexicalist and the lexicalist approaches create structural slots for an entire clause before lexical insertion starts.

*models*, in contrast to slot-and-filler models.[7] Thus, there are largely two contrasting views on how structures grow in sentence production: structure-driven models that tend to assume extensive hierarchical planning before articulation onset, and word-driven models that tend to assume little hierarchical planning before articulation onset. These two views assume rather different grain sizes of structural processing in generation.

### *2.3 Comparing parsing and generation*

Both the structure-driven and word-driven models of sentence production are at odds with widely supported parsing models, in terms of the grain size of structure building and/or in terms of the manner of structure building. Most parsing models assume that the grain size of structure building is small, and that both top-down and bottom-up structure building procedures are involved in building a single sentence structure (e.g., Johnson-Laird, 1983, Abney & Johnson, 1991, Resnik, 1996, Staub & Clifton, 2006; Lewis & Vasishth, 2005; among others). In contrast, structure-driven models in generation assume a larger grain size of structure building, and hence do not align well with widely assumed parsing models in terms of the unit of structure building. Word-driven models fare better in terms of the alignment of the grain size of structure building; articulation can start before the entire structure of a clause is specified, meaning that structural processes can interleave frequently with post-structural processes up to articulation. However, neither the structure-driven nor the word-driven models align well with the widely supported parsing models in terms of the method of structure building. Namely, parsing models often assume that both top-down and bottom-up structure building methods are involved in parsing a single sentence, while generation models either assume that structures are built in a radically top-down method or that structures are built in a radically bottom-up method.

If these misalignments between parsing and generation are accurate, then it poses a problem to the single system view. The single system view requires that both parsing and generation have the same method of structure building, meaning that whether a specific part of a sentence is elaborated via a top-down or bottom-down structure building processes. Also, the processor identity claim requires that both parsing and generation utilize the same sized unit of structure building. In order to investigate these misalignments, I re-evaluate the assumptions about how structures are built in sentence production. Before I do so, however, I first review the evidence that parsing involves both top-down and bottom-up structure building methods, and the evidence that parsing is strongly incremental, i.e., involves small structural units.

### 2.3.1 Manner of structure building in parsing

Since the old days of psycholinguistics, it is known that neither a purely bottom-up nor purely top-down approach can adequately capture human parsing behavior. For instance, Miller & Chomsky (1963, see also Johnson-Laird, 1983) argued that a psychologically plausible model of parsing should involve both bottom-up and top-down structure building methods.[8] This claim was derived from the observation that left-branching structures (2-1) and right-branching structures (2-2) are about equally easy to process.

(2-1) *John's brother's cat despises rats.*

(2-2) *This is the dog that chased the cat that bit the rat that ate the cheese.*

Contrary to this observation, a purely bottom-up parser cannot parse (2-1) without incurring a substantial memory cost. This is because each phrase cannot be structurally integrated together until the end of the sentence is reached in purely bottom-up parsing, (assuming a traditional phrase structure grammar). To the extent that holding multiple unintegrated items in memory is costly, purely bottom-up parsing should predict difficulty in parsing (2-1). On the other hand, a purely top-down parser cannot parse (2-2) without incurring substantial memory cost. This is because a large chunk of structure must be built even before the first word can be matched in purely top-down parsing. However, observationally, sentences (2-1) and (2-2) are both easy to parse. Based on this observation, Miller & Chomsky (1963) claimed that the human parser must be performing both bottom-up and top-down structure building. Along this line, Johnson-Laird (1983), Abney & Johnson (1991), and Resnik (1996) all argued that this observation suggests the involvement of both top-down and bottom-up parsing, albeit on somewhat different grounds (see Resnik, 1996 for discussion).

---

[8] Even without considering psychological plausibility, a purely top-down parser can also be excluded as a plausible parser of natural language. For instance, it cannot deal with left-recursive rules (Kimball, 1975) in which the right hand side of the phrase structure rule contain the symbol of on the right hand side of the rule, e.g., NP -> NP S. This left recursion does occur in natural languages.

One parsing strategy that is often argued to approximate human parsing is left-corner parsing, where both top-down and bottom-up structure building are involved (Johnson-Laird, 1983; Abney & Johnson, 1991; Resnik, 1992). The major advantage of this is that the parser can accommodate the incoming lexical input on a word-by-word basis without making too many implausible predictions. A purely bottom-up parser often cannot accommodate input on a word-by-word basis, while a purely top-down parser makes many implausible predictions in the absence of evidence. A general rule of thumb is, as pointed out by Johnson-Laird (1983), predicting the beginning of a constituent is harder than predicting the following parts of the constituent. Left-corner parsing embodies this rule of thumb. It processes the beginning of a constituent in a bottom-up fashion while processing the following part in a top-down fashion. This strategy minimizes the risk of making too many implausible predictions while maintaining immediate incorporation of incoming words into a coherent structural representation. This left-corner strategy is adopted in modern computational models of sentence comprehension such as in Lewis & Vasishth (2005).

It is relatively uncontroversial that at least some bottom-up structure building is involved in human parsing. Otherwise,  Thus, I specifically focus on the evidence that top-down structure building is involved. The past few decades of experimental studies made it clear that human parsing involves a fair amount of top-down structure building. The first line of evidence comes from a study on syntactic category prediction at the phrasal and lexical levels. First, Wright & Garrett (1984) embedded the lexical decision task in context, and compared the response time for the bolded words in the following sentences.

(2-3) a. *The towers on the horizon indicate that the barriers isolate* ***translation.***

(2-3) b. *The crowd near the church indicates that an important funeral* ***translation.***

They found that reaction times were faster for (19a) than for (19b). Given that neither sentential context can be used to predict the concept of *translation*, this difference can be attributed to the prediction of syntactic category - a sign of top-down structure building.

More recently, Staub & Clifton (2006) used eye-tracking to extend Wright & Garrett's finding. Namely, they found that comprehenders not only predict at the single-word level, they also make phrasal level predictions - a sign of more extensive top-down structure building. They compared reading times for the following types of sentences.

(2-4) *a. Either Linda bought the red car or her husband leased the green one.*

(2-4) b. *Linda bought the red car or her husband leased the green one.*

(2-4) c. *The team took either the train or the subway to get to the game.*

(2-4) d. *The team took the train or the subway to get to the game.*

In (20a) and (20c), the coordination with *or* is predictable based on the presence of e*ither*, and moreover the position of *either* predicts the type of coordination (S-coordination in (20a) and NP-coordination in (20c)). They found that the reading time for the latter part of the coordination in (20a) and (20c) is faster than that in (20b) and (20d), respectively. Furthermore, they found that comprehenders were less surprised by S-coordination, with the presence of *either* at the beginning in (20a). These results suggest that, based on *either*, comprehenders predict the structure of the latter part of the coordination, and use those predictions to facilitate processing.

Finally, Lau et al. (2006) investigated syntactic category prediction using ERPs. They examined the brain response to the following types of ungrammatical sentences (this presents only the relevant part of their design).

(2-5) a. *Although Erica kissed Mary's mother, she did not kiss Dana's **of** the bride

(2-5) b.*Although the bridesmaid kissed Mary, she did not kiss Dana's **of** the bride

They compared the brain response to the preposition *of* (bolded). The difference between these two sentences is following. In (2-5a) at *Dana's*, comprehenders can recover the elided noun *mother*, so they do not have to (or even prefer not to) predict an overt noun to appear next. In contrast, in (2-5b), such an ellipsis analysis is unavailable, and thus comprehenders cannot help but predict a noun. Thus, when the preposition *of* appears next, the category prediction is violated only in (2-5b). According to this logic, they found an attenuated early left-anterior negativity (eLAN) response to (2-5a) compared to (2-5), which is taken to be an index of phrase structure violation (Friederici et al., 1993; Hahne & Friederici, 1999; Neville et al., 1991). The important aspect of this study is that *of* in both (2-5a) and (2-5b) is an ungrammatical continuation involving a phrase structure violation, but nevertheless the eLAN is attenuated in (2-5a). This suggests that comprehenders actively predict the structure of an upcoming noun in (2-5b) but not in

(2-5a), supporting the existence of the category-level prediction, a natural consequence of top-down structure building.

Of course, I do not argue that the human parsing mechanism is a left-corner parser exactly. Indeed, any context-free parsing mechanism cannot possibly be the exact model of human language, which is known to be at least mildly context-sensitive (e.g., Shieber, 1985). For the current purpose, however, it suffices to claim that both top-down and bottom-up methods are involved in building structural representations of a sentence in an interleaved fashion. Left-corner parsing is a convenient example for demonstrating this point. The actual human parsing mechanism is likely to be much more complex, and it remains an empirical question when the parser built a particular part of structure via bottom-up vs. top-down methods. Nevertheless, this mixed method of structure building in parsing and the above-described views on how structure is built in sentence production do not align, and this discrepancy causes a challenge to the single system view as advocated in Chapter 1. In order to maintain the single system view, it is necessary to pursue either of the following options. The first option is to re-evaluate structure building in parsing and show that it involves larger structural units than usually assumed and that it involves either top-down or bottom-up structure building methods and not both. The second option is to re-evaluate structure budding in generation and show that it involves smaller structural units than often assumed in structure-driven models and that both top-down and bottom-up structure building methods are utilized in generation. I take the second approach in this chapter.

2.3.2 Size of structure building in parsing

Most psychologically-oriented models of parsing assume that structure building is incremental, meaning that structural analysis is performed as each new piece of input arrives, without waiting for further input. There is a strong body of evidence suggesting that parsing is incremental, and evidence ranges from speech-shadowing (Marslen-Wilson, 1975), self-paced reading (Aoshima, Yoshida & Phillips, 2009), eye-tracking (Tannenhaus et al., 1995), and ERPs (Kutas & Hillyard, 1980). Among those studies, studies on head-final language processing, like Japanese sentence processing (Mazuka & Itoh, 1995; Kamide et al., 2003; Aoshima et al. 2009), are particularly informative on this issue. Among the studies and observations that suggest that that sentence processing is incremental, Kamide et al. (2003) and Aoshima et al. (2009) provide particularly compelling evidence. First, Kamide et al. (2003) investigated the following types of sentences in Japanese, using visual world eye-tracking (Tannenhaus et al., 1995).

(2-6) a. *weitoresu-ga kyaku-ni tanoshige-ni humburger-o ….*

waitress-NOM customer-DAT happily hamburger-ACC

(2-6) b. *weitoresu-ga kyaku-o tanoshige-ni…*

waitress-NOM customer-ACC happily

They showed that participants anticipatorily fixated on a picture of the hamburger in the visual scene more given (2-6a) than (2-6b), before the hamburger is actually heard. This pattern suggests that, based on the dative case marking of the second noun phrase in (2-6a), Japanese speakers can anticipate that another noun is upcoming. To the extent that the interpretation and prediction requires structural representation, this suggests that some kind of syntactic representation is already available before the verb is heard. However, as Aoshima et al. (2009) pointed out, it is not clear whether this evidence shows that the connected syntactic structure is represented before the verb. For instance, comprehenders might keep track of the valency and saturation status of the yet-to-be-heard predicate, without contracting a detailed syntactic structure over which more subtle hierarchal relations can be specified. Aoshima et al. (2009) provided evidence suggesting that the pre-head structure is detailed enough to even allow binding relations to be computed. This suggests that the syntactic structure over which a tree-geometric notion like c-command is available.

These pieces of evidence suggests that structure building in parsing occurs strongly incrementally, small chunk by small chunk even in a language where the head of a clause is put in the final position. This incremental nature of structural analysis suggests that the size of structural units utilized in parsing is small, unlike in some structure-driven models of generation that assume that structural units are large.

## *2.4 Experiment 1: Timing of lexical vs. structural planning*

The comparisons between parsing and generation above reveal two discrepancies between parsing and generation models, one with respect to the manner of structure building, and the other with respect to the size of the structural units. Experiment 1 aims to test whether these discrepancies are empirically unwarranted, by examining the relative time-course of syntactic and lexical processes. First, Experiment 1 tests whether the planning of syntactic structure associated with a verb phrase occurs at the beginning of an utterance, as claimed by some models that assume a broader scope of planning for syntactic structure (Bock & Cutting, 1992; Ford & Holmes, 1978; Garrett, 1988), or whether it occurs at a point immediately before the relevant part of the verb phrase is uttered, as claimed by models that assume a limited scope of planning (Kempen & Hoenkamp, 1987; V. Ferreira, 1996). Second, Experiment 1 also examines whether structural planning precedes or follows lexical planning of the relevant portion of the utterance. It specifically tests whether syntactic planning of the internal structure of the verb phrase precedes or follows lexical planning of verb phrase internal argument nouns. This is informative about whether syntactic planning precedes lexical planning, as claimed by models that build structure top-down; e.g., Garrett, 1975) or whether lexical planning precedes syntactic planning, as claimed by models that build structure bottom-up; e.g., V. Ferreira, 1996).

To address these questions, in this experiment I examine the timing of structural and lexical repetition effects with respect to the arguments of ditransitive verbs. The timing at which structural and lexical repetition effects on verb phrase arguments can be observed is informative about the timing at which the repeated syntactic and lexical processes of verb phrase arguments occur. The assumption that I adopt here is that syntactic and lexical processes are faster and take less time when they are repeated than when they are not. Comparing the relative timing of such facilitatory effects of syntactic vs. lexical repetition reveals the relative timing of structural vs. lexical processes. The radically top-down approach generally assumed in structure-driven models predicts that syntactic repetition effects should appear before a sentence begins to be articulated, well before lexical repetition effects can appear. In comparison, the radically bottom-up approach as generally assumed in word-driven models predicts that lexical repetition effects should precede structural repetition effects with respect to the same part of a sentence. Finally, the mixed approach that I advocate here predicts that structural repetition should appears earlier than lexical repetition effects, but not as early as the sentence onset. This is because, in a mixed approach, like in left-corner parsing, the internal structure of a verb phrase is planned before the arguments of verbs are planned, but it is not extensively planned at the beginning of sentence.

In the current experiment I specifically examined the production of ditransitive sentences with non-alternating verbs, and examine how production processes for those sentences are affected by recent exposure to the same (Prepositional object: PO) or different (Double object: DO) ditransitive structures, and by recent exposure to the same or different theme or patient argument nouns. If structural processes are extensively performed at the beginning of production processes, it is expected that the speed-up effect due to structural repetition should be observed at the beginning of the sentence, i.e., as a shorter sentence onset latency. On the other hand, if the structural processes only occur immediately before the relevant portion of the sentence, it is expected that the speed-up effect due to structural repetition should not be observed until a later point in the sentence.

I exclusively used non-alternating ditransitive verbs in Experiment 1. This choice is because the dependent measure suitable for investigating the current timing question is the latency and duration of each word. These measures were necessary to assess the current question about the scope of syntactic planning, because the traditional measure of syntactic priming, i.e., proportion of PO/DO usage, does not contain any temporal information about the underlying processes. Given that the dependent measure was onset latency and duration, it was necessary to compare the same utterances across conditions. Using non-alternating verbs increases the consistency of the utterances, thereby making it possible to compare timing measures across conditions without losing too many data points.

### 2.4.1 Method

*Participants*

30 undergraduate students from the University of Maryland community participated for class credit. They were all native speakers of English.

*Materials and Design*

70 pictures of events corresponding to non-alternating ditransitive verbs (e.g., *deliver*) were created by combining pictures of human characters and object pictures Those verbs were chosen based on Levin’s (1993) English verb classification. The verb that participants were instructed to use was printed at the bottom of the picture. A sample picture is shown in Figure 2-2.


.

Figure 2-2: An example picture stimulus used in Experiment 2. Target sentence: *The clown is delivering the ruler to the waitress.*

Additionally, a total of 350 sentences were created (70 PO sentences, 70 DO sentences, 140 transitive sentences, and 70 intransitive sentences), to be used as prime sentences that were presented before the picture description phase. The PO sentences (e.g., *John gave the book to him*) were used to create the PO prime condition, and DO sentences (e.g., *John gave him the book*) were used to create the DO prime condition. The lexical contents for these prime sentences were exactly matched. Half of the transitive sentences contained an object noun that either corresponded to the goal entity (Goal prime; e.g., *John saw the waitress*) or the theme entity (Theme prime; e.g., *John saw the ruler*) in the paired picture. The rest of the sentences were the intransitive baseline condition, which was neither syntactically or lexically similar to the target ditransitive sentences. The set of subject nouns was matched across all five conditions. In addition, 70 filler trials also consisted of sentence-picture pairs, but they differed in the content of the pictures and sentences. They contained transitive, intransitive, and ditransitive pictures with an appropriate verb printed below the picture. All of the ditransitive pictures corresponded to a verb that allows alternation between PO and DO structures. These ditransitive pictures were sometimes paired with prime sentences with a DO structure, in order to encourage the use of DO structures for describing those ditransitive filler pictures. The filler trials were interspersed randomly between the experimental trials. Five stimulus lists were created and presented to participants according to a Latin-Square design. No participant saw the same picture or the same sentence more than once in the experiment.

The current study did not adopt a factorial design, and had two separate sub-experiments. First, the syntactic priming conditions contained PO prime and DO prime conditions. These conditions are the main interest of the current study, and they test specific hypotheses about the scope of structural planning as described above. Second, the lexical priming conditions contained theme prime, goal prime, and intransitive baseline sentences. A rationale for including an intransitive baseline in this comparison is provided below in the Discussion. This is a secondary interest, but it is informative about (i) whether region-by-region duration measures, which are rarely used in sentence production research, can be informative about the timing of lexical priming, and (ii) whether lexical planning of the direct and indirect object precedes or follows the syntactic priming effect. Regarding (ii), if lexical planning of direct/indirect objects precedes the relevant syntactic planning, the pattern would be consistent with the bottom-up structure building view of sentence production, with respect to the construction of verb phrases. If the lexical repetition effect does follow the structural repetition effect, it provides evidence for top-down structure building with respect to the construction of verb phrases.

*Procedure*

Participants first studied a booklet containing the human characters and objects used to create experimental as well as filler picture stimuli, with the target name corresponding to each picture. They studied it until they felt comfortable with each picture and sentence. After this familiarization session, the structure of each trial was explained to participants. They were instructed to repeat back the sentences and describe the pictures as fluently and accurately as possible. They were also instructed to use the verb printed on each picture in a sentence when describing the pictures. Following the instructions, they were asked to perform some practice trials that had the same structure as the experimental trials. The experimental session followed this practice session.

Each experimental trial was structured as follows. First, the participant saw a fixation cross at the center of the screen for 300ms. Following a 200ms blank screen, a prime sentence was presented at the center of the screen. Participants spent as much time as they needed to repeat back the sentence, and pressed the space key when ready. At the same time as the key press, another fixation cross appeared on the screen for 300ms. Following a 200ms black screen a picture stimulus appeared on the screen for 5000ms. A blank screen was presented until another key press, which ended the trial. The speech onset time relative to the picture onset, as well as the duration of every word in the utterance, were measured using the automatic speech alignment algorithm *P2FA* (Yuan & Liberman, 2008). Any trials that deviated from the target sentences or that included hesitations or repairs were excluded (11.8% of experimental trials). Any trials with a speech onset time of less than 30ms and more than 3000ms were removed from the data. Also, any trials with a single word duration longer than 2000ms or less than 30 ms were removed as well.[9] This accounted for 3.1% of the error-free experimental trials. These onset and duration measures were all log-transformed and then submitted to statistical models.

### 2.4.2 Results

*Syntactic priming conditions*

The difference in mean onset latency between PO and DO conditions for each word is shown in Figure 2-3. There were two specific hypotheses with respect to the timing of the syntactic priming effect. First, models that assume an extensive scope of structural planning predict that the priming effect should occur at the onset of the sentence. Second, models that assume smaller planning units predict that it should occur immediately preceding the articulation of PO/DO structures. Reflecting these two hypotheses, there were two regions of interest, Region 1 (Onset) and Region 5 (Main verb).

[9] This is because a word duration of 30ms is implausibly short. A few trials did show such an implausibly short word duration, probably due to an error in the speech alignment algorithm.

For each of these two regions, a mixed effects model with a maximal random effects structure in the sense of Barr et al. (2013) was constructed. In addition to the experimental factors, the logged sum of other regions was included as a predictor, which improved the fit of the model according to a maximal likelihood ratio test ($p < 0.001$). This was not surprising, because the speed of articulation at a particular region of an utterance correlates highly with the speed of articulation at other regions of the same utterance. The main interest here is in relative speed up and relative slow down, and so the model needed to account for general speech rate.

*Onset latency*

The analysis of onset latencies revealed no effect of prime type (Bonferroni adj. $p > 0.5$).

*Time spent on the verb region*

The analysis of the time speakers spent on the verb region (onset of the verb until the onset of the following determiner) revealed that a DO prime increased the time speakers spent on this region significantly ($\beta = 0.05$, $SE = 0.02$, $|t| = 2.44$, adj. $p < 0.05$).

*Time spent on other regions*

Although the other regions were not included as regions of interest in the initial analyses, exploratory analyses were conducted on those regions. None of those regions showed significant differences between the PO and DO prime conditions ($p$s > 0.5)

*Lexical priming conditions*

The difference in mean onset latency between the Theme prime and Intransitive baseline conditions for each word is shown in Figure 2-4, and the difference in mean onset latency between the Goal prime and Intransitive baseline conditions for each word is shown in Figure 2-5. For the analyses of the lexical priming conditions, I did not have any specific regions of interest determined in advance. Thus, I fully analyzed the entire regions in the target sentences. However, this creates a problem of multiple comparisons, and Bonferroni correction inflates Type 2 errors dramatically. In the current setting, Type 2 errors are as misleading as Type 1 errors because both of the two competing hypotheses predict a facilitative effect of lexical repetition priming effect *at some region*, and the null hypothesis is really not viable under any theory of sentence production, unless one assumes that priming should fail to yield facilitation in time. Thus, I provide the results of the entire analyses without Bonferroni correction. However, the reader should note that this part of the experiment should be seen as an exploratory, not a confirmatory study.

*Onset latency*

The analysis of onset latencies revealed no effect of prime type ($p > 0.5$).

*Time spent at each region*

Dummy coding with the intransitive prime condition as the reference was used for the analyses of each region. The analysis at region 6 (the second determiner region) revealed a significant difference between the intransitive prime and the theme prime conditions ($\beta = -0.07$; $SE = 0.03$; $|t| = 2.10$; $p < 0.05$). That is, participants spent less time between the onset of the second determiner and the following noun in the theme prime condition relative to the intransitive condition. Note that this difference is not so clear in the means in Figure 2-4 below. This is probably because the mixed effects model, but the not the figure, accounts for the duration of other regions (which was included as a predictor in the model). Supporting this conjecture, removing that predictor made the effect non-significant ($p > 0.1$).

The analysis at the Region 8 (*to*) revealed a significant difference between the intransitive prime and the goal prime conditions ($\beta = -0.10$; $SE = 0.04$; $|t| = 2.38$; $p < 0.05$). That is, participants spent less time between the onset of the second determiner and the following noun in the goal prime condition relative to the intransitive condition. No other region showed a significant effect of prime type ($p > 0.2$).

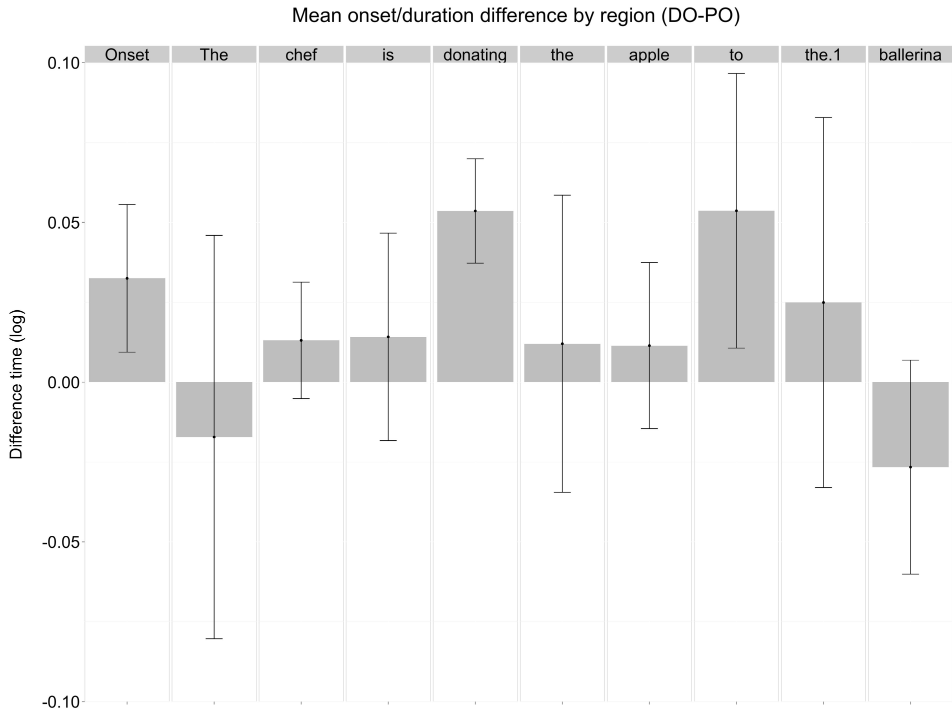



Figure 2-3: The effect of structural repetition in Experiment 1. Mean difference in onset/duration by region between the DO and PO prime conditions. Error bars are standard error of the means.

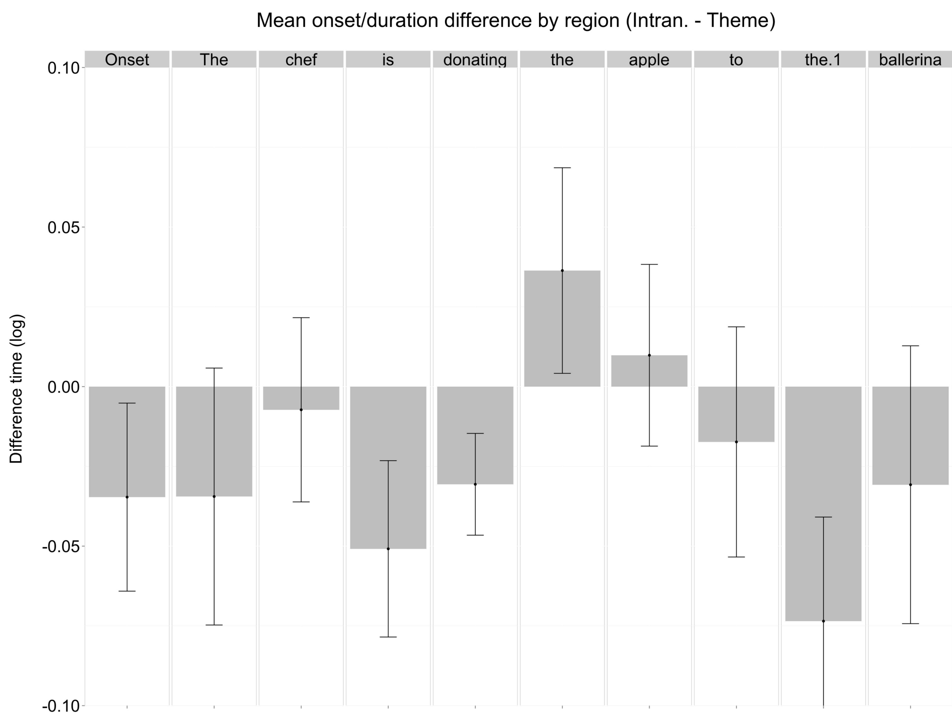



Figure 2-4: The effect of lexical repetition on theme arguments in Experiment 1. Each bar represents the mean difference in onset/duration by region between the intransitive baseline and theme prime conditions. Error bars are standard error of the means.

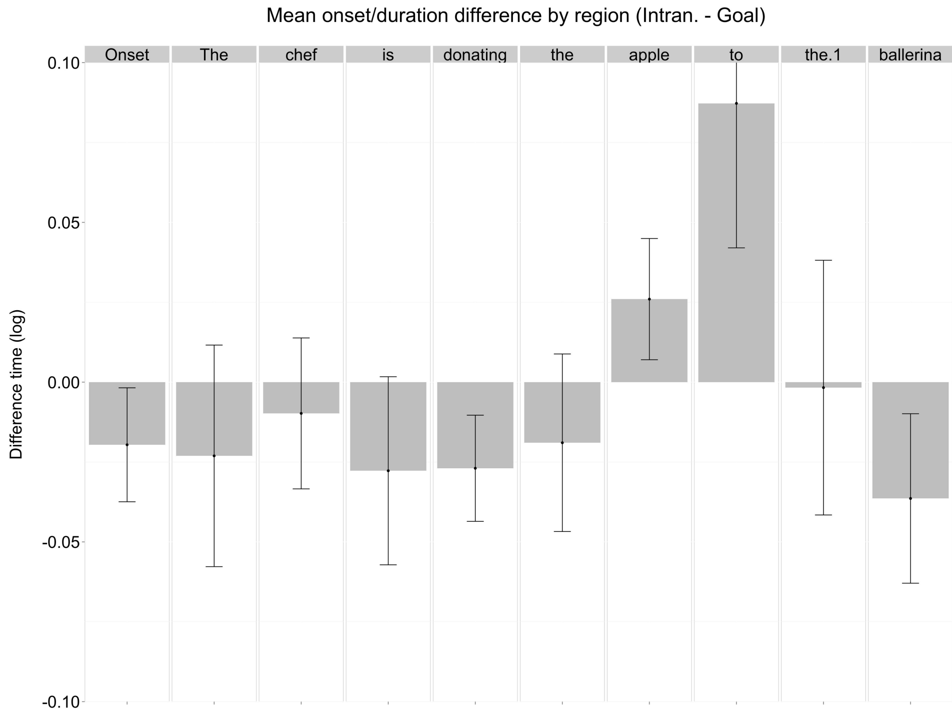



Figure 2-5: The effect of lexical repetition on goal arguments in Experiment 1. Each bar represents the mean difference onset/duration by region between the intransitive baseline and goal prime conditions. Error bars are standard error of the means.

### 2.4.3 Discussion

Experiment 1 examined the timing of syntactic and lexical priming with respect to the verb phrase internal parts of sentences, in order to infer the relative timing of syntactic and lexical processes with respect to ditransitive verb phrases. First, the speed-up due to the PO prime sentences, or equivalently the slow-down due to DO prime sentences, suggests that the syntactic repetition effect appeared after participants started uttering the verb. In contrast, there was no evidence suggesting such a speed contrast at the sentence onset, nor at any point of the sentence other then the verb region. This supports the hypothesis that syntactic planning is not extensive, but unfolds in smaller steps. One way of interpreting this is that the grain size of structural units is relatively small. Namely, it suggests that the structural planning unit does not contain the subject noun phrase and the internal structure of verb phrase at the same time. Indeed, it seems that the internal structure of the verb phrase is planned after the verb head has already started to be articulated. Thus, the size of the structural unit may be even smaller than a single verb phrase.

Furthermore, the analyses of the effect of lexical repetition priming revealed that the facilitatory effect of lexical repetition with respect to the indirect and direct object occurred later than the syntactic repetition effect. This pattern suggests that, although syntactic planning is not extensively done at the onset of the sentences, it still occurs before lexical planning of the relevant parts of the sentences. Neither the radically top-down or bottom-up approaches to sentence production expect this pattern, and I argue that this reflects interleaving between syntactic planning and lexical planning, and that syntactic planning generally precedes lexical planning at each increment of a sentence. Thus, the current results suggest a view of sentence production that neither falls under models that postulate radically bottom-up structure generation (e.g., V. Ferreira, 1996; Levelt, 1989) or top-down structure generation (e.g., Garrett, 1975). This fits well with the dominant view in parsing, where top-down and bottom-up structure building interleave in the analysis of individual sentences.

It must be acknowledged that the current results have a potential alternative explanation. Namely, it is possible that speakers, after encountering DO prime sentences, tend to have DO structural commitments at the beginning of sentence production. This DO structural commitment needs to be retracted later in the sentence in the experimental trials because it is incompatible with non-alternating PO verbs, which were used in all the experimental trials. Since retracting a structural commitment should take time, this could explain the increased duration of the verb in the DO prime condition compared to the PO prime condition. If this explanation is correct, the current results do not necessarily suggest that the scope of hierarchical planning is smaller than often assumed in some models of sentence production. This explanation cannot be conclusively ruled out based on the current pattern. One possible future study that can test this possibility against the current explanation would involve repeating the current experiment using alternating ditransitive verbs. If the alternative possibility presented here is correct, the structural repetition effect at the verb should disappear, because speakers do not need to retract the structural commitment that they made early on, since whichever structure the speakers commit to, the alternating verbs are compatible with either option.

*Aligning parsing and generation*

Based on the current results, I argue that generation involves small-sized units of structure building, but it still involves top-down structure building processes. Given that the evidence from the availability effects reviewed above suggests the involvement of bottom-up structure building processes in production, the resulting model of the generator assumes a mix of both bottom-up and top-down methods in structure building. This view resembles the widely supported model of parsing, i.e., left-corner parsing models. Given this view of the generator, it is possible to identify the parser and the generator in terms of the size of structural units and in terms of the vertical order of structure building. First, the size of structural units in parsing is usually assumed to be small, certainly smaller than an entire clause, and the current results suggest that it is also small in generation, contrary to the assumption of many structure-driven models of sentence production. Second, the vertical order identity view claims that the manner of structure building in parsing and generation is identical, e.g., the partial structures of sentences that are built in a top-down method in parsing should also be built in a top-down method in generation. Of course, these claims do not decisively confirm the single system hypothesis as described in Chapter 1, as many other aspects of parsing and generation need to be aligned as well. However, the current results at least show that the general principles of structure building, both in terms of the manner of structure building and the size of structural units, can plausibly be aligned between parsing and generation.

# 3 The interaction between structure and lexicon in parsing and generation

Chapter 2 discussed the manner and unit size of structure building in sentence production, and argued that both aspects can plausibly be aligned to parsing. The discussion led to the view that the syntactic structure generally precedes lexical processes, i.e., involves top-down structure building, except at the left-edge of phrases as in left-corner parsing. However, it is still not clear how the structures built in a top-down fashion plays a role in the subsequent lexical processing in parsing and generation. In other words, the relationship between structural and lexical processes beyond the mere temporal ordering is yet to be specified. In Chapter 3, I aim to specify how the structural processes constrain the following lexical processes both in parsing and generation, and assess whether the mechanism characterizing such an interaction can plausibly be identical between parsing and generation. This attempt will lead to an introduction of what I call 'syntactic gating' mechanism, which strongly limits the lexical processes based on the category cue provided by the syntactic representation. In describing this mechanism, I will also aim to resolve some of the challenges outlined in Chapter 1. Namely, I will argue that the syntactic gating mechanism is a key mechanism in explaining how the parser and the generator can plausibly build structures in the same order in the same way despite distinct input. I will also argue that the same notion of incremental structure building can be maintained for parsing and generation if one assumes this syntactic gating mechanism.

### *3.1 Structural effects on lexical processes*

A number of previous proposals address the relationship between structural and lexical processes both in comprehension and production. In comprehension, Wright & Garrett (1984) showed that comprehenders can judge the lexical status of a target word in a syntactic context more quickly when the category of the target word has a syntactic category that fits with the syntactic context, as reviewed in Chapter 2. This result may suggest that syntactic structures that are built in a top-down fashion facilitate subsequent lexical recognition process. In production, researchers like Garrett (1980) argued that the structural plan controls the selection of lexical items. Garrett (1980) explained the overwhelming tendency of substitution and exchange speech errors to result in errors that obey what is called the syntactic category constraint - the strong tendency that the misproduced words almost always belong to the same syntactic category as the intended words. Models like Garrett (1980) and its successors assume that syntactic processes constrain subsequent lexical processes in a facilitative fashion. Thus, some researchers in both parsing and generation propose that structural processes constrain lexical processes, though a detailed explanation of this facilitative interaction is elusive. In this chapter, I aim to provide a more detailed characterization of how structural processes constrain subsequent lexical processes, and whether such a structure-lexicon interaction can be characterized in the same terms between parsing and generation.

### *3.2 Mechanism*

Before characterizing the mechanism that defines the interaction between syntactic and lexical processes, I must first discuss the mechanism for lexical processes. For lexical processes, I specifically adopt two widely assumed views of lexical retrieval. First, I take lexical retrieval to be a form of content-addressable memory access. This is a common assumption in lexical processing models (e.g., McClelland & Elman, 1985), although in many cases the term content-addressable memory is not used in describing the model. The basic characteristic of content-addressable memory is that the memory access is based on the content of the memory items, rather than the address or the relative location of items in memory. This memory architecture is susceptible to retrieval interference (e.g., Gillund & Shiffrin, 1984), when some cues used to retrieve the target memory item partially overlap with non-target items. This compares with some models of memory access in which the address in memory is used as a primary resource for retrieving a memory item (e.g., Sternberg, 1969).

Second, I assume that there is a syntactic level of representation for lexical items, along with many word production models (e.g., Levelt, 1989; Roelofs, 1992; Levelt, Meyer & Roelofs, 1999; Kempen & Huijbers, 1983). This is often called *lemma* representation (Kempen & Huijbers, 1983) and it is canonically assumed in the word production literature. The main feature of this lemma representation is that it does not contain any phonological information, but it does contain syntactic features like grammatical gender, subcategorization, and syntactic categories. The existence of such a phonology-free representation is supported by evidence from tip-of-the-tongue phenomenon in which the meaning and the gender feature, but not the phonological shape of a target word is not known to the speaker experiencing the tip-of-the-tongue effect (Vigliocco, Antonini & Garrett, 1997). Event Related Potential (ERP) studies utilizing lateralized readiness potentials (Van Turennout, Hagoort, & Brown, 1998) also suggest that the retrieval of gender features temporally precedes the retrieval of the phonological features of a target word. These pieces of evidence suggest that there is a stage at which the syntactic representation of a word is retrieved without its phonological content, and this level of representation is called the *lemma*. A representational level analogous to lemma representations, however, is not usually assumed in the comprehension literature. This is understandable, because in comprehension lexical retrievals are normally based on sound input. Because lexical retrievals are based on sound input, it makes little sense to assume a processing stage where only syntactic information is represented without phonological information. However, a recent studies on predictive sentence comprehension (Ito et al., 2016) suggests that semantic features of a lexical items are accessed before the phonological features. Thus, I assume the existence of lemma-like representations in which syntactic information of a word is represented independently from phonology, both in comprehension and production.

When those two views on lexical retrieval are combined, lexical access can be viewed as a content-addressable memory access based on a combination of phonological, conceptual, and syntactic features. As syntactic features are part of the content of lexical representation, they should be available for use as a type of retrieval cue.

Once syntactic features can be used as a retrieval cue for specific target lexical item, the lexical retrieval system can interface naturally with the syntactic structure that is built in a top-down fashion. Top-down structure building can construct a category node that needs to be filled with a lexical item before the lexical retrieval process for that position starts. The presence of such an empty category node can provide a syntactic category cue for the subsequent content-addressable lexical retrieval operation. That is, I hypothesize that the top-down structure building process constrains subsequent lexical retrieval processes, just like phonological analysis and conceptual generation constrains lexical retrieval processes in comprehension and production, respectively.

In the current view on lexical retrieval in which multiple sources of information are deployed, a critical question is how those information sources interact with each other. That is, in content-addressable lexical access that assumes multiple cue types, the interaction between the different cue types needs to be characterized. For instance, it is conceivable that syntactic cues and phonological cues independently contribute to lexical access, in which case phonologically relevant non-target items in the lexicon should interfere with target item retrieval due to the partial match of the phonological or conceptual features. This is in accordance with the principle of similarity-based retrieval interference (Gillund & Shiffrin, 1984; Van Dyke & McElree, 2006) On the other hand, it is also conceivable that some cue types are somehow more privileged than the others, in a sense that the privileged cues define how other cue types are used for retrieval. For instance, syntactic cues might be privileged retrieval cues that constrain the use of phonological or conceptual cues, in such a manner that phonological or conceptual partial match does not cause interference if syntactic features fail to match between the retrieval cue and the target item. Such an interactive relationship between different cue types might be beneficial, because privileged types of cues may serve as a ‘filter’ that minimizes the interference due to the partial match of non-target items in other cue types.

In sentence production, there are some reasons to think that syntactic cues are more privileged than conceptual cues. As I mentioned above, the syntactic category constraint in speech errors (Nooteboom, 1969; Fromkin, 1971) can be construed as evidence that lexical retrievals are constrained to only the relevant syntactic category. Under the view that speech errors partially result from improperly resolved retrieval interference, the syntactic category constraint suggests that retrieval interference occurs only within a single syntactic category. That is, only nouns are relevant when retrieving a noun, and only verbs are relevant when retrieving a verb.

In addition, sentence production requires a rapid succession of lexical retrievals, and this poses a working memory problem, regardless of whether one adopts an interference based view or a capacity based view of working memory. It is an intuitive assumption that conceptually important parts of a message tend to come to mind earlier than conceptually minor parts of the message; however, a conceptually important part of a message, for instance the predicate, might not be allowed to come early in a sentence, due to grammatical or other constraints.[10] This mismatch between message-level order and lexical-level order requires either that speakers buffer the retrieved words or they must somehow know when lexical retrieval for each fragment of message needs to be performed. The buffering option is unattractive, because it could create a large amount of memory interference. Alternatively, speakers might somehow postpone certain retrieval operations so that buffering is not needed. This, however, presupposes a mechanism that tells the speakers when to retrieve which part of the sentence, consistent with the grammar of the language and the current sentence context.

I propose that this smart postponing might be possible if syntactic cues serve as a privileged type of cue for content-addressable lexical retrieval. I will call this mechanism the *syntactic gating* mechanism. In this view, syntactic cues are privileged in two senses. First, they are privileged in the sense that the retrieval process for a particular position of a sentence does not begin without the presence of such a cue. Second, they are privileged in the sense that other cue types becomes relevant only when syntactic category cues match. That is, syntactic cues gate candidate words in retrieval.

This mechanism contributes temporal order to successive lexical retrieval operations. As I argued in Chapter 2, top-down structure building in speaking is not

---

[10] Of course, some languages are fortunate enough to allow relatively free word order, in which case there may be a rough correspondence between conceptual importance and order of utterance. Indeed, speakers do exploit word order flexibility to align word order with conceptual prominence (e.g., McDonald, Bock, & Kelly, 1993). However, in many languages under many circumstances, word order is constrained by syntax, among other factors such as pragmatics. For instance, even when the predicate of a propositional message is normally assumed to be important, and therefore can also be assumed to come to mind early, it is often rather odd to front the verb in a sentence in many languages, including English and more flexible word order languages like Japanese.

extensively done at an early point in speaking, and occurs relatively locally as speakers utter a sentence. Thus, this mechanism provides the guidance for when a lexical item of a certain category needs to be retrieved, according to the trajectory of syntactic structure building. Structure building unfolds in time, creating the structural positions to be filled with lexical items, and this top-down structure building process gives temporal instructions for when a lexical retrieval process for a particular position in a sentence needs to be performed. Of course, thematic roles like agents and patients will also be used to select the appropriate lexical items for a particular structural position. However, such information is not an inherent property of lexical items, and thus is not used directly for content-addressable lexical memory access.

One of the predictions from the view that assumes the syntactic gating mechanism is that retrieval interference should occur only within the same syntactic category. That is, distractor non-target words create interference only when they share the syntactic category feature with the target word. In Experiment 2 and Experiment 3 I tested this prediction, using a modified version of the Picture-Word Interference (PWI) task (Lapker, 1979; Schriefers, Meyer, & Levelt, 1990) which I call the Sentence-Picture Interference (SPI) task.

### *3.3 Experiment 2: Syntactic gating in production I*

Experiment 2 investigates the effect of syntactic structure, specifically category labels, on lexical retrieval in sentence production. It specifically tests whether retrieval interference due to conceptual similarity between non-target and target lexical items can be eliminated when syntactic categories mismatch between them.

Before describing the details of the current experiment, it is worth mentioning that there are previous studies investigating the effect of syntactic category on lexical retrieval in the PWI literature (Pechmann & Zerbst, 2002; Vigliocco et al., 2005). Those studies suggest that distractor words that match in syntactic category with the target word cause greater interference on target word retrieval. For example, Vigliocco et al. (2005) demonstrated that syntactic category overlap and semantic similarity had additive interference effects on target naming latency in a PWI task in Italian. Specifically, they asked speakers to describe action pictures using an inflected verb (e.g., t*ossire: coughs*) while ignoring the written distractor verb (e.g., r*uttare*: *burp*) or a noun phrase (e.g., *lo starnuto: the cough*). They found that the interference effect was obtained only when a conceptually similar verb distractor was presented, and not when a conceptually similar noun distractor was presented. Importantly, they reported that this noun-verb contrast was not obtained when the participants were asked to describe the picture using the citation form of the noun or verb. They interpreted this pattern to mean that the phrasal-level processes required to produce inflected verbs make the category information relevant, so that the within-category distractor caused extra interference only in phrasal production. This additive effect is not predicted by the syntactic gating mechanism that prevents the category-irrelevant lexical items from causing interference, because the syntactic gating mechanism makes a syntactically mismatching item irrelevant for the lexical processing at hand. Thus, the syntactic gating mechanism should predict an under-additive, rather than additive, effect of syntactic category and semantic similarity.

However, PWI tasks have important limitations, which make it very hard to manipulate syntactic categories in an interpretable fashion. To manipulate the syntactic category of the distractor in a PWI task, morpho-phonological and/or semantic features that correlate with syntactic category must also be changed. This is especially true for Vigliocco et al.'s materials, as they used entirely different sets of distractor items between the noun distractor and verb distractor conditions. Even if the stems were matched, this problem does not resolve, as, for instance, *cough* as a verb and *cough* as a noun have a different meaning. *Cough* as a noun, for instance, refers to the sound of coughing, not necessarily to the act of coughing; in contrast, *Cough* as a verb cannot refer to the sound of coughing. The close correlation between morphophonology/semantics and syntactic category makes it difficult to interpret the PWI results.

Furthermore, even when these morphophonological and semantic confounds are ignored, there is another problem, which arises from the assumption of automaticity in PWI. The assumption of automaticity is that distractor stimuli are processes automatically so they cannot be entirely ignored. Since distractor stimuli cannot be ignored, they exert facilitatory or inhibitory effects on the processing required for the main task. Certainly, as long as the distractors are kept to a minimal complexity, the assumption of automaticity might hold. However, it is not clear if this assumption holds when distractor *phrases* rather than distractor *words* are used to manipulate syntactic category, as in Vigliocco et al., (2005). This is because it is far from clear if speakers process written phrasal input automatically enough that they engage in combinatorial processes, which are a prerequisite for identifying the syntactic category of the distractor. To the extent that PWI or other Stroop-like tasks rely on the automaticity assumption (Stroop, 1935) the results of PWI studies using complex distractor stimuli must be interpreted with caution.

Instead of PWI, the current experiment uses a novel Sentence-Picture Interference (SPI) task, which is created based on PWI. In the SPI task, instead of isolated distractor words, participants are presented with a distractor sentence. However, instead of ignoring the distractor, they are asked to memorize the distractor sentence. After the memorization phase, participants are sometimes presented with a picture, which they are instructed to describe. In other trials they are prompted to repeat back the memorized sentence verbatim. This procedure ensures that the memory trace of the distractor sentence is active until the beginning of the picture description task. This paradigm side-steps the depth of processing problem for distractors, and furthermore it allows researchers to manipulate the syntactic environment of distractor words without changing other aspects, as shown below.

Using the SPI task, I assessed whether the interference effect due to conceptual/ semantic similarity of the distractor can be blocked when the distractor does not share a syntactic category with the relevant word in the target utterances. Specifically, I manipulated the syntactic category of a key word in the distractor sentence by changing the syntactic environment in which it occurred. For example, compare the following sentences.

(3-1) a. *John told me that the boy is sincerely praying.*

(3-1) b. *John told me about the boy's sincere praying.*

In these sentences, the word *praying* has maximally similar, if not identical, meanings. Nevertheless, it functions as a verb in (3-1a) but as a noun in (3-1b). This manipulation isolates the syntactic category from semantic and morphophonological dimensions, thereby making it possible to see the pure effect of syntactic category on the following picture description task. If this manipulation effectively eliminates the lexical interference effect due to semantic similarity, it provides strong evidence that syntactic structures constrain lexical processes. In other words, it provides strong evidence for a syntactic gating mechanism for lexical retrieval.

### 3.4.1 Methods

*Participants*

48 native speakers of English participated for either class credit or monetary compensation.

*Materials and Design*

24 pictures of events corresponding to either optionally intransitive verbs (e.g., *cook*) or obligatorily intransitive verbs (e.g., *sneeze*) were selected from the UCSD International Picture Database. An example picture is shown in Figure 3-1ab below.

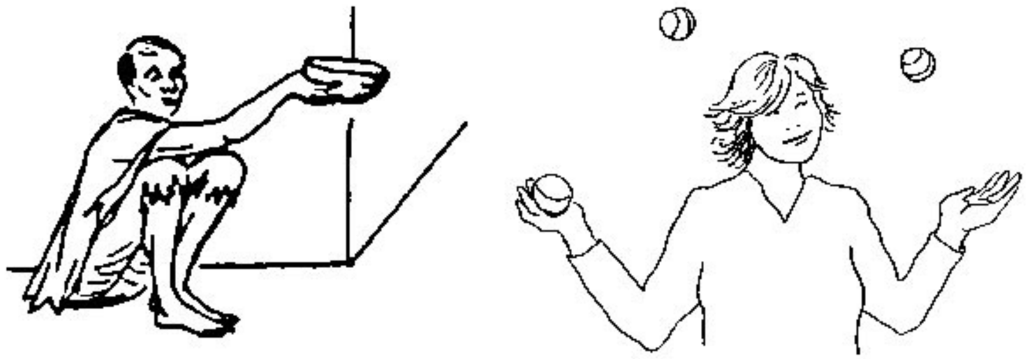


Figure 3-1ab: Sample picture stimuli used in Experiment 2.

For each target picture, a semantically related distractor verb was selected from the set of target verbs for the other pictures. These distractors always corresponded to one of the other target verbs, to maximize the chance of obtaining an interference effect (Roelofs, 1992). Semantic relatedness was defined based on the cosine distance measure from latent semantic analysis (Landauer & Dumais, 1997).

There were two types of distractor sentence, as in (25).

(3-2) a. *John told me that the boy is sincerely praying.*

(3-2) b. *John told me about the boy's sincere praying.*

In (3-2a), the last word *praying* is a verb. In (3-2b), it is a noun. These sentences were paired with a picture with a conceptually similar action (e.g., *begging*), or a dissimilar action (e.g., *juggling*). These sentences were also used as filler trial once, i.e., with a verbatim recall instruction, instead of a picture description task. This was done to prevent participants from predicting the following event in a trial (picture or instruction to repeat the sentence) based on the structure of the distractor sentence.

Thus, the current study had a 2 x 2 design, with Sentence Type and Relatedness both as within-subjects, within-items factors. In total, there were 192 trials, half of which were filler trials where participants were asked to repeat the sentence instead of describing a picture. Each participant saw the same picture four times, once in each condition, in a random order.

*Procedure*

Participants first studied a booklet containing the pictures that they were going to see in the experimental session, with the target sentences corresponding to each picture. They studied the booklet until they felt comfortable with each picture and sentence. The experimental session followed these familiarization sessions. Each experimental trial was structured as follows. First, the participant saw a fixation cross at the center of the screen for 500ms. Then, the distractor sentence was presented at the center of the screen. Participants spent as much time as they needed to memorize the sentence, and pressed the space key when ready to proceed. One second after the key press, in a target trial a picture from the studied set appeared on the screen for 3000ms. A 2000ms blank screen separated each trial. A schematic illustration of a sample experimental trial is presented in Figure 3-2 below. In a filler trial, the picture presentation was replaced with the word 'repeat' on the screen, to which participants responded by repeating back the sentence they had memorized. The screen remained blank until the participants pressed the space key again. The speech onset time from the picture onset to the utterance onset was measured using a simple amplitude threshold detection algorithm in MATLAB. Any trials with a speech onset time less than 400ms and more than 5000ms were removed from the data. This accounted for 3.7% of the all trials. These measures were log-transformed and then submitted to statistical models.

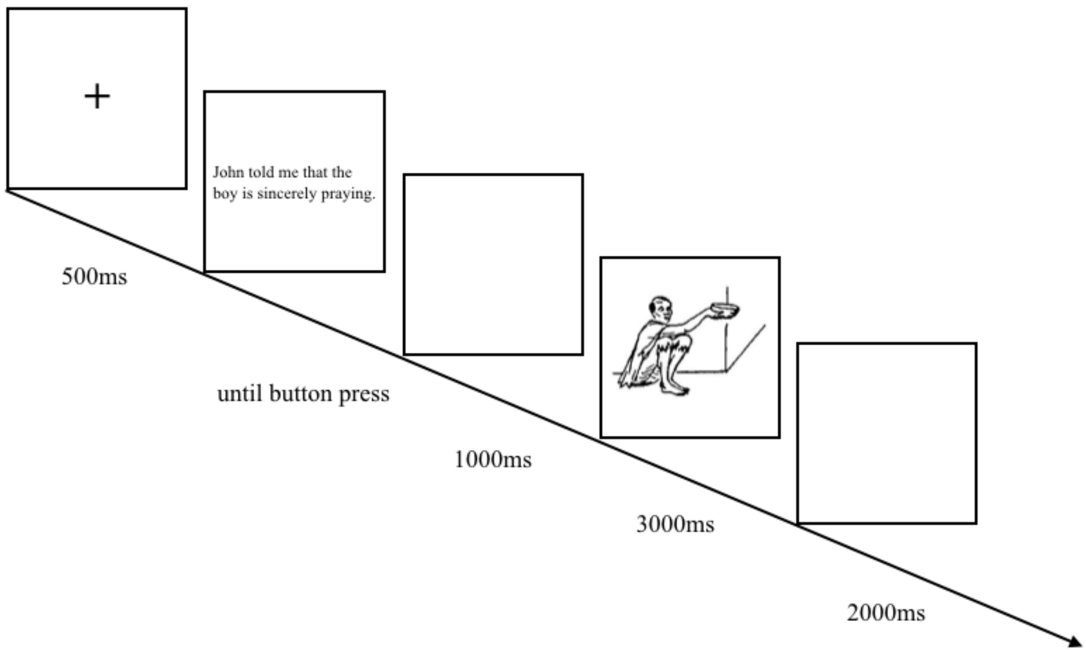



Figure 3-2: A schematic illustration of a sample experimental trial in Experiment 2.

3.4.2 Results

The mean onset latency by condition is shown in Table 3-1. A mixed effects model with a maximal random effects structure in the sense of Barr et al (2013) was first constructed. However, no random slope for interaction terms were included, as they led to convergence failures and the by-subjects random slope for Relatedness was removed via model simplification (Baayen, Davidson & Bates, 2008) due to the high correlation between random effects, a diagnostic of over-parametrization (Bates et al., 2015). In addition to experimental factors, the centralized trial number was included as a predictor, which improved the model fit significantly.

| | Mean latency | | Onset interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |
| Noun | 921 [9] | 915 [9] | 6 |
| Verb | 953 [10] | 916 [9] | 37 |

Table 3-1: Mean speech onset latencies in Experiment 2. Standard error of means (SEM) in square brackets.

The analysis revealed a significant effect of Relatedness ($\beta = 0.01$, SE = 0.003, $|t| = 2.27$, $p < 0.05$) and a marginally significant effect of Sentence Type ($\beta = -0.007$, SE = 0.003, $|t| = 1.95$, $p = 0.059$). The interaction between Verb Type and Relatedness was significant ($\beta = -0.07$, SE = 0.03, $|t| = 2.16$, $p < 0.05$). A following planned comparison revealed that distractor relatedness affected onset latency when the distractor category was a verb ($t1(47) = 4.03$, adj. $p < 0.001$; $t2(23) = 2.44$, adj. $p < 0.05$) but not when the distractor category was a noun ($t(47) = 0.32$; adj. $> p = 0.9$; $t2(23) = 0.48$, adj. $p > 0.9$).

3.4.3 Discussion

Experiment 2 showed that distractor sentences containing conceptually similar distractor verbs (e.g., *John told me that the boy is sincerely praying*) delayed the utterance onset of the target sentence (e.g., *He is begging*). Importantly, this interference effect was not observed when the distractor sentences contained conceptually similar distractor nouns (e.g., *John told me about the boy's sincere praying*). This pattern suggests that conceptually similar lexical items do not interfere with retrieval processes of the target lexical items.

Experiment 2 showed that nouns do not interfere with verbs; however, this only supports half of the prediction based on the syntactic gating mechanism. The syntactic gating mechanism predicts the absence of interference effect in the production of not only verbs but also nouns, among other lexical categories. It is not guaranteed that this category effect should carry over to situations where speakers produce nouns while being distracted by a verb. Thus, it needs to be tested whether the current pattern generalizes to the production of other categories. However, it is unclear how even to design an experiment that is comparable to the current experiment and yet involves the production of nouns. A task in which speakers simply produce an object noun is not suitable, because verbalization of a noun canonically involves substantial changes in lexical meaning. There thus needs to be some task in which speakers produce action nouns, but it is not clear how to have speakers produce a noun given an action picture. Experiment 3 presented below is an attempt to overcome this challenge.

### *3.5 Experiment 3: Syntactic gating in production II*

Experiment 2 showed that nouns do not interfere with verbs, but verbs do interfere with verbs. This contrast is hard to explain via conceptual or morphophonological differences, because the sound and meanings of the distractors were (near-)identical in the category-match and the category-mismatch conditions. It is a further empirical question whether verbs do not interfere with nouns. Experiment 3 tested this question.

3.5.1 Methods

*Participants*

24 native speakers of English participated for either class credit or monetary compensation. None had participated in Experiment 1.

*Materials and Design*

The same set of pictures was used as in Experiment 1. In this experiment, however, a colored square was placed in the lower right corner of each picture, The size of the square varied randomly across pictures, but not across the same picture in different conditions. Example pictures are shown in Figure 3-3 below. This colored square was used to elicit sentences in which the depicted action is expressed using a nominal, as in *his begging is yellow* and *her juggling is blue* (see Procedure for how these utterances were elicited).

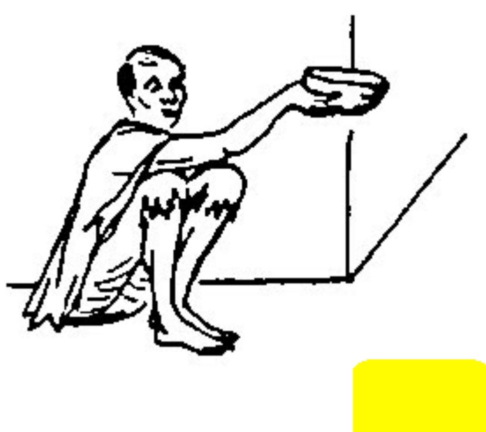


Figure 3-3: Sample picture stimuli used in Experiment 2.

*Procedures*

Procedures were the same as Experiment 1. The only difference was the instruction for the sentence description part of the task. The instruction was something like the following: *Imagine yourself in a hypothetical world where you perceive a color for each action. You know that some people, specifically people who have what's called synesthesia, perceive colors for things like numbers and letters. In your case, you perceive a color for each action. Depending on the kind of action and depending on who does it, you perceive different colors. Your task is to report the color of each action, using a full sentence of the form X's Ving is red/blue/etc…*

3.5.2 Results

The mean onset latency by condition is shown in Table 3-2. The same mixed effects model with a maximal random effects structure was constructed, but all the random slopes except the by-item random slope for Relatedness were removed via model simplification due to the high correlation (> 0.8) between the random effects. In addition to the experimental factors, the centralized trial number was included as one of the predictors.

| | Mean latency | | Onset interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |

| | | | |
|---|---|---|---|
| Noun | 1046 [12] | 1011 [13] | 35 |
| Verb | 1011 [12] | 1011 [12] | 0 |

Table 3-2: Mean speech onset latencies in Experiment 3 (noun description). Standard error of means in square brackets.

There was no significant effect of Relatedness or Sentence Type ($p > 0.15$). The interaction between Verb Type and Relatedness was significant ($\beta = 0.01$, $SE = 0.005$, $|t| = 2.16$, $p < 0.05$). Following planned comparisons revealed that distractor relatedness affected onset latency when the distractor category was a noun ($t1(23) = 2.89$, adj. $p < 0.01$; $t2(23) = 2.17$, adj. $p = 0.08$)[11] but the effect of relatedness was clearly not significant when the distractor category was a verb ($t(23) = -0.93$; adj. $p > = 0.6$; $t2(23) = -0.19$, adj. $p > 0.9$).

3.4.3 Combined analysis

An analysis combining the data from Experiments 2 and 3 was conducted. For this analysis, a categorical factor Experiment (2 vs. 3) was added to the model. A mixed effects model with a maximal random effects structure was constructed, but the by-subjects random slope of Relatedness, as well as the by-item random slopes for interaction terms and the by-item random slopes for Experiment were removed via model simplification due to the high correlation (> 0.8) between the random effects. The resulting model revealed a main effect of Relatedness ($\beta = 0.008$, $SE = 0.004$, $|t| = 2.14$, p

[11] Note that this marginal significance is due to the Bonferroni's correction, which can be overly conservative for a planned comparison. For instance, see (Keppel,1991)

$< 0.05$), a main effect of Experiment ($\beta = -0.05$, SE = 0.019, $|t| = 2.77$, $p < 0.01$), an interaction between Distractor Category and Experiment ($\beta = -0.006$, SE = 0.003, $|t| = 2.32$, $p < 0.05$), and a three-way interaction between Relatedness, Distractor Category, and Experiment ($\beta = -0.009$, SE = 0.003, $|t| = 3.01$, $p < 0.01$)

### *3.6 Discussion: Experiments 2 and 3*

The results of Experiments 2 and 3 are summarized in Figure 3-4 below. In both experiments the interference effects from conceptually similar distractors disappeared when the distractor did not match the target in syntactic category. This was so even though the conceptual similarity between distractor words and the target words was kept maximally constant across the different syntactic category conditions. Under the assumption that the semantic interference effect reflects lexical competition (e.g., Schriefers et al., 1990; but see Mahon et al. 2007), this pattern suggests that syntactic categories are a strong determinant of when and what lexical items should be considered as candidates for retrieval, even when they are divorced from default ontological categories.

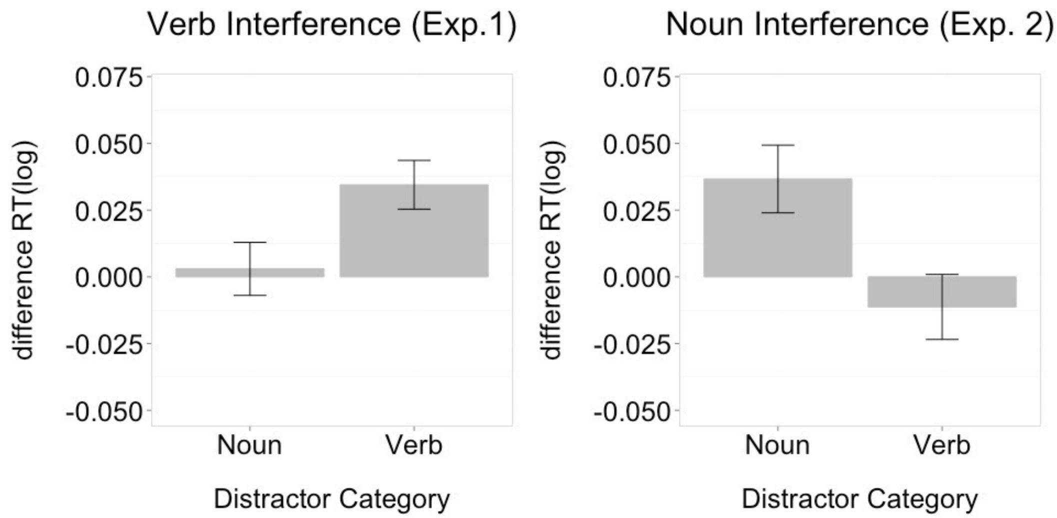



Figure 3-4: Difference RTs (log) by distractor category in Experiment 2 (left) and Experiment 3 (right).

*Relation to previous studies*

As briefly discussed above, there are some relevant studies in the PWI literature. First, Pechmann & Zerbst (2002) showed that distractors from the same syntactic category create greater interference effects relative to those from a different syntactic category. Vigliocco et al. (2005) replicated this with a different set of materials in Italian, while also showing that this effect is additive to the effect of semantic similarity. Vigliocco and colleagues also showed that the category effect obtains when speakers are producing inflected verbs, and not when producing the uninflected citation form of the verb. Both of these experiments obtained a rather different pattern of results than the current experiments. Specifically, they showed a main effect of syntactic category that independently contributed to the net interference effect. This independent contribution of syntactic category effect on interference was absent in the current results.

However, those studies have several problems that make them hard to interpret. First, Janssen et al. (2010) demonstrated that Pechmann & Zerbst's study was confounded with imagability that covaries with syntactic category. Second, Vigliocco et al.'s experiment used phrasal distractors in order to reliably manipulate the syntactic category of the distractors. As discussed above, the use of phrasal distractors is problematic given the automaticity assumption underlying the PWI task. Additionally, Vigliocco et al.'s study had mismatching conjugations (of Italian 3rd person singular) between the distractor verbs and target verbs in the inflected production conditions. Italian has three types of verbs, which show different phonological realizations of 3rd person singular morphology. When the target verb and the distractor verb differ in terms of how the 3rd person singular feature is expressed phonologically, an interference due to mismatching conjugation could arise. This might contribute to the observed effect of syntactic category. Additionally, when the target verb and the distractor verb share the same phonological realization of the 3rd person singular feature, a facilitative effect due to matching conjugation could result. Such interference and/or facilitation due to conjugation (mis-)match could reasonably affect speech onset time. This can explain why Vigliocco et al. found a main effect of category mismatch only in the inflected verb production task, because mismatching conjugations between the distractor and the target is relevant only when the target verb needs to be inflected. Indeed, when the stimuli used in Vigliocco et al. (2005) are examined, the number of matching vs. mismatching conjugation distractors were disproportionate between the semantically similar vs. semantically dissimilar distractor conditions. Thus, the pattern obtained in Vigliocco et al (2005) might be due not to syntactic category, but to the confound from the conjugation match/mismatch between the target and the distractor verbs. Thus, the previous studies,do not provide definitive evidence that within-category distractor words create interference independently from semantic similarity.

Turning to the mechanism underlying the current pattern, there is an existing proposal that is similar in spirit to the syntactic gating mechanism I proposed above. Dell et al. (2008) argued that syntactic category restricts the competition between words, explaining why substitution and exchange errors rarely occur between words that do not share the same category. They called this constraint the 'syntactic traffic cop,' and implemented it using a connectionist network. In a nutshell, they proposed that, in the lexical network there are nodes that represent syntactic categories. Those nodes are activated in an appropriate syntactic environment, and they possesses excitatory connections to every lexical node of that category. This enhances the activation of the items with the relevant syntactic category, creating extra competition between within-category words but not between across-category words. They argued that this model can capture Pechmann & Zerbst's and Vigliocco et al.'s results.

However, Dell and colleagues' model does not explain the current findings. First, the current experiment did not obtain a main effect of the syntactic category manipulation. Dell and colleagues' model predicts a main effect of syntactic category manipulation, due to the relatively highly weighted excitation received from the category nodes. In addition, their model should predict that conceptually similar between-category words should still compete to some extent.[12] In contrast, the current results showed no evidence of competition between semantically related words from different categories.

Of course, Dell et al.'s model may be modified to capture the current effect. For instance, adding inhibitory connections between category nodes and lexical items from the non-matching category might help to capture the current results. However, there is a serious problem that must be overcome when one attempts to explain this data in a spreading activation framework (Dell, 1986). The problem is that the nouns and verbs used in the current experiments carry the exact same phonology and near identical semantic content, and that the noun version of each target word is derived form the verb in a completely productive rule of deverbalization. This is problematic because the productive rule of deverbalization requires the network to (i) duplicate the lexical node for every verb with a unique set of connection to the syntactic category nodes, and (ii) learn the appropriate weights of the connection between the duplicated lexical nodes with the appropriate category nodes reliably based on few occurrences, without relying on the phonological or semantic content of words. Thus, the current results pose challenges to the (localist) spreading activation lexical network model, both in terms of the number of nodes required to represent lexicon, and in terms of the learning problem due to the relative low frequency of the deverbal form of some verbs, which share identical phonology and near-identical semantic content to the associated verb.

---

[12] This prediction actually depends on a specific activation function parameter.

Finally, it is worth mentioning that the progressive form of verbs and deverbal nouns used in this experiment is different in meaning, phonological paradigm, or both. Regarding the meaning difference, they have for instance differential possibilities for later being referred to by a nominal pronoun.[13] Deverbal nouns can be later referred to by the pronoun *it*, but progressive verbs usually cannot be. However, the relevant question here is whether such a subtle difference can completely eliminate the otherwise robust conceptual similarity interference effect. The core meaning of verbs and deverbal nouns are still shared to the large extent, and it would be surprising if such a subtle difference can explain the current pattern. Also, verbs and nouns have different morphophonological paradigms, in the sense that they have different sets of possible inflections (e.g., verbs can be inflected with past tense *-ed*, nouns cannnot be inflected by it). Thus, morphophonological paradigm covaries with syntactic category. However, again, it is not clear how this can explain the current results. In addition, separating morphophonological paradigm from syntactic category is *in principle* impossible, at least in most if not all of the natural languages. Given that the alternative explanation of the current pattern based on the difference in morphophonological paradigms is unavailable, I argue that the current contrast in the conceptual interference effect between nouns and verbs is due to the difference in syntactic category.

*Implications to lexical access model*

---

[13] I would like to thank Alexander Williams for pointing this out to me.

One important aspect of the current results is that the semantically similar distractors showed no sign of interference compared to semantically dissimilar distractors when they belonged to a different syntactic category then the target word. This pattern is unexpected under the view that the noun and verb version of the target words are derived from category-less roots in online processing (e.g., Barner & Bale, 2002). Under such an underspecification view, there should exist a retrieval process for category-less roots. This root retrieval process should be subject to similarity-based interference. That is, in any processing theory in category-less root is the target of retrieval process, some amount of interference due to conceptually similar distractors is expected. Unless one assumes that this root retrieval operation is somehow exempt from conceptual similarity-based interference, the complete elimination of the interference effect in the category-mismatching conditions in the current experiments is unexpected.

There are some ways to accommodate the current pattern in models that assume no category features in lexical representation. First, it is possible to argue that the interference effect observed in the current experiment reflects post-retrieval integrative processes. In sentence production, a lexical item presumably needs to be integrated to a sentential structure before it can be uttered. It is possible that this structural integration process is subject to similarity-based interference, and that the interference effect in the current experiment reflects it. This explanation, however, also has to assume that the interference effect observed in the current experiment is fundamentally different from the semantic interference effect reported in many previous PWI experiments (e.g., Lapker, 1979; Schriefers et al., 1990). In most PWI experiments, speakers do not produce a sentence. Hence, the integration account of interference effects is not available. Given the similarity between the nature of the manipulation in the current SPI experiment and previous PWI experiments, I take the integration account of SPI interference effect to be unlikely.

Second, it is also possible to assume that a category-less root contains no semantic information that can be interfered with by conceptually similar distractors. This is indeed akin to an assumption made in what is called the Distributed Morphology framework (Halle & Marantz, 1993), the theoretical basis of the processing models that assume no category features in lexical representations. In the Distributed Morphology framework, the meaning of a word is assigned after the root is combined with a categorizer, a functional lexical item that defines the syntactic category of the resulting item (see Harley & Noyer, 1998). If this assumption transfers to processing as well, conceptually-based interference might occur only after a root combines with a categorizer because roots do not carry meaning by themselves. This assumption, however, is rather non-standard in processing models. I am aware of no model of lexical memory in which a lexical item carries no meaning until after some combinatorial operations are performed. Thus, I take this possibility to be rather premature to accept in processing theory.

As both of the two approaches under the processing models that assume no category feature in lexical representation are not easily acceptable, I adopt the position that the category feature is an integral part of the lexical representations that the content-addressable retrieval operation can target. This position is a necessary component of the proposal that I described in the section 3.1, because the content-addressable retrieval operation cannot feasibly use a category feature as a retrieval cue if the category feature is not represented as a content of lexical representation. In effect, the current proposal does not assume that the derivational assumptions of Distributed Morphology is a straightforward reflection of processing models.

Finally, it must be acknowledged that the current tasks may allow speakers to use a structural template in producing target sentences. Speakers in the current experiments can know what kind of structures need to be used before identifying the content of the presented picture because the structure of the target sentence does not change in the current experiment. This might allow speakers to reuse the structural template without actually computing the structure of the utterance for each trial. If this is the case, then the structural representation of course exists before lexical processes can begin. The current task thus might be biased to show evidence of the temporal precedence of structural processes relative to lexical processes. Despite this caveat, however, it is reasonable to claim that, syntactic category strongly constrains the lexical processes at least when structural representation is present before lexical processes. This syntactic constraint eliminates the interference effect when the non-target distractor items do not match in syntactic category features with the target. Since Experiment 1 in Chapter 2 independently showed that structural processes precedes the lexical processes in a less formulaic speech task, it can be argued that, in speaking generally, the syntactic category feature limits the space of potential retrieval candidates for content-addressable memory access. This is the sense in which I take syntactic category to be a privileged retrieval cue in lexical access, and this privileged nature of syntactic category in lexical access is the central claim of the syntactic gating mechanism that I proposed above in Section 3.1.

So far, the arguments have been largely restricted to sentence production. However, as I hinted in Section 3.1, the mechanism that characterizes the interaction between structural and lexical processes should also exist in sentence comprehension, under the single system hypothesis. I will discuss the relevance of the syntactic gating mechanism in the following section, with a specific focus on the predictive part of comprehension. The reason that I focus on predictive comprehension is because it involves a process that may be characterized as analogous to sentence production. Predictive comprehension involves retrieval of lexical items based on message-level representations extracted from the contextual input. In this sense, it is a part of comprehension that might plausibly exhibit similar effects of the syntactic gating mechanism if it exists in sentence comprehension. However, before I discuss the potential involvement of the syntactic gating mechanism in predictive comprehension, it is necessary to characterize the mechanism underlying predictive comprehension itself. Section 3.6 serves this purpose. It presents a case where comprehenders arrive at a correct prediction surprisingly slowly, in comparison to previously reported evidence. I will discuss the implications of this slow-prediction on the mechanism underlying prediction in Section 3.6. Finally, Section 3.7 discusses the effect of syntactic category on prediction. It presents 'fast' cases of prediction, as compared to the evidence presented in Section 3.6. This speed contrast needs explanation, and the explanation that I offer in Section 3.7 implicates the syntactic gating mechanism.

### *3.6 Predictive comprehension*[14]

The world I live in is to a certain extent regular. I exploit this regularity by using our memory to predict future events, in order to efficiently process and respond to incoming sensory input. This predictive nature of cognition has been suggested as a fundamental principle of human brain function (e.g., Bar, 2007). In the domain of language, comprehenders actively predict various properties of upcoming words in a sentence, including their semantic (Altmann & Kamide, 1999), grammatical (van Berkum et al., 2005), phonological (DeLong et al., 2005) and possibly even orthographic (Dikker, Rabagliati & Pylkkänen, 2009) properties, from various types of information such as thematic relations (Altmann & Kamide, 1999; Kamide, Altmann & Haywood, 2003), discourse pragmatics (Nieuwland & van Berkum, 2006) and even speaker identity (van Berkum et al. 2008). Prediction is taken to be one of the primary reasons why comprehension is fast and robust despite degradation in the input. However, despite previous successes in demonstrating prediction in comprehension, it has been proven difficult to infer the underlying cognitive processes because of its impressive speed. The current research thus aims to reveal the processes by which comprehenders generate predictions, by examining a case where prediction is measurably slow. Such a slow case of prediction gives us an opportunity to 'look inside' the computations involved in prediction during online comprehension.

[14] Section 3-6 is a lightly edited version of the manuscript for a journal article (Momma, Sakai, Lau & Phillips, in prep). For this reason, the analyses conducted on Experiment 4 and Experiment 5 are different from each other.

A measure that is commonly used to examine prediction is Event Related Potentials (ERPs), stimulus-locked brain potentials that are extracted from electroencephalographic (EEG) recordings. This is not the only measure that is successful in demonstrating prediction. For instance, other particularly useful measures include visual-world eye-tracking (Altmann & Kamide, 1999; Tanenhaus & Spivey-Knowlton, 1996). However, I use ERPs here because I are interested in a potentially slow effect. Slow effects might not appear to be slow in VWET because the visual stimuli typically used in visual world eye-tracking paradigm dramatically simplify the prediction problem that comprehenders face. I use a well-studied language-related ERP component called the N400. N400 is a negative-going brain potential that peaks at around 400ms after a word's onset. This component has sometimes been thought to reflect integrative semantic processes (Brown & Hagoort,1993; Hagoort, 2004), but here I adopt the position that it reflects how predicted the word is given the current context. I have several reasons for adopting this position (for an overview see Federmeier et al., 2012 and Lau et al., 2008), but the most important reason for current purposes is that the amplitude of the N400 shows an inverse correlation with the predictability of a word. The less predictable the word, the larger the N400 amplitude,, independently of whether the word is semantically anomalous (e.g., Kutas & Hillyard, 1984).

There exist, however, systematic cases that violate this overwhelming generalization. For instance, a*rgument role reversal sentences (*henceforth simply role reversal sentences*)* systematically fail to elicit a difference in N400 amplitude relative to a non-reversed, canonical sentence baseline, at least in most languages most of the time.

Such sentences instead elicit an ERP effect called P600 effect, which is typically associated with syntactic processing (Osterhout & Holcomb, 1992). In a seminal study by Kim & Osterhout (2005), the verbs in sentences like *the hearty meal was devouring*… and in its unreversed plausible version like *the hearty meal was devoured* elicited no reliable N400 but a reliable P600 difference has been found between. In these two sentences the verb differs greatly in predictability, but the lack of an N400 effect suggests that the verb was equally expected in either case. This lack of N400 effect and the presence of a P600 effect in the role-reversal manipulation has been replicated multiple times across different laboratories and in different languages (Spanish: Stroud & Phillips, 2012; Japanese: Oishi & Sakamoto 2010; Chinese, Chow & Phillips, 2013; Dutch, Kolk et al., 2003; van Herten et al., 2005; for a related phenomenon in a visual-world eye-tracking study, see Kukona et al., 2011).

On the view that N400 reflects prediction, role reversal sentences provide a valuable tool for studying prediction processes. This is because they create contexts that make the same continuation either very likely or very unlikely, while using exactly the same set of words. This allows us to test the distinct contribution of the words and their roles in the sentence. Using this property, a recent study in Mandarin Chinese (Chow et al., submitted) revealed that the N400's insensitivity to the role-reversal manipulation can be overcome by increasing the distance (number of words intervening) between argument nouns and verbs. When a temporal adverbial like *last week* intervenes between the pre-verbal nouns and the verb in Chinese sentences like *The cop the thief last week arrested* and a role-reversed version like *The thief the cop last week arrested*, the N400 is larger for the role-reversed sentences. In contrast, when the temporal adverbial is fronted, with no words intervening between the second noun (*cop)* and the verb (*arrested)*, no difference in N400 was observed, just like in previous role reversal studies in various languages. Based on this pattern, Chow and colleagues argued that argument roles are slow to affect verb prediction.

This observation begs the question of why argument roles are slow to affect verb prediction. This section is devoted to answering this question. As an initial step I identify at least two broad classes of possible accounts. Roughly, the two accounts differ in whether the delayed effect of role reversal is attributed to the bottom-up process of building context representations or to the top-down process of generating predictions. The first account attributes the delay to the process of building context representation. If the context information is not correctly represented, an appropriate verb prediction is of course not possible. This would mean that the slow verb prediction in role reversal sentences is not due to some constraint of the top-down prediction mechanism *per se*, but simply due to the slowness of the bottom-up processing of the context. If this is the case, role reversal sentences are not too informative about the top-down mechanism that give rise to predictions. In contrast, the second account attributes the delayed effect to the top down process of predicting a verb using the contextual input. Under this account, argument roles are assigned correctly and rapidly, but the prediction mechanism cannot immediately use it for some reasons. The first aim of this study is thus to distinguish between these two accounts.

A strong test for these accounts arises when the contextual input is maximally transparent in terms of argument roles and when the complexity of the contextual input is reduced to a bare minimum, so that any delayed effect cannot be attributed to a delay in bottom-up processing of the input. To create such a test, Japanese is an ideal language because it allows us to create maximally simple role reversal sentences with explicit morphological markings of the arguments' grammatical functions, and hence by strong correlation their argument roles. Thus, in the current study, I compare Japanese speakers' brain response to canonical sentences like *bee-NOM(inative) sting* or *fly-ACC(usative) swat* and their role reversed variants like *bee-ACC sting/fly-NOM swat*. In these sentences, *sting*/*swat* is likely after *bee-NOM/fly-ACC*, but it is unlikely after *bee-ACC/ fly-NOM*. This maximally simple context with explicit case marking is useful in assessing whether the delayed verb prediction is due to a delay in bottom-up context analysis or due to slow top-down predictive processes. Furthermore, the pure effect of time on verb prediction can be assessed by simply manipulating the time interval between the onset of the first word and the onset of the second word (Stimulus Onset Asynchrony (SOA)).

### 3.6.1 Methods

*Participants*

23 participants from Hiroshima University (3 replaced; Male = 12; Female =11) participated in the experiment for monetary compensation. All participants were right-handed and were native speakers of Japanese with no significant early (before age of 10)

exposure to any other languages. The mean age was 20.73 (range = 18-26). Prior written consent was obtained for each participant.

*Materials*

80 noun-verb pairs that have a typical agent-predicate relation (agent pairs, e.g., *bee-sting*) and another 80 noun-verb pair that have a typical patient-predicate relation (patient pairs, e.g., *fly-swat*) were constructed in Japanese. These word pairs were converted into plausible sentences of Japanese by inflecting the noun with the appropriate case marker, either nominative (*-ga*), which generally marks the subject (and hence agent in the current cases) of a sentence or accusative (*-o*), which generally marks the object (and hence patient) of a sentence. Here I exploited three properties of Japanese. First, Japanese has explicit case morphology, so the grammatical function (which has a strong correlation with argument roles) is encoded not by word order but by simple inflectional marking. Second, Japanese allows liberal argument dropping, so not mentioning one of the arguments does not render the sentence ungrammatical. Third, Japanese is a strongly verb-final language, so the verb naturally follows all nouns. This property makes the verb-final construction a default, unmarked structure unlike in the many of other role-reversal studies. These three properties conspire to make two-word noun-verb sentences grammatical and easily role reversible without involving two arguments or changing the form of the verb. In addition to this manipulation, the Stimulus Onset Asynchrony (SOA) between the noun and the verb was manipulated. The short SOA condition had an 800ms SOA, and the long SOA condition had a 1200ms SOA. These two manipulations were fully crossed to yield the following four conditions: Canonical-short SOA; Canonical-

long SOA; Reversed-short SOA and Reversed-long SOA. A total of four stimulus presentation lists were created according to the Latin-Square method, such that each participant saw exactly one version of the sentence created from a particular word pair template. Each list contained four blocks of trials, each containing an equal number of trials per conditions in a pseudo-random order. The order of these four blocks was randomized for each participant so as to avoid list-specific effects.

*Procedure*

Each participant read a total of 320 two-word sentences (160 experimental, 160 fillers) while their brain activity was measured via EEG. Psychopy (Pierce, 2007) was used for stimulus presentation. At the beginning of each trial, a fixation cross was presented at the center of the screen for 500 ms. Following a 400ms blank screen a simple two-word sentence was presented in a word-by-word manner. Each word was presented for 400ms and separated by either a 400ms or an 800ms blank screen (experimental variable). 1000ms after the offset of the second word, a screen that read 'Plausible?' appeared. Participants were instructed to judge the plausibility of the sentence they just read after seeing this screen.

*Electrophysiological Recording*

Twenty-nine tin electrodes were placed on the scalp by an elastic cap. Electrodes were also placed below and above the right eye to monitor the ocular movements and blinking. Impedance was kept below 5kOhms for all scalp electrode sites whenever possible.

*Data Analysis*

EEGLAB (Delorme & Makeig, 2004) and ERPLAB toolboxes (erpinfo.org/erplab) were used for analysis of the EEG data. EEG data were filtered offline using a 4th order Butterworth filter with a cutoff between 0.1Hz-30Hz. Subsequently, EEG data were epoched and baseline corrected with a 200ms pre-stimulus baseline. Artifact rejection was then performed using a combination of 1) a simple voltage threshold algorithm with a voltage threshold of +- 100 microvolts over all electrodes, 2) a moving window algorithm with a voltage threshold of 75 microvolts, a window size of 200ms, and a window step of 50ms in the vertical Electrooculogram. This affected 15.1% of experimental trials, roughly equally distributed across conditions (Reversed-Short: 15.5%; Reversed-Long: 16%; Canonical-Short: 15%; Canonical-Long: 15.5%). Participants with fewer than 20 trials (< 50%) remaining in any of the conditions were excluded from the analysis. 3 participants were replaced due to this exclusion criterion.

Event-related potentials were computed separately for each participant and each condition for the 800ms after the onset of the critical verb. Statistical analyses on average voltage amplitudes were conducted separately for two time windows chosen based on previous literature (Kutas & Hillyard, 1984): 300-500ms for the N400 and 500-800ms for the P600. I used R (R Development Core Team, 2010) for all statistical analyses reported here.

For the analysis of critical verbs, I computed a repeated measures Type III ANOVA on the mean amplitudes of all electrode sites in each time-window. I then conducted a quadrant analysis following Lau et al. (2012) to examine the topographic distribution of the effects of interest[15].

I further analyzed the responses to pre-critical noun phrases. This was done mainly to assess the effect of case-marking, i.e., nominative vs. accusative. For this analysis I did not exclude one of the participants that I excluded for the above analysis. This is because I had sufficient trials in both the nominative and accusative conditions (80 trials), and thus each condition had enough surviving trials. However, excluding this participant did not change the statistical patterns reported here. Note that this comparison is independent from the effect of plausibility on the critical verbs. This is because both plausible and implausible conditions contained equal numbers of nominative and accusative marked noun phrases. The same time windows (300ms-500ms, 500ms-800ms) were analyzed.

### 3.6.2 Results

*Behavioral Results*

Participants reliably judged the canonical sentences to be plausible, and the role reversal sentences to be implausible, with overall accuracy of 91.5%. The response accuracy for each condition is the following: Plausible-Short: 95.4%; Implausible-Short:

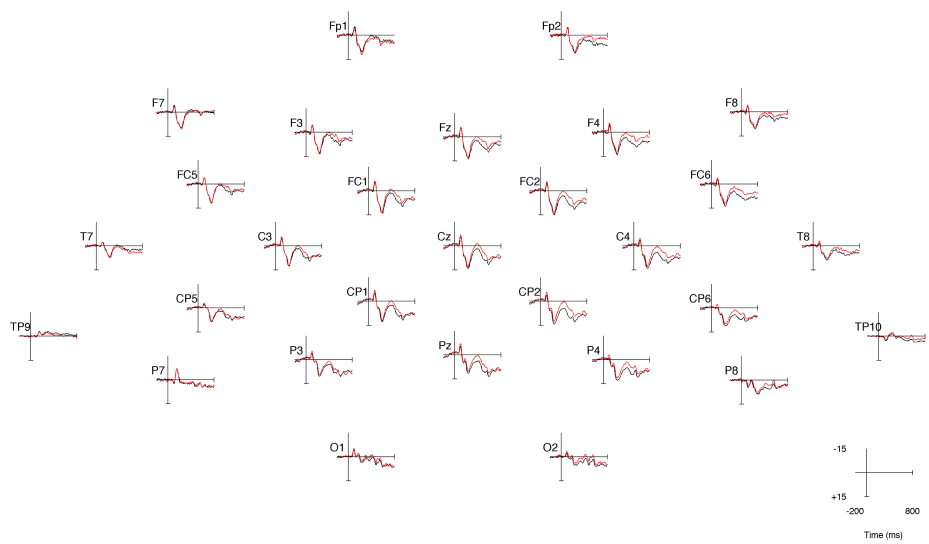
---

[15] Note that this analysis differs from Experiment 5. This is because the current experiment is written for a journal manuscript, and because I wanted to compare the current effect with Lau et al. (2012), in which the quadrant analysis revealed that the predictive effects on N400 was most prominent in the RP quadrant.

88.1%; Plausible-Long: 95.0%; Implausible-Long: 86.5%. For filler trials, the overall accuracy was 94.6%.

*ERP Results*

Critical Verb: The ERP responses to the critical verb for each SOA condition are shown in Figure 3-5 below.



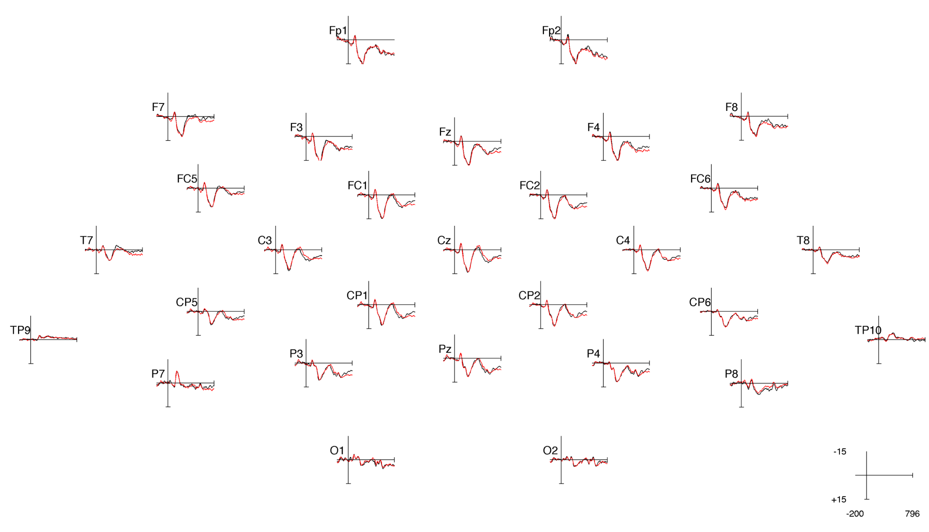



Figure 3-5: Grand average ERP to verbs in the critical sentences in the long SOA conditions (top) and in the short SOA conditions (bottom).

In the 300-500ms time-window after the critical verb onset, the repeated measures ANOVA analysis across all electrode sites revealed a main effect of plausibility ($F(1,22) = 5.69$, $p = 0.03$). The effect of duration was not significant ($F < 1$). The effect of plausibility interacted with duration ($F(1,22) = 4.92$, $p < 0.05$). A planned comparison revealed that the effect of plausibility was significant in the long SOA condition ($t(22) = 2.70$, $p = 0.01$), but not in the short SOA condition ($t(22) < 1$).

A subsequent quadrant analysis that was conducted to examine the topographic distribution of the effects revealed that the interaction was significant at Right Anterior (RA) quadrants ($F(1,22) = 6.31$, $p < 0.05$) and Right Posterior (RP) quadrants ($F(1,22) = 4.32$, $p < 0.05$), but not at Left Anterior (LA) or Left Posterior quadrants.

In the 500-800ms time window, the overall ANOVA analyses revealed no significant main effect or interaction involving the plausibility factor, though the interaction between plausibility and duration was marginally significant at the right anterior quadrant ($F(1,22) = 3.82$, $p = 0.06$). Figures 3-10ab show the ERPs to the critical verbs at representative midline electrode Pz. The topographic map of the difference amplitude in two time-windows of interest is plotted in Figure 3-11.

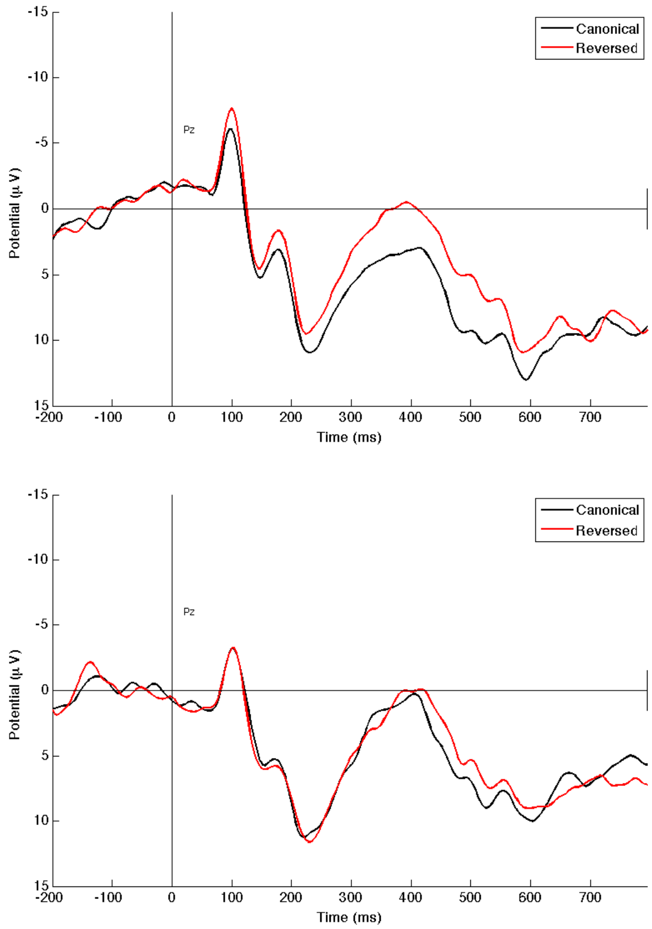



Figure 3-6ab: ERP waveform for the verbs in critical sentences at midline electrode Pz comparing canonical (black) and reversed (red) orders in the long (a) and short (b) SOA conditions.

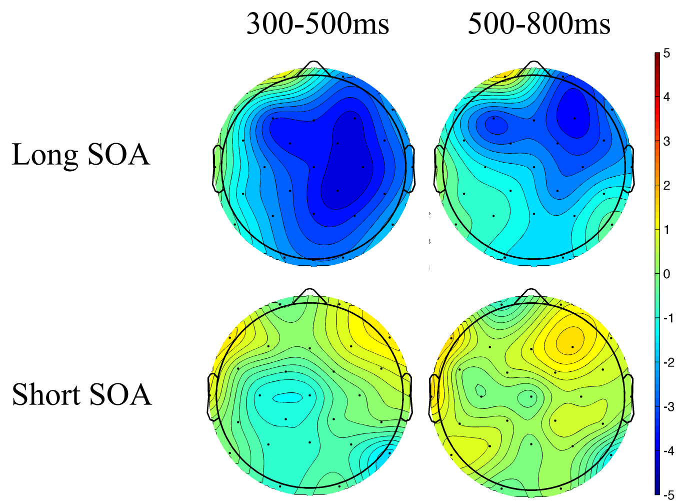



Figure 3-7: Topographic map of the ERP amplitude difference (reversed - canonical) in each time-window at the critical verb.

*ERP Results (filler sentences)*

As a point of comparison, I also analyzed ERPs to semantically anomalous two-word sentences that were used as fillers. These sentences were created by recombining the words from plausible noun-verb sentences. For instance, plausible noun-verb pairs like *boxer-NOM fight*/*ice-NOM melt* were recombined to yield the anomalous *boxer-NOM melt*/*ice-NOM fight*. The brain responses to the plausible and recombined versions were compared.

In the 300-500ms time window, an overall ANOVA revealed a main effect of plausibility ($F(1,22) = 35.16$, $p < 0.01$). The interaction between plausibility and duration was not significant ($ps > 0.14$). A planned comparison revealed that the effect of plausibility was significant in the long SOA condition ($t(22) = 5.90$; $p < 0.01$) and in the short SOA condition ($t(22) = 3.29$; $p < 0.01$).

A subsequent quadrant analysis revealed that the main effect of plausibility was significant at all quadrants (LA: $F(1,22) = 7.89$, $p = 0.01$; LP: $F(1,22) = 23.02$, $p < 0.01$; RA: $F(1,22) = 19.37$, $p < 0.01$; RP: $F(1,22) = 47.64$, $p < 0.01$). The main effect of duration was significant only at LA ($F(1,22) = 8.81$), $p < 0.01$) and RA ($F(1,22) = 4.98$, $p < 0.05$). No quadrant showed an interaction between plausibility and duration (all ps > 0.1)

In the 500-800ms time window, an overall ANOVA revealed a main effect of plausibility ($F(1,22) = 4.65$, $p < 0.05$). The interaction between plausibility and duration was not significant (ps > 0.7). The main effect of plausibility was significant at quadrants RA ($F(1,22) = 9.04$, $p < 0.01$) and RP ($F(1,22) = 7.34$, $p = 0.01$), but not at quadrants LA or LP (Fs < 2).

Figure 3-8ab shows the grand average ERP to the critical verb at the representative midline electrode Pz at each SOA. A topographic map of the amplitude difference in the two time-windows is shown in Figure 3-9.

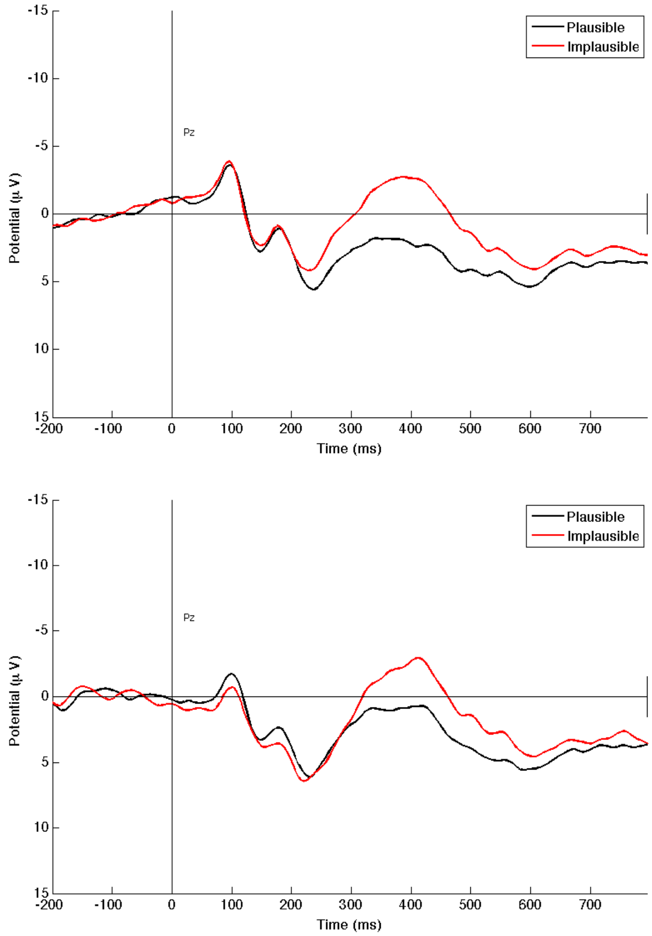



Figure 3-8ab: Grand average ERP at the verb in filler sentences at midline electrode Pz comparing plausible (black) and implausible (red) in the long (top) and short (bottom) SOA conditions.

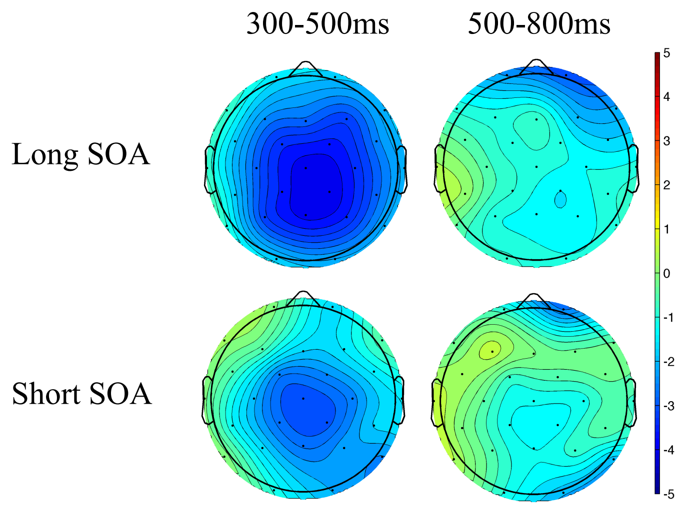



Figure 3-9: Topographic map of the amplitude difference (implausible - plausible) in each time window at the verb in filler sentences.

*ERP Results: Pre-critical nouns*

In the 300-500ms time window, the main effect of case was marginally significant ($F(1,22) = 3.09$, $p = 0.09$) in overall electrode analysis.

A subsequent quadrant analysis to test the topographic distribution of the effect revealed that the main effect of plausibility was significant at quadrants LP ($F(1,22) = 5.87$, $p < 0.05$) and RP ($F(1,22) = 7.02$, $p = 0.01$), but not at quadrants LA and RA ($Fs < 1$).

In the 500-800ms time window, neither the overall ANOVA nor analyses in individual quadrants revealed an effect of case marker, though there was a marginally significant difference at the RP quadrant ($F(1,22) = 3.56$, $p = 0.07$).

Figure 3-14 shows the grand average ERP to the critical verbs at the representative midline electrode Pz at each SOA. A topographic map of the amplitude difference in the two time-windows of interest is shown in Figure 3-15.

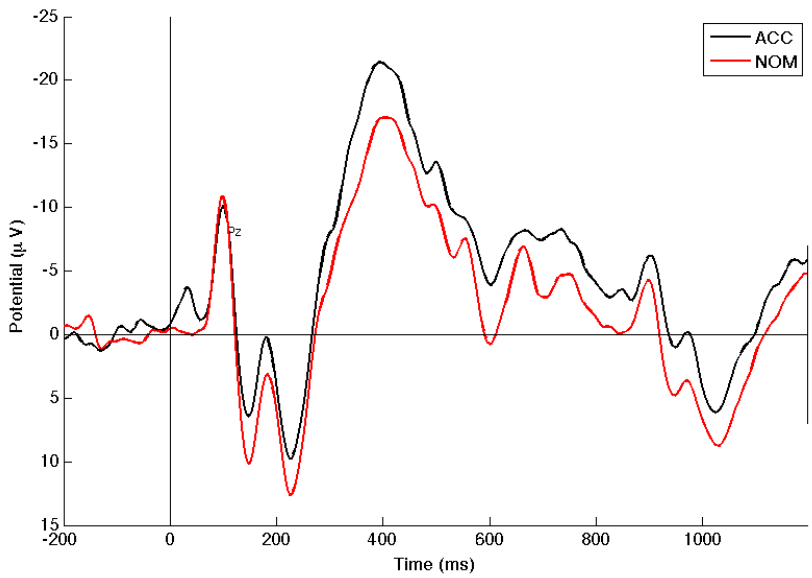



Figure 3-10: Grand-average ERP at the noun phrase in midline  sentences at electrode Pz comparing accusative (black) and nominative conditions.

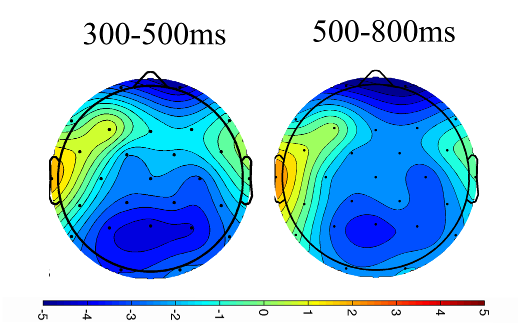



Figure 3-11: Topographic map of the amplitude difference (nominative - accusative) in each time-window at the sentence initial noun.

### 3.6.3 Discussion

I tested whether the N400's insensitivity to role reversal sentences reflects a delay in bottom-up context analysis or top-down prediction processes. I did so by using Japanese sentences whose context is minimally complex and maximally transparent in terms of argument roles. The results are that the N400 did not differentiate the verbs in the canonical and role reversed sentences in the short SOA (=800ms) conditions. This suggests that argument role information is indeed initially ineffective for modulating the N400, even when the context contains only one word whose grammatical function is morphologically encoded. Furthermore, I tested whether the N400's insensitivity can be overcome by simply adding time. In contrast to the short SOA conditions, there was a clear N400 difference between the verbs in the canonical and the role-reversed sentences in the long SOA (=1200ms) conditions, suggesting that adding pure time refines the prediction based on argument roles. No such timing contrast was found in filler sentences that lacked a role reversal manipulation (see also Anderson & Holcomb, 1995 for additional evidence that N400 is usually not affected by SOA manipulation). I thus argue that the delayed effect of argument roles can be attributed to top-down predictive processes.

Before I discuss the potential mechanisms underlying prediction, it is worth revisiting the question of whether the bottom-up interpretation is still possible in light of the current results. Namely, is it still possible that the N400 insensitivity in the short SOA condition reflects a failure to promptly analyze the context, rather than a failure to predict verbs? I give three reasons why I think this is not plausible given the current results. First, if the N400 insensitivity is due to a failure to encode the context, one has to also argue that the identification of the correct argument role of a single word takes more than 800ms, even when no additional stimuli are presented during that time. This is implausible given the highly incremental nature of sentence comprehension, especially in a language whose speakers have life-long experience of identifying argument roles without verbs, as Japanese speakers do. Second, I observed in the current study that the N400-like effect differentiated nominative- and accusative-marked nouns in the pre-critical region. This effect was significant in the 300ms-500ms interval but not in the 500-800ms interval, suggesting that the brain activity related to this component reached its maximum relatively quickly. This pattern is incompatible with the position that comprehenders fail to immediately encode the case information in the context before the critical verb arrives[16]. Although I cannot confidently identify this effect as an N400 effect, I can be confident that it is a cognitive (rather than perceptual) component given its latency, and it therefore suggests that the case marker is cognitively encoded early on. Finally, as I will see in Experiment 5, a minimal case marker change does affect N400 in

[16] Of course, evidence that case marking is encoded is not in itself evidence that argument roles are also encoded. After recognizing case marking, comprehenders must perform one more step of computation to map grammatical functions to argument roles, because the mapping between grammatical function and argument roles is not one-to-one, e.g., in passive sentences. However, in half of the critical materials used here (e.g., *bee-ACC sting*), this option is unavailable because passive verbs cannot co-occur with accusative arguments. In addition, in the other half of the materials, it is highly unlikely that comprehenders assign the theme/patient role to the nominatively marked noun, because (i) passives are non-default structures (ii) no inanimate noun is used in the experimental sentences, and (iii) the current experiment contained no passive sentences. Thus, the current experiment strongly encouraged comprehenders to assign the default argument role corresponding to each case marker.

Japanese without delay (within 750ms SOA to be precise), when the case change alters the syntactic category of likely upcoming words. This suggests that comprehenders can in principle incorporate case information to modulate prediction of an upcoming word without much delay. For these reasons, I argue that it is highly unlikely that comprehenders fail to assign argument roles within 800ms.

*Prediction as memory retrieval*

Why are argument roles not initially effective for verb prediction? In order to answer this question, I adopt the view that the goal of the lexical prediction mechanism is to retrieve a (set of) word(s) that satisfy all the contextual constraints prior to the bottom-up input, using the context stimuli as a retrieval cue. That is, lexical prediction is essentially a cue-based retrieval process. Given such a conceptualization of prediction, the question is why retrieving a verb from arguments' roles takes so much time. I offer three possible accounts below. In describing each account, I use an example sentence from the current material, *bee-ACC swat/sting*, for illustrative purposes. In this example sentence, the retrieval cue is *bee-ACC*, and the target item in memory is a verb that takes *bee* as a typical patient, for instance *swat*, and the non-target competitor is *sting*.

First, it is possible that the prediction system suffers from similarity-based interference (e.g., Van Dyke & McElree, 2006) in predictive word retrieval. In this account, a general property of memory retrieval models (e.g., Gillund & Shiffrin, 1984) predicts that retrieving a target word should be harder when the retrieval cue has more associated words in memory. So, to the extent that the features of *bee-ACC* is associated with those of other non-target words, such as *sting*, the retrieval of *swat* is more laborious. Such an association between the retrieval cue *bee-ACC* and the non-target competitor *sting* is not surprising, given that the retrieval cue *bee-ACC* consists of a complex feature (bee as a patient), a part of which is shared with any bee-related words. This retrieval interference is eventually resolved as the system converges to the perfect-matching item *swat*, but interference due to partially matching features between the the retrieval cue and non-target word may explain why it takes relatively long time to generate appropriate verb prediction. Under this explanation, interference will be resolved properly eventually given enough time, but such an interference resolution may or may not catch up with the speed of actual input.

Second, the prediction system may be performing retrieval operations successively as more information accumulates. If this successive retrieval is performed both before and after argument role information is encoded, then later retrieval processes based on the combination of the lexical and argument role information may be hindered by the products of earlier retrieval operations based solely on lexical information. In the current example, the first retrieval operation is performed with *bee* only, and another retrieval operation based on *bee-ACC* follows. Only the second retrieval operation can arrive at the contextually appropriate verb *swat*. This head-start for the retrieval based solely on the lexical cue hinders the subsequent retrieval based on the combination of the lexical and case cues. This approach may be computationally insensitive, but it is nevertheless appealing given that the prediction system in real life must operate at many point in a sentence, and there is no a priori reason to think that prediction happens only at word boundaries.

Finally, the slow effect of argument role information may reflect the organization of semantic memory itself, rather than retrieval interference. Under this account, argument roles are not features that are encoded in verbs' lexical representation, and hence argument roles cannot be used as retrieval cues for content addressable (i.e., direct) memory access. Of course, argument roles must eventually be used for selecting an appropriate verb, but they might only be used in more laborious search-based memory access algorithms where verb candidates that are first retrieved via a direct access mechanism are evaluated for contextual fitness. In the current example, the patient role given by the accusative case marker in *bee-ACC* itself cannot be used as a retrieval cue for *swat* as *swat* does not itself contain a feature that corresponds to a bee-as-a-patient or a patient role itself, and thus only the lexical information *bee* is used for a direct access mechanism. As a result, *sting* and *swat* are both the output of the direct access mechanism, and each can then be evaluated against the context bee-ACC, perhaps by actually computing the compositional meaning and then comparing the computed meaning with world knowledge.

A puzzling finding in the current study is that I did not find a significant P600 effect that has been found consistently in previous studies. One likely culprit for this, I suggest, is that our sentences were short and simple compared to other studies. Assuming that P600 reflects some integrative process, whether syntactic (Kaan et al., 2000; Gouvea et al., 2010) or semantic (Brouwer et al., 2012), it may be the case that a sufficiently rich representation must already be constructed before the word in question appears. Although

I cannot be certain from the current experiment, it is possible that this partly explains why I did not get a P600.

To sum up, in Experiment 4, I showed that the N400's insensitivity to role reversal sentences is observed even when contexts are maximally simple and argument roles are maximally transparent. I furthermore showed that this insensitivity can be overcome simply by adding extra time before the verb, with no linguistic stimuli intervening. This pattern suggests that some computation that is involved in verb prediction can under some circumstances take considerable amounts of time, and I suggested that this is due to constraints in the memory mechanism that are implicated in lexical prediction (either the memory structure itself or the retrieval operation). This account grounds the problem of lexical prediction in the larger framework of memory retrieval, and provides descriptive tools to model lexical prediction as a cognitive phenomenon that involves complex computation with a temporal dimension.

Having discussed in depth the underlying mechanisms of prediction and the computational limits associated with them, now I turn to discuss the relevance of the syntactic gating mechanism as promised above. As hinted already in Chapter 2, the relevance of the syntactic gating mechanism in predictive comprehension derives from the fact that the informative part of a sentence with respect to making a specific lexical prediction is arbitrarily far away from where the predicted lexical item is syntactically allowed to occur. Experiment 4 did not tap into this problem, but note that the difficulty of distinguishing between two semantically attractive verbs (e.g., *sting*, given *bee-ACC*) and syntactically appropriate verbs (e.g., *swat*, given *bee-ACC*) cannot be distinguished by syntactic category. Both of them are verbs. This illustrates the loose interfacing between the lexico-semantic memory system and relational aspects of syntactic representation. This apparent loose interfacing contrasts with the robust interaction between syntactic category representations and the lexico-semantic memory system postulated in the syntactic gating mechanism. This selectively robust interfacing between category and lexico-sematic memory, if it exists in comprehension as well, highlights the special role that syntactic category plays in solving the binding problem between structure and content in predictive comprehension. Experiment 5 addresses this.

## *3.7 Experiment 5: Syntactic gating in predictive comprehension*

Experiment 5 examined the effect of category expectations on lexical processing in sentence comprehension. It examines the existence of a mechanism like the syntactic gating mechanism in predictive comprehension.

### 3.7.1 Method

*Participants*

30 participants from Waseda University participated in the experiment for monetary compensation. All participants were right-handed and were native speakers of Japanese with no significant early (before age of 10) exposure to any other languages. Prior written consent was obtained for each participant.

*Materials and Design*

240 sentences frame were constructed. (3-3a) is an example plausible grammatical sentences. Using this version as a base, the second case marker and the verb

identity was changed to manipulate the syntactic fit and conceptual fit of the following verb. The second case marker of the each frame was changed to create the category mismatching conditions (e.g., 3-3b & 3-3d). Orthogonally, the first verb was changed to a verb that is not conceptually associated with either of the two preceding nouns, in order to create the poor fit conditions. Each condition consisted of two types of sentences (15 sentences each), one in which accusative case is syntactically more appropriate, and another in which genitive case is syntactically more appropriate. Although the category mismatch sentences are ungrammatical, notice that they cannot be judged as ungrammatical at the critical word, and filler sentences ensured that participants could not infer the grammaticality of the sentence at the critical word based on the sequence of case marker and category (see below).

(3-3) a. Good fit, category match

(i) daiku-ga ie-o **tateru** tameni ashiba-o kumitateta.
carpenter-NOM house-ACC **build** in-order-to scaffold made
'A carpenter made a scaffold in order to build a house'

(ii) rakunoka-ga yagi-no **miruku-o** utta kasegi-de saishi-o yashinatta.
Dairy-farmer-NOM goal-GEN **milk-ACC** sold earnings wife-and-children supported
'A dairy farmer supported his wife and children by selling goat's milk'

(3-3) b. Good fit, category mismatch

(i) * daiku-ga ie-no **tateru** tameni ashiba-o kumitateta.
carpenter-NOM house-GEN **build** in-order-to scaffold made

(ii) * rakunoka-ga yagi-o **miruku-o** utta kasegi-de saishi-o yashinatta.

Dairy-farmer-NOM goal-ACC **milk-ACC** sold earnings wife-and-children supported

(3-3) c. Poor fit, category match

(i) daiku-ga ie-o **shiraberu** tameni ashiba-o kumitateta.
carpenter-NOM house-ACC **investigate** in-order-to scaffold made
'A carpenter made a scaffold in order to investigate a house'

(ii) rakunoka-ga yagi-no **niku-o** utta kasegi-de saishi-o yashinatta.
Dairy-farmer-NOM goal-GEN **meat-ACC** sold earnings wife-and-children supported
'A dairy farmer supported his wife and children by selling goat's meat'

(3-3) d. Poor fit, category mismatch
(i) * daiku-ga ie-no **shiraberu** tameni ashiba-o kumitateta.
carpenter-NOM house-ACC **investigate** in-order-to scaffold made

(ii) * rakunoka-ga yagi-o **niku-o** utta kasegi-de saishi-o yashinatta.
Dairy-farmer-NOM goal-ACC **meat-ACC** sold earnings wife-and-children supported

First, note that (3-3a-d) each represent one condition though they contained two types of sentences. The reason for each condition having two types of sentences is to avoid baseline differences between the critical word due to differences in case marking. As you can see, when two versions are averaged over, each condition contains the exact same proportion of accusative vs. dative case marking, avoiding the baseline problem.

In these sentences, the third word (bolded) is the critical word, and two dimensions of this word were orthogonally manipulated. First, I manipulated whether the category of the word matches with the syntactic expectation. When Japanese speakers hear/read a sequence of a nominative noun phrase and an accusative noun phrase, they overwhelmingly expect to hear/read a verb. In contrast, when they hear/read a sequence of a nominative noun phrase and a genitive noun phrase, they overwhelmingly expect to hear/read a noun. This expectation is either satisfied (in the category match conditions), or violated (in the category mismatch conditions). Orthogonally to this factor, I manipulated the conceptual fitness of the word given the content of the two preceding nouns. This was determined by a separate cloze task (n = 28). The conceptually fitting words were the highly probable continuations in the cloze test given the appropriate case marking, with an average cloze of 50%. On the other hand, the conceptually poor fitting words never appeared in the cloze test, with an average cloze of 0%.

In sum, the current experiment adopted a 2 x 2 design, with the factors Conceptual Fitness (Good fit vs. Poor fit) and Category Match (Match vs. Mismatch).

In addition, in order to prevent participants from inferring that the sentence was ungrammatical as soon as they see the verb following a genitive case marker (-*no*, as in 3-3b/3-3d), 16 filler sentences were included like the following.

(3-4) kangoshi-ga byonin-no tsukatta beddo-o soujisita.
nurse-NOM patient-GEN used bed-ACC cleaned
A nurse cleaned the bed that the patient used.

This sentence, due to the existence of so-called *nominative-genitive conversion* (e.g., Harada 1971) in the subject of the embedded clause, is a grammatically well-formed sentence. The number of this type of filler sentence is more than the number of the experimental sentences with genitive case-verb sequences (15 sentences). Thus, participants should not have been able to reliably predict the grammaticality of the relevant sentences at the verb position in experimental conditions like (3-3b & 3-3d).

Furthermore, in order to prevent participants from inferring that the sentences were ungrammatical as soon as they saw two accusative case noun phrases in succession, 16 filler sentences were like the following.

(3-5) demotai-ga shushokanteimae-o purakado-o kakagenagara koshinshita.
protesters-NOM prime-minister's-residence-in-front-of-ACC placard-ACC while-displaying marched
'A group of protesters marched in front of the prime minister's residence, displaying placards'

This sentence, due to the existence of a null subject embedded clause in Japanese, is a grammatically well-formed sentence. Again this type of filler sentence outnumbered the sentences with two accusative-marked nouns that ended up being ungrammatical in the experimental conditions (15 sentences). Thus, participants should not have been able to reliably predict the grammaticality of the relevant sentences at the noun position in experimental conditions like in (3-3b & 3-3d).

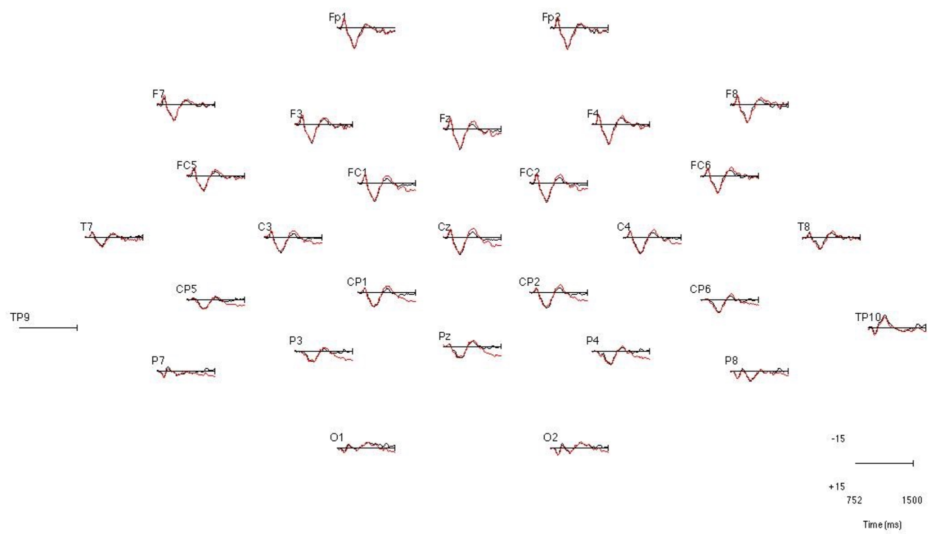
*Procedure*

Each participant read a total of 240 sentences (120 experimental, 120 filler) while their brain activity was measured via EEG. Psychopy (Peirce, 2007) was used for stimulus presentation. At the beginning of each trial, a fixation cross was presented at the center of the screen for 500 ms. Following a 400ms blank screen a sentence was presented in a word-by-word manner. Each word was presented for 400ms and separated by either a 350ms blank screen 1000ms after the second word offset or a screen that read 'Natural?' appeared (in English). Participants were instructed to judge the naturalness of the sentence they had just read after seeing this screen.

3.7.2 Results

*ERP Results*

Critical Verb: The ERP responses to the critical verb for each SOA condition are shown in Figure 3-12 below.



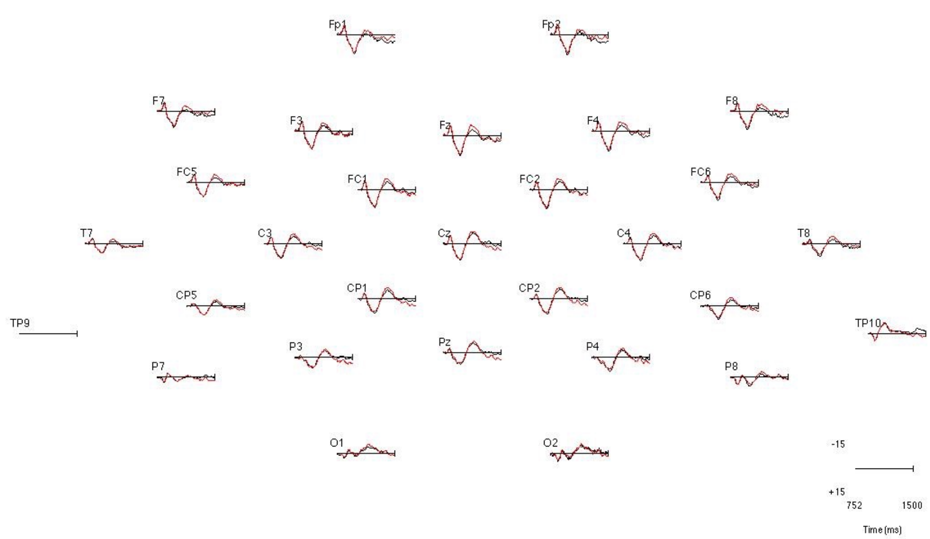



Figure 3-12: Grand average ERP to the critical words comparing category matching and category mismatching words, in conceptually fitting (top) and unfitting (bottom) conditions.

In the 350-450ms time-window after the critical word onset, a repeated measures ANOVA analysis at midline electrodes (Cz, Pz, Fz), medial electrodes (FC1, FC2, C3, C4, CP1, CP2), lateral electrodes (F3, F4, FC5, FC6, CP5, CP6, P3, P4), and peripheral electrodes (Fp1, Fp2, F7,F8, T7, T8, P7, P8, O1, O2) were conducted. These analyses involved Conceptual Fit, Category Match, and Electrode Site as factors. At the midline electrodes, the three way ANOVA analysis revealed a significant main effect of Conceptual Fit ($F(1,29) = 13.03$, $p <0.01$), a significant effect of Category Match ($F(1,29) = 4.39$, $p < 0.05$), a main effect of Electrode Sites ($F(2,58) = 7.37$, $p < 0.01$), and a significant interaction between Conceptual Fit and Electrode Sites ($F(2,58) = 8.48$, $p < 0.01$). No other effects were significant ($p > 0.05$).

At the medial electrodes, the three way ANOVA revealed a significant effect of Conceptual Fit ($F(1,29) = 16.21$, $p < 0.001$), a significant effect of Category Match ($F(1,29) = 8.88$, $p < 0.01$), a significant effect of Electrode Site ($F(5,145) = 2.37$, $p <0.05$) and a significant interaction between Conceptual Fit and Electrode Sites ($F(5,145) = 3.75$, $p < 0.01$). No other effects were significant ($p > 0.05$).

At the lateral electrodes, the three way ANOVA revealed a significant effect of Conceptual fit ($F(1,29) = 4.45$, $p < 0.05$), a significant effect of Category match ($F(1,29) = 4.51$, $p < 0.05$), and a significant interaction between Conceptual fit and Electrode site ($F(5,145) = 3.75$, $p < 0.01$). No other effects were significant ($p > 0.05$).

At the peripheral electrodes, the three way ANOVA revealed a significant effect of Electrode Sites ($F(9,261) = 7.13$, $p < 0.001$) a significant interaction between Conceptual Fit and Electrode Sites ($F(9,261) = 5.24$, $p < 0.001$). No other effects were significant ($p > 0.05$).

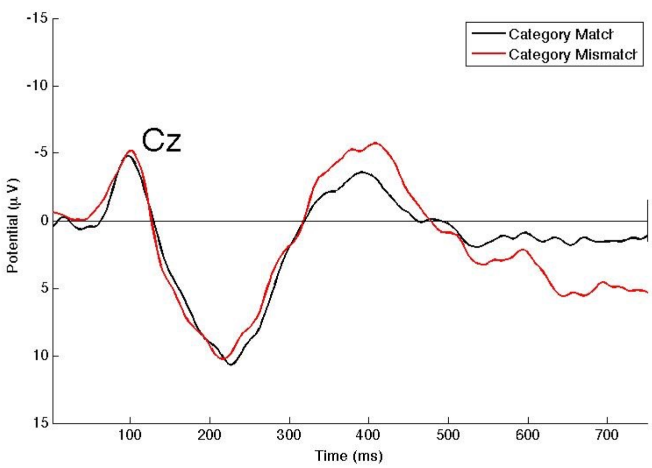
In the 500-700ms time window, the same analyses as in the earlier time window were conducted. At the midline electrodes, the three way ANOVA analysis revealed a significant main effect of Category Match ($F(1,29) = 20.66$, $p < 0.001$) and a significant main effect of Electrode Sites ($F(2,58) = 3.19$, $p < 0.01$). A significant interaction between Conceptual Fit and Electrode Sites ($F(2,58) = 5.14$, $p < 0.01$) as well as between Category Match and Electrode Sites ($F(2,58) = 6.96$, $p < 0.01$) were found. No other effects reached significance ($ps > 0.2$).

At the medial electrodes, the three way ANOVA analysis revealed a significant effect of Category match ($F(1,29) = 14.44$, $p < 0.001$) and a significant effect of Electrode site ($F(5,145) = 3.09$, $p < 0.05$). No other effects were significant ($p > 0.05$).

At the lateral electrodes, the three-way ANOVA revealed a significant effect of Category Match ($F(1,29) = 9.91$, $p < 0.01$), a significant interaction between Conceptual Fit and Electrode Sites ($F(7, 203) = 3.30$, $p < 0.001$), a significant interaction between Conceptual Fit and Electrode Sites ($F(7,203) = 3.30$, $p < 0.05$) and between Category Match and Electrode Sites ($F(7,203) = 7.03$, $p < 0.001$). No other effects were significant ($p > 0.05$).

At the peripheral electrodes, the three way ANOVA revealed the significant effect of Electrode site ($F(9,261) = 20.90$, $p < 0.001$), a significant interaction between

Conceptual fit and Electrode site ($F(9, 261) = 3.50$, $p < 0.001$), a significant interaction between Conceptual fit and Electrode site ($F(9,261) = 3.30$, $p < 0.05$) and between Category match and Electrode site ($F(7,203) = 4.85$, $p < 0.01$). No other effects were significant ($p > 0.05$).



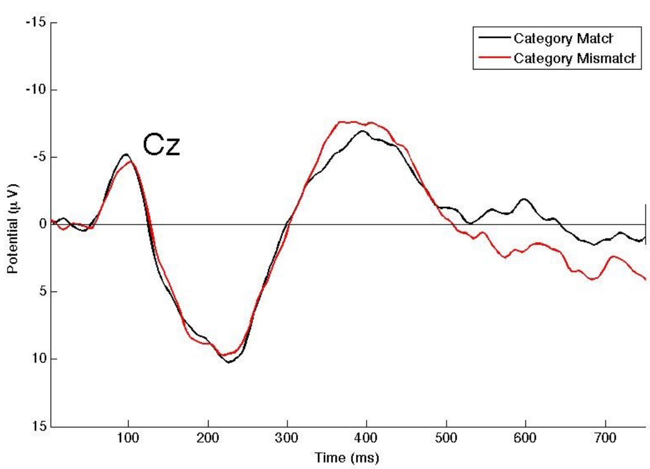
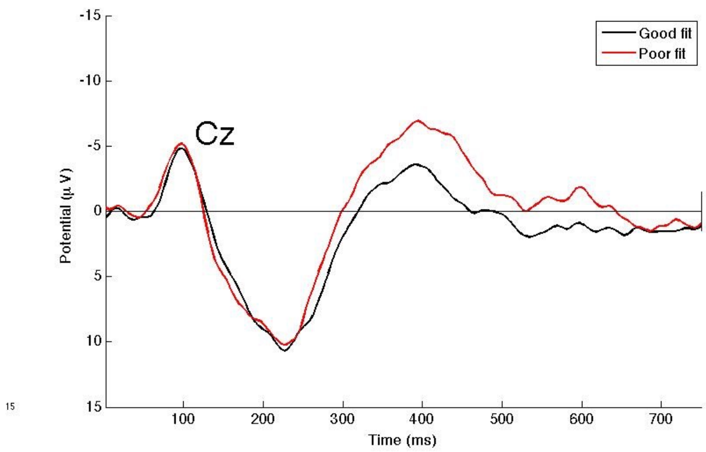



Figure 3-13ab: Grand average ERPs at the critical word at the midline electrode Cz. It compares the response to category matching (black) and mismatching (red) words, with conceptually well fitting (top) and conceptually poorly fitting (bottom) conditions.



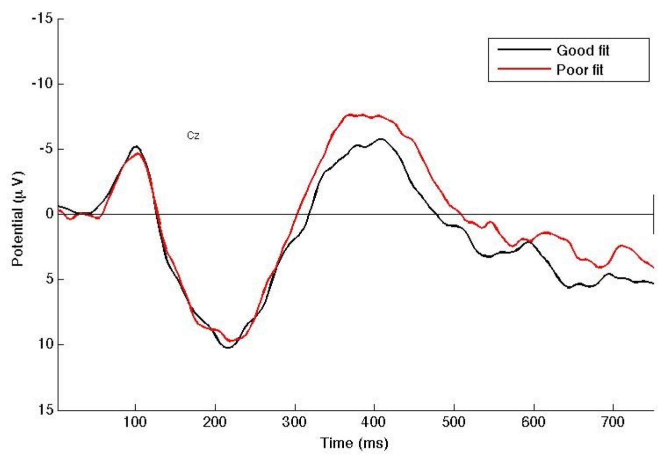




Figure 3-14ab: Grand average ERPs at the critical word at the midline electrode Cz. It compares the response to conceptually well-fitting (black) and poor-fitting (red) words, with category matching (top) and category mismatching (bottom) conditions.

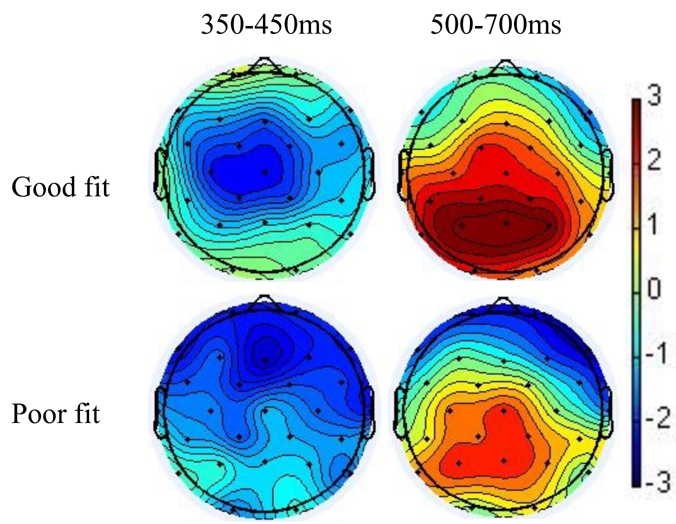



Figure 3-15: Topographic map of the amplitude difference (mismatch - match) in each time-window at the critical word.

### 3.7.3 Discussion

In Experiment 5, I tested whether category expectations affect the prediction of lexical items, even when the conceptual association between the contextual words and critical words are held constant as in Experiment 4. I manipulated the case marker immediately preceding the critical word to change the likely syntactic category of the following word, and this category expectation was either met or violated in the actual stimuli. The violation of such a category expectation increased the N400 response (in addition to the P600 response, see below), a characteristic brain response associated with ease of lexical-semantic memory retrieval (e.g., Federmeier & Kutas, 2000; DeLong et al., 2005). This increase in N400 response was independent from the conceptual fit of the critical word.

Before I discuss the underlying mechanism of the current effects, one might wonder whether the category effect in the current study simply reflects a difference in cloze probability. To put it differently, how is the current effect anything different than the effect that has been demonstrated by many researchers (e.g., Kutas & Hillyard, 1983, among many others)? There are two patterns in the data that cannot be easily accounted for by a simple cloze probably explanation. First, in the conceptual poor fit conditions, the case manipulation did not change the cloze probably. This is because the baseline words of the relevant category already had a cloze probability of zero, i.e., at floor. Changing the case cannot lower the cloze probability any further. Of course, it is possible that, if a larger sample of people is tested in the cloze test, it might give a small difference because some people might come up with the conceptual unfitting words that I used. However, I consider it unlikely that such a small potential difference, if any, could explain a fairly large effect (by ERP standards) that creates a 2μv difference. Thus, a simple cloze probability explanation does not capture the category effect straightforwardly. Furthermore, a similar radical change in lexical probability due to case marker changes did not increase the N400 amplitude in Experiment 4 (in the short SOA condition). Admittedly, there are many differences between Experiment 4 and the current experiment. For instance, I did not conduct a cloze test for Experiment 4, because there was a constraint that the sentence needed to be exactly two words with transitive verbs. Experiment 4 and the current experiment had different behavioral tasks and different lengths of sentences, etc. However, despite those differences, they share one essential aspect: both involve simple case marker changes and nothing else. These case marker changes, both in Experiment 4 and the current experiment, have a clear effect on the likely continuation of the sentences. Nevertheless, the N400 profile was contrastive in the two experiments. In Experiment 4, case change induced little/no N400 effect, while in Experiment 5, case change clearly did induce an N400 effect. Note that the SOA, if anything, was shorter in the current experiment than in Experiment 4. This contrast needs to be explained, and cannot easily be captured via a simple cloze probability change associated with the case marking change. I argue instead that the N400 contrast between the two experiments is due to the difference in the syntactic consequences associated with the case changes.

Now I turn to a discussion of how the current results can be explained in terms of the syntactic gating mechanism proposed above. In the current experiment, no significant interaction between category match and conceptual fit was found. At first glance, this additivity of the syntactic and conceptual effects may seem to be in conflict with the syntactic gating mechanism, because the syntactic gating mechanism should block the retrieval of category-irrelevant words regardless of their strength of conceptual association with the context. However, it needs to be emphasized that the N400 probably does not reflect lexical retrieval only, but also reflects a conceptual level of processing as well. Indeed, the N400 has been demonstrated to be elicited by non-linguistic visual stimuli, like videos of real world events (Sitnikova, Kuperberg, Holcomb, 2003). As the conceptual level of representation cannot by definition contain features like syntactic category, the syntactic gating mechanism cannot possibly eliminate any conceptual-level effects. Thus, it is not surprising that the category mismatch effect was additive to the effect of conceptual fit. The point that I would like to emphasize is that the current effect suggests that the violation of syntactic category expectation has an effect on a lexico-semantic index (N400 amplitude), suggesting an interaction between syntactic category prediction and lexico-semantic memory retrieval. This interaction would not be possible without the existence of some category representation prior to lexico-conceptual processes (i.e., the existence of top-down structure building). The mechanism that I described as the syntactic gating mechanism may capture that interaction.

Finally, in addition to the N400 effects, the syntactic category mismatch increased the P600. This is entirely expected. The category mismatch by definition results in the retraction of the default phrase structure parse, and accordingly, induces P600 effects (e.g., Osterhout & Holcomb, 1992). This result shows that participants are indeed sensitive to the case marking changes, and inferred the syntactic consequence rapidly. Note, however, that this P600 effect does not necessarily indicate that participants regarded the sentence as ungrammatical as soon as they saw the critical word. Indeed, the current experiment was designed such that the grammaticality of the sentence could not be judged at the critical words, as explained in the Materials and Design section. It also has been shown that P600 does not necessarily index the detection of a syntactic anomaly, but reflects syntactic processing cost (e.g., Kaan et al., 2000; Gouvea et al., 2010). Thus, the P600 effect obtained at the critical word does not imply that the participants were treating the sentence as ungrammatical at that point.

### *3.8 Conclusion*

In this chapter I introduced 4 experiments (2 in production; 2 in comprehension) that examined the interaction between syntactic processes and lexical processes. Experiments 2 and 3 showed that category representation strongly constrain lexical retrieval in sentence production, in such a way that category-irrelevant words do not interfere with the retrieval process despite their conceptual similarity to the target. Experiments 4 and 5 showed a timing contrast in the effectiveness of case markers on lexical prediction, and suggested that a case manipulation immediately affects the neural index of lexicon-semantic retrieval processes (i.e., N400) only when the case marker changes the upcoming category. This pattern suggests a close interaction between syntactic category representations and lexico-semantic memory access in predictive comprehension, and it is consistent with the view introduced as the syntactic gating mechanism. Despite the distinct task and measurements deployed in Experiments 2-3 and Experiments 3-4, the underlying mechanism that characterizes the interaction between syntactic and lexical processes can be characterized analogously. This needs not be the case if the parser and the producer are the distinct mechanisms.

# 4 The timing of verb planning in sentence production

This chapter investigates in detail the order in which structures are built in production. I devote this chapter to studying production because, unlike comprehension, it is far from clear if production processes proceed from left-to-right, given that the message components likely do not come in the same order as they are uttered. Specifically, I focus on examining the timing at which verbs/predicates are planned in verb medial and verb final utterances, in English and Japanese. The reason why I focus on verbs is that they are the 'glue' for other parts of sentences, and hence potentially have a functional significance that other parts of sentences do not have. Given this functional significance both at the message and the syntactic levels, verbs that appear late in a sentence may have a different timing of processing than, for instance, nouns. Thus, studying the timing of verb planning across constructions and across languages may be particularly revealing about the order in which sentence pieces are stitched together as speakers build sentences.

Additionally, the question about the order in which different parts of a sentence are planned has important implications into how incrementality and memory costs should be defined in sentence production. As I will review in Section 4.10 below, the notions of incrementality are currently defined in comprehension and production substantially differently, and it needs to be shown that the same notion of incrementality can be applied in describing the behavior of the production system. In this chapter, I will argue that the notion of incrementality defined as connectedness (e.g., Sturt & Lombardo, 2005) fares better than the oft-invoked notion of incrementality in sentence production models, which I name cascading incrementality.

To look ahead, the generalization that emerges from the experiments introduced in this chapter is that verbs are consistently planned before articulation of their internal arguments, but not before articulation of their external arguments, regardless of whether the target language is English or Japanese. As I will see in the General Discussion, this selective advance planning of verbs before internal argument noun articulation can be explained once I jointly consider the top-down nature of structure building, the modest size of structural units, and the syntactic gating mechanisms, all of which were introduced and motivated in Chapter 3.

## *4.1 The Awful German Language and the timing of verb planning*[17]

*… well, in a German newspaper they put their verb away over on the next page; and I have heard that sometimes after stringing along the exciting preliminaries and parentheses for a column or two, they get in a hurry and have to go to press without getting to the verb at all. Of course, then, the reader is left in a very exhausted and ignorant state.* - A quote from The Awful German Language by Mark Twain (1880)

Uttering a sentence requires that a single message be converted into a sequence of words. There is widespread agreement that this conversion occurs incrementally, meaning

[17] Sections 4-1 to 4-6 are lightly edited versions of sections from Momma, Slevc & Phillips (2015).

that speaking can begin before the entire utterance is planned. A strong version of the incrementality hypothesis (e.g., Brown-Schmidt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2006; Griffin, 2001; Iwasaki, 2011; Schriefers, Teruel & Meinshausen, 1998) posits that the sentence plan is, most of the time, developed in the same order that the words are uttered, i.e., the first word is planned first, the second word second, and so on. This strong version of the incrementality hypothesis holds that the sentence plan is developed on a just-in-time basis, such that words are planned right before they are uttered. This approach might be beneficial for reducing memory demands and avoiding interference between multiple activated lexical items, but it requires that utterance plans are flexibly adjusted to the word order demands of a language, so as to avoid 'looking-ahead' in planning. The best tests of this strong incrementality arise in situations where words that play a central role in the organization of a sentence are forced to appear late in the utterance, due to the word order constraints of a language. Here I examine such a case, focusing on evidence for look-ahead effects in Japanese, a language whose word-order constraints force verbs to appear in utterance-final position.

Linguistic and psycholinguistic evidence suggests that verbs are critical for structural processing. This points to the early encoding of verbs in utterance formulation. Accordingly, many influential models of sentence production (e.g., Bock & Levelt, 1994; Ferreira, 2000; Kempen & Hoenkamp, 1987; Levelt, 1989) adopt the view that the verb's syntactic representation (i.e., *lemma*; Kempen & Huijbers, 1983) guides structural processes, and thus they predict that a verb's lemma is selected before the relevant structural processes are performed. Most explicitly, Ferreira (2000) argued that the selection of a verb's lemma (or the head of any phrase) must be performed before phonological encoding of the first phrase of a sentence is finalized.

Despite the emphasis on verbs' early encoding in models of sentence production, the experimental evidence from tests of this issue is equivocal at best. Some suggestive evidence for advanced verb selection comes from Kempen & Huijbers (1983), who found that changing the target verbs between experimental blocks delayed speech onset for both verb-subject and subject-verb sentences in Dutch, though with a greater delay for verb-subject than subject-verb sentences. This suggests that either verbs' conceptual representation, their lexical representation, or both are at least partially planned in advance, although the results are ambiguous between these three interpretations. In addition, Schnur, Costa and Caramazza (2006) and Schnur (2011), using the multi-word picture-word interference task (Meyer, 1996), found that distractors that are phonologically related to the target verb in SV and SVO utterances facilitated speech onset, suggesting that even the phonological representation of a verb is encoded in advance. However, this evidence is hard to interpret because facilitative effects may alter normal planning in favor of advance planning of verbs, due to potential feedback activation from the phonological level to higher levels, and because these findings conflict with existing evidence that suggest a narrow scope of planning at the phonological level (e.g., Griffin 2001; Meyer, 1996; Meyer, Slederink & Levelt, 1998; Wheeldon & Lahiri, 1997). Nevertheless, the studies by Schnur and colleagues are broadly consistent with models that posit the advanced selection of verbs.

On the other hand, the most direct experimental test of the issue failed to provide evidence for advanced verb selection. Schriefers and colleagues (1998) used an extended version of the picture-word interference paradigm in German. In five experiments, participants were presented with action pictures with distractor words. The participants' task was to complete a lead-in fragment by describing a picture as quickly as possible using either a verb-initial (e.g., *laugh – the girl*) or verb-final (e.g., *the girl - laugh*) clause, while ignoring distractors (e.g., *cry*). Schriefers and colleagues reasoned that if verb selection occurs prior to utterance onset then the Semantic Interference (SI) effect, a relative delay of utterance onset due to semantic relatedness between target and distractor verbs, should obtain even in verb-final utterances. However, distractors that were semantically related to the target verbs elicited SI effects only in verb-initial utterances (although see Hwang & Kaiser, 2014 for evidence from English to the contrary[18]). These findings led some researchers to conclude that verb selection is not necessarily performed in advance (Allum & Wheeldon, 2007; Iwasaki, 2011). This chapter attempts to reconcile the theoretical motivation for advance verb selection and the previous empirical failure to demonstrate it by examining the timing of verbs' lemma selection at two different points in sentences: before the internal argument noun and before the external argument noun's articulation.

---

[18]Hwang and Kaiser (2014) showed a semantic interference effect from the verb in English SVO sentences, suggesting that advance verb selection can occur before subject articulation. This pattern contrasts with the results from Schriefers et al. (1998). However, Hwang and Kaiser's choice of unrelated distractors was problematic in multiple respects, making it hard to interpret the results. First, the related and unrelated distractors were not matched in terms of number of repetitions throughout the experiment. Each of the related distractors was presented twice, while the unrelated distractors were presented only once, leading to a potential independent advantage for related distractors. Second, the related and unrelated distractors were not matched in terms of lexical frequency. The mean corpus log frequency of the related distractors were 7.74 (SEM = 0.19) while that of the unrelated distractors were 7.11 (SEM = 0.24) and this difference was marginally significant in two tailed t-test that does not assume equal variance, even though the number of items was very small (4 for the related condition and 8 for the unrelated condition). Finally, Hwang and Kaiser used adjectives as unrelated distractors. They argued that same-category distractors (i.e., verb distractors) might cause interference and thus they should be avoided, but this is exactly the reason why the same grammatical category distractors should be used for related and unrelated condition alike. Otherwise, the interference effect cannot be reliably identified as a semantic interference effect. It cannot be identified as a grammatical category mismatch effect either, because there is no guarantee that verb distractors and adjective distractors have the same effect on the production of the sentence-initial noun, since adjectives and verbs have very different distributional relations to nouns. Therefore, I consider it premature to take the verb interference effect in Hwang & Kaiser's English experiment as evidence for advance verb planning before subject nouns in English. Indeed, I failed to replicate Hwang & Kaiser's semantic interference effect using balanced related vs. unrelated verb distractor in an English SVO sentence production task even though the distractors were shown to be effective in a single verb naming context (see Experiment 9 below).

### *4.2 Linguistic contrasts between subjects and objects*

Linguistic analyses suggest that objects are more strongly dependent on verbs than subjects in many respects. First, it has sometimes been argued that a verb and its internal arguments constitute the verb's argument structure while external arguments are not part of such a structure (cf. Marantz, 1984; Kratzer, 1996; 2002). This claim is derived from the observation that the choice of the object, but not the subject, has a significant impact on the meaning of the verb. Second, verbs do not select subjects in the same sense that they select objects – a subject is obligatorily present regardless of the properties of the verb (Chomsky, 1981; see Lasnik, 2003, for a discussion), while the presence/absence of an object depends on the subcategorization property of the verb (e.g., Haegeman, 1991). Third, subjects and objects are considered to receive case from different sources. The case-assigner of object nouns is generally considered to be the lexical head V(erb), while that of subject noun's is considered to be the functional head I(nflection) in nominative-accusative languages (Chomsky, 1981). Finally, objects have a closer constituency relationship with the verb. In a transitive sentence, the verb and the object noun phrase together form a verb phrase, while a subject noun phrase and a verb do not by themselves form a syntactic constituent under most accounts (e.g., GB, Chomsky, 1981; HPSG, Pollard & Sag, 1994; LFG, Bresnan, 2001). These linguistic analyses all suggest that object nouns are more closely associated with verbs than subject nouns, both syntactically and semantically.

### *4.3 Experiment 6: Japanese SV/OV - control experiment I*

Given the difference between subject and object nouns in terms of their dependency on verbs, it is possible that verb selection might be required before object articulation, but not before subject articulation. To test this hypothesis, I adapted the extended picture-word interference paradigm (Meyer, 1996; Schriefers et al., 1998) to Japanese, with some modifications.

In Experiments 6-8, I exploited two properties of Japanese: strict head-finality and liberal argument dropping. Specifically, in order to probe the timing of verb selection both before subject and object articulation based on speech onset latency, one needs to be able to naturally elicit two types of sentences: one starting with an object noun phrase, another starting with a subject noun phrase, preferably in canonical word order. This condition can naturally be met with Japanese. Japanese allows complete sentences consisting either of subject-verb (SV) or object-verb (OV) sequences. The OV structure allows us to test whether verbs are selected prior to object noun phrase articulation. SV and OV utterances are both naturally producible grammatical sentences in Japanese, and they can be closely matched in terms of the length of their initial noun phrase. Therefore, they are well suited for comparing the status of advanced verb planning before subjects and objects are uttered.

Experiments 6-8 used a similar picture-word interference design and are closely related to each other. Experiment 8 was a sentence production task that tested the key question of the current study: whether verbs are planned before object articulation but not before subject articulation. Experiment 6 was a single-word production task eliciting only verbs, to verify the effectiveness of the picture-distractor pairs used in Experiment 8. Example stimuli used in Experiments 7 and 8 are illustrated in Figures 4-1 and 4-2 below. Experiment 7 was also a single-word production task eliciting only nouns, to verify that any interference effects observed in Experiment 8 were reflections of verb planning rather than noun planning. These three experiments in concert are necessary to attribute the presence/absence of a verb-based semantic interference effect to the presence/absence of the advanced selection of verb lemmas.

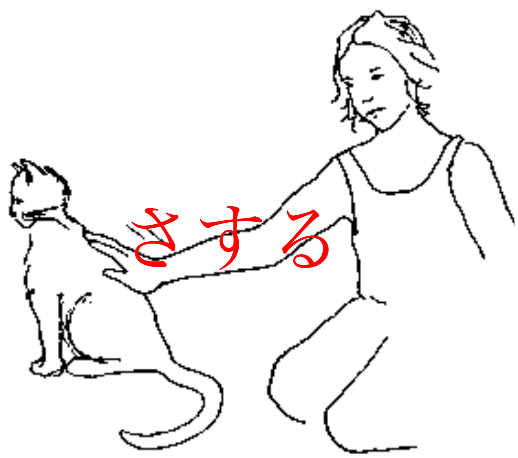



Figure 4-1: A sample picture stimulus for transitive verbs/sentences in Experiments 7-8. Target utterance: neko-o naderu (cat-ACC pet). Distractor: sasuru (rub).


Figure 4-2: A sample picture stimulus for intransitive verbs/sentences in Experiments 7-8. Target utterance: inu-ga hoeru (dog-NOM howl). Distractor: naku (cry).

### **4.3.1 Methods**

*Participants*

Twenty-four students from Hiroshima University in Japan participated in Experiments 6, 7 and 8 in exchange for 500 yen.

*Materials*

Twenty-four action pictures were chosen from the UCSD International Picture Naming Database (Szekely et al., 2004): half corresponding to transitive verbs, and the other half corresponding to intransitive verbs. For each picture two types of distractors were chosen, one semantically related to the target verb and the other unrelated to the target verb. The related distractors were chosen based on two native Japanese speakers' intuitive judgments, and they involved synonym, antonym, or co-hyponym relations. Importantly, the related distractor words were all used as unrelated distractors for other pictures, with the effect that the set of distractor words was identical in the related and unrelated conditions. The unrelated distractors were also matched to the target word in terms of transitivity. This ensured that uncontrolled parameters of the distractor words such as frequency, length and orthographic complexity could not differentially affect the speech onset latencies in the related and unrelated conditions. Neither the related nor unrelated distractors had a systematic phonological relationship to the target words.

*Procedure*

Participants were tested in a sound-attenuated, dimly lit room with an experimenter present. In a familiarization phase they first saw each picture and produced the associated target verb, and then practiced the action-naming task using pictures and distractors that were not included in the experimental set. In the experimental session for Experiment 1, participants were instructed to produce one word that described the action depicted by each picture as soon and as accurately as possible. On each trial a fixation-cross appeared at the center of the screen for 500 ms with a brief click sound (used for calculating speech onset), and then an action picture appeared simultaneously with a written distractor word. The distractor word disappeared after 300 ms, while the picture remained for 1500 ms. A 3000 ms blank screen separated the trials. Participants saw each picture twice, in the related and unrelated conditions, in different blocks. The ordering of pictures within a block was randomized for each participant, and the presentation of a picture with the related distractor in the first vs. second block was counterbalanced across participants.

*Analysis*

For each trial, speech onset latency was manually measured, specifically, the interval between the click sound and the onset of speech minus 500 ms, using Praat (Boersma & Weenink, 2012). The measurer was blind to the conditions, although he could identify the target utterances. Any trials with disfluencies, audible non-speech noise before utterance onset, or speech onset of more than 2000 ms were excluded. In addition, for each participant, trials with response times more than two standard deviations away from that individual's mean RT were excluded. In total, 12.7% of the data points were excluded (7.3% due to errors and 5.4% due to trimming[19]). Both the mean and the dispersion (standard deviation) of each condition was analyzed. The analysis of dispersion was conducted here because (i) the mean analysis of Experiment 8 reported below does not yield unequivocal results and this analysis can serve as a point of comparison and (ii) there are good reasons to think that the semantic interference effect, or incongruity effects in any Stroop-like task, might not be best captured by a difference in the sample means (Heathcote, Popiel & Mewhort, 1991; Scaltritti, Navarrete & Peressotti, 2014)[20].

[19]Analysis without trimming yielded the same results.

[20]I have not conducted a full ex-Gaussian distributional analysis because I did not have enough trials per condition to obtain a stable estimate of the ex-Gaussian parameters. Standard deviation can capture a simple measure of the distributional parameter that is not at all ($\sigma$) or not strongly ($\tau$) reflected by sample means.

#### 4.3.2 Results

Mean RTs (of participant averages) in milliseconds as a function of Verb type and Relatedness in Experiment 6 are shown in Table 4-1.

| | Relatedness | |
|---|---|---|
| **Verb type** | **Related** | **Unrelated** |
| **Intransitive** | **786 [18]** | **760 [16]** |
| **Transitive** | **794 [17]** | **761 [17]** |

Table 4-1. Mean speech onset latencies based on participant means in milliseconds as a function of Verb type and Relatedness in Experiment 6: Standard error of means in square brackets.

A mixed effects model with a maximal random effects structure was constructed, followed by model simplification based on a maximum likelihood ratio test, with the alpha level of 0.1.[21] As a result, the random slope of any factor by items or by subjects was not significant (all *p*s > 0.15). Also, I tried forward selection of the model and the resulting model did not change. Thus, I report models with by-subjects and by-items intercepts. The R code corresponding to the LMER model building can be found in Appendix 1. This analysis revealed an effect of relatedness ($\beta$ = 31.03, *SE* = 10.93, |t| =

[21]Barr et al. (2013) argued against this data-driven approach for selecting random effect structures in confirmatory hypothesis testing. For this experiment, including the maximal random effects structure following Barr et al. (2013) did not change the statistical pattern (but see Experiment 8 for diverging results and discussion).

2.84, $p < 0.01$). In both models, neither the effect of transitivity nor the interaction between Relatedness and Transitivity was significant (all *p*s > 0.85)

In order to make a comparison between this experiment and Experiment 8, I conducted planned comparisons with Bonferroni correction between related conditions and unrelated conditions for transitive and intransitive verbs. Both transitive ($\beta = 33.55$, $SE = 10.96$, $z = 3.06$, adjusted $p < 0.01$) and intransitive verbs ($\beta = 28.26$, $SE = 10.77$, $z = 2.62$, adjusted $p < 0.01$) independently showed a reliable semantic interference effect.

A repeated measures ANOVA analysis in which each cell represented a participant's (for F1) or an item's (for F2) standard deviation (see Heathcote et al., 1991 for an analogous method of analysis in the Stroop task) revealed that the standard deviations in the related and unrelated conditions were significantly different ($F1(1, 23) = 21.36$ , $p < 0.01$; $F2(1, 23) = 7.71$, $p = 0.01$), without any interaction with verb type ($p > 0.5$). This suggests that the semantic relatedness of the distractor has an effect on the dispersion of the reaction time distribution. This semantic interference effect on dispersion is visualized in Figures 4-3 and 4-4 by means of vincentile plots (e.g., Vincent, 1912; Staub, 2010), using only five bins due to the relatively small number of trials per condition per participant. These plots suggest that the semantic interference effect is only reliably identified in the right tail of the distribution (i.e., for slow responses), which is normally reflected by the $\tau$ parameter of an ex-Gaussian distribution, and has the following relation to the sample standard deviation: $S=\sqrt{\sigma^2+\tau^2}$ . I however cannot identify whether the significant effect of relatedness resides in $\sigma$ or $\tau$ from the current experiment, and it is possible that the semantic interference effect might reside in both parameters.

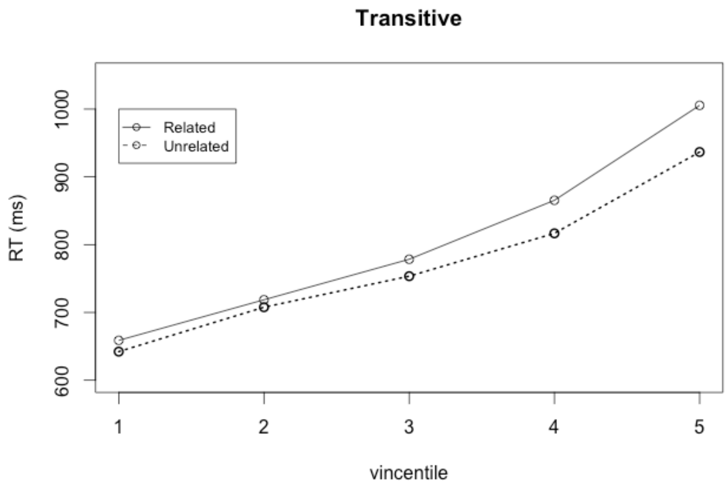



Figure 4-3: Vincentile plot for the transitive condition in Experiment 6, with 20% percentile increments.

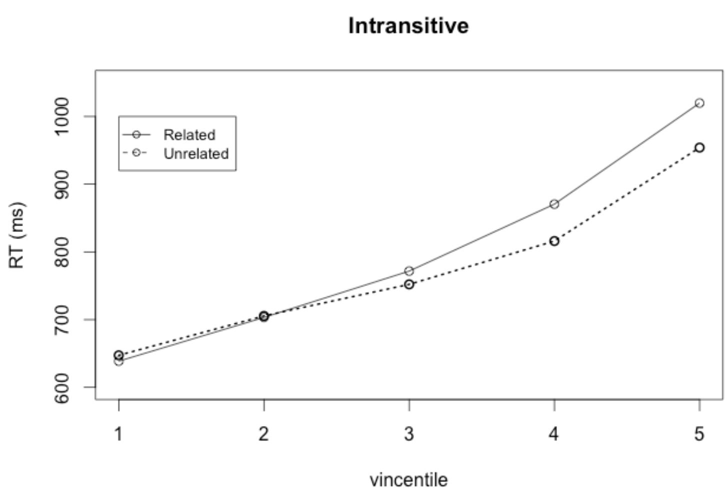



Figure 4-4: Vincentile plot for the intransitive condition in Experiment 6, with 20% percentile increments.

### 4.3.4 Discussion

The aim of Experiment 6 was to test whether the set of verb distractors would reliably elicit a semantic interference (SI) effect. The results clearly show that the verb distractors were indeed effective in eliciting an SI effect regardless of transitivity, both in mean and standard deviation measures (for a discussion of how these measures relate to the ex-Gaussian parameters, see Heathcote et al., 1991). Hence, the same verb-distractor pairs were suitable for testing sentence production in Experiment 8.

## *4.4 Experiment 7: Japanese SV/OV - control experiment II*

### *4.4.1 Methods*

*Participants*

The same set of participants that participated in Experiment 6 participated in Experiment 7, except that one participant was excluded due to recording failure. The order of Experiments 1 and 2 was counterbalanced across participants.

*Materials*

24 new pictures from the IPNP were chosen to elicit the bare target nouns that were also used in Experiment 7, i.e., without associated actions. No actions were depicted in the pictures, unlike in Experiment 6. The distractor words, as well as the pairing between target nouns and distractor words, were kept constant as in Experiments 6 and 8.

*Procedure & Analysis*

The procedure was identical to Experiment 6, except that participants were instructed to describe the object depicted in the pictures in one word i.e., a noun. The same R code that was used Experiment 1 was used for analysis in this experiment, the only difference being that the factor *transitivity* was changed to *noun type*.

### 4.4.2 **Results**

Speech onset latency did not significantly differ across conditions. No effect of noun type, relatedness, or interaction was found (*p*s > 0.35). Mean RTs (of participant averages) in milliseconds, as a function of Noun type and Relatedness in Experiment 2, are shown in Table 4-2 below. A standard deviation analysis as in Experiment 1 revealed no effect of Noun type, Relatedness, or interaction between them (all ps > 0.15).

| Noun type | Relatedness | |
|---|---|---|
| | **Related** | **Unrelated** |
| **Subj. Nouns** | **652 [16]** | **660 [16]** |
| **Obj. Nouns** | **642 [17]** | **647 [17]** |

Table 4-2: Mean speech onset latencies based on participant means in millisecond as a function of Noun type and Relatedness in Experiment 7. Standard error of means in square brackets.

### **4.4.3 Discussion**

The results of Experiment 7 suggest that there was no systematic relationship between the target nouns and distractor verbs that affected utterance onset latency. This is important for ensuring that the nouns that were used in Experiment 8 have no accidental relationship to the distractor verb in a way that could have altered the pattern of semantic interference effects.

## *4.5 Experiment 8: Japanese SV/OV - main experiment*

### <u>4.5.1 Methods</u>

*Participants*

The same set of participants participated in Experiment 8 after completing Experiments 6 and 7, following a short break.

*Materials*

The same set of picture-distractor pairs was used with identical presentation parameters as in Experiment 6. The target nouns preceding the target verbs were matched in mean length (number of moras: 2.33 for the transitive condition and 2.5 for the intransitive condition). The transitive condition targeted 11 inanimate nouns and 1 animate noun, and the intransitive condition targeted 8 inanimate nouns and 4 animate nouns. As participants saw the same set of pictures and distractors in Experiment 1, they had seen each picture four times by the end of Experiment 8.[22]

*Procedure & Analysis*

The same procedure was followed as in Experiments 6 and 7, except that participants were instructed to describe the action depicted in the pictures in sentential forms using two words, i.e., a noun, inflected with a case particle, followed by a verb.

---

[22]This number of repetitions is lower than in many other PWI studies, which often involve 10 or more repetitions per picture. (e.g., Schriefers et al., 1990; 1998)

This elicited an SV sentence for the intransitive action pictures, and an OV sentence for the transitive action pictures. The same analysis procedure was adopted as in Experiment 6 and 7. In total, 13.3% of the trials were excluded (9.4 % due to errors, and 4.9% due to trimming).

### 4.5.2 Results

Mean RTs (of participant averages) in milliseconds as a function of Sentence type and Relatedness in Experiment 8 are shown in Table 4-3 below.

| | **Relatedness** | |
|---|---|---|
| **Sentence type** | **Related** | **Unrelated** |
| SV | **766 [17]** | **764 [18]** |
| OV | **764 [21]** | **736 [16]** |

Table 4-3: Mean speech onset latencies based on participant means in millisecond as a function of Sentence type and Relatedness in Experiment 8; standard error of means in square brackets.

The same mixed effects model analysis procedure was applied as in Experiment 6. After model simplification based on maximum likelihood ratio tests, the model with both by-subjects and by-items intercepts revealed a marginally significant interaction ($\beta$ =

25.20, *SE* = 13.08, |*t*| = 1.93, *p* = 0.054)[23]. The same R code was used as in Experiment 6, except that the factor T*ransitivity* was now changed to *Sentence Type*.

There were no main effects of relatedness or sentence type (*p*s > 0.7). As in Experiment 1 I conducted planned comparisons with Bonferroni correction between related and unrelated conditions for OV and SV sentences. These analyses showed that verb-related distractors significantly delayed sentence onsets in OV sentences (β = 26.53, *SE* = 9.11, *z* = 2.91, adj. *p* < 0.01), but not in the SV condition (*p* > 0.95).

Also, a post-hoc comparison of noun frequencies between OV and SV conditions was conducted to ensure that the differential effect of relatedness on OV and SV conditions was not attributable to the properties of the nouns. Based on the Tsukuba Web Corpus (http://corpus.tsukuba.ac.jp/), the mean log frequencies of the nouns did not differ between the transitive (M = 10.80; SEM = 0.39) and intransitive (M = 9.63; SEM = 0.58) conditions, although there was a trend toward a difference in a two-tailed t-test (*p* = 0.11). A further test using individual items found no correlation between noun frequency and the amplitude of the effect of semantic relatedness (r = 0.07, *p* > 0.7).

As in the verb-only experiment (Experiment 1), I also analyzed the difference in the standard deviation of the Related vs. Unrelated conditions. This analysis can be thought of as a coarse measure of the dispersion of an ex-Gaussian distribution that reflects both σ and τ parameters. This analysis revealed a fully significant interaction between relatedness and sentence type on standard deviations (i.e., dispersion, F1(1, 23)

---

[23]A model with maximal random effect structure (Barr, 2013) yielded non-significant interaction (p = 0.14). However, see the standard deviation analysis and vincentile analysis below.

= 6.07, $p = 0.02$; $F2(1, 22) = 11.41$, $p < .01$), with no significant main effects (ps > 0.13). Planned comparisons revealed that in the OV conditions, the dispersion was larger in the related condition (M = 112; SD = 60)[24] than in the unrelated condition (M = 90, SD = 27; $t1(23) = 2.44$, adj. $p < 0.05$; $t2(11) = 2.73$, adj. $p < 0.05$) while in the SV conditions, the standard deviations were not different between the related condition (M = 105, SD = 31) and the unrelated condition (M = 112, SD = 35; $t1(23) = 1.28$, adj.$p > 0.4$; $t2(11) = 2.32$, adj.$p > 0.08$), indicating that the semantic interference effect on dispersion was only present in the OV utterances. The directionality of this dispersion effect is as expected; Heathcote et al. (1991) also reported that the interference effects in Stroop tasks manifested as increased standard deviations. Thus, this pattern corroborates the central tendency analysis I conducted above. This semantic effect on dispersion is visualized by means of vincentile plots in Figures 4-5 and 4-6.

[24] This larger SD value (60) was due to one participant showing an especially large standard deviation in RT. Removing this participant lowers the standard deviation to 33, but does not change the statistical results.

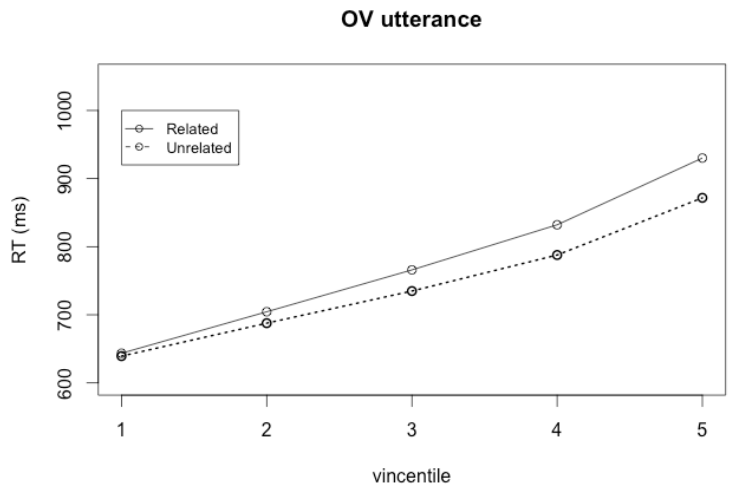



Figure 4-5: Vincentile plot for the OV condition in Experiment 8, with 20% percentile increments.

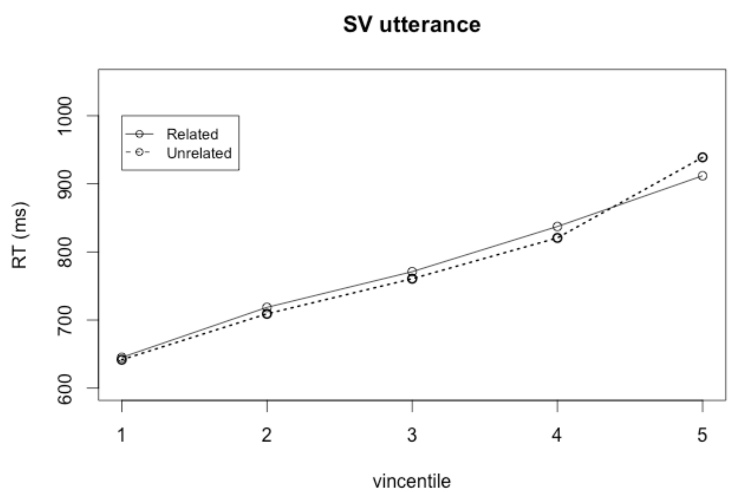



Figure 4-6: Vincentile plot for the SV conditions in Experiment 8, with 20% percentile increments.

## *4.6 Discussion: Experiments 6, 7 and 8*

Experiment 8 showed a semantic interference effect in object-initial sentences but not in subject-initial sentences. Specifically, distractors that were semantically related to the sentence-final verb interfered with the articulation of the sentence-initial noun, but only when the sentence-initial noun was a direct object. Note that the same set of picture-distractor pairs was used as in Experiment 6, where SI effects were reliably obtained for transitive and intransitive verbs alike. Therefore, I can be confident that the contrast observed in Experiment 8 is a consequence of the sentence production task, and not simply a result of poor selection of picture-distractor pairs. Note also that previous research tends to report more consistent semantic interference effects for intransitive verbs than transitive verbs (Schnur et al., 2002; Tabossi & Collina, 2004). Therefore, it is unlikely that the observed contrast between the SV and OV conditions was due to a smaller amplitude of semantic interference in the SV conditions.

The contrast between SV and OV sentences is consistent with our hypothesis that verbs are selected before the onset of objects but not before subjects, reflecting the closer dependency of objects on verbs. The contrasting pattern of SI effects casts the previous failure to find an SI effect in verb-final utterances in German (Schriefers et al., 1998) in a new light. Specifically, it seems premature to draw the conclusion that a producer can dispense with verb information when processing all of a verb's arguments in verb-final utterances. Certainly, our results do not show that speakers retrieve the verb's syntactic information before articulating the object in every utterance. However, our results do show that Schriefers et al.'s (1998) findings are, in fact, consistent with the claim that verb choices guide structural processes in sentence planning. That is, the absence of an SI effect in Schriefers et al.'s data and in the SV condition here, combined with the presence of a SI effect in the OV condition here, likely reflects the selective nature of the advance planning mechanism for verbs.

These results could reflect a potential difference in the relative 'thematic fit' (cf. McRae, Spivey-Knowlton, & Tanenhaus, 1998) between the target nouns and the target verb (*dog-NOM howl*/*cat-ACC pet*) vs. target nouns and related distractor verbs (*dog-NOM cry*/*cat-ACC rub*). To address this concern, I conducted a post-hoc norming study (n = 11) where participants rated the thematic fit between the target nouns and each type of verb (e.g., *dog-NOM howl* vs. *dog-NOM cry* for SV, *cat-ACC pet* vs. *cat-ACC rub* for OV utterances) on a 7-point Likert scale, similar to McRae et al. (1998). A mixed effects analysis analogous to that used in the main experiments was conducted (this model included random intercepts for participants and items as well as random slopes of Relatedness for participants and items) and revealed a marginally significant interaction between verb type (target verb vs. related distractor verb) and sentence type (SV vs. OV; $\beta = -0.87$, SE = 0.50, $|t| = 1.74$, $p = 0.10$) as well as significant main effects of verb type ($\beta = 3.14$, SE = 0.39, $|t| = 6.02$, $p < 0.01$) and sentence type ($\beta = 1.26$, SE = 0.43, $|t| = 2.94$, $p < 0.01$). This marginally significant interaction partially supports the reviewer's concern. However, I find no evidence for a relationship between the magnitude of semantic interference (in Experiment 8) and thematic fit difference (between noun-target verb and noun-distractor verb) in an item based correlation analysis (n = 24, r = -0.14, p =

0.66). Thus, it is unlikely that the differing semantic interference effects for SV vs. OV sentences can be attributed entirely to the difference in thematic fit.

The results of Experiments 6-8 indicate that semantic interference (SI) from distractors related to non-initial verbs – a marker of advanced verb selection – is obtained selectively before object noun articulation (Experiment 8). This selectivity is not likely to be due to the ineffectiveness of the semantic distractors for intransitive verbs (Experiment 6) or to an accidental relation between the preceding nouns and the distractors (Experiment 7). The SI effects reflected by sample means in each experiment are shown in Figure 4-7 below, and the same pattern was observed for the SI effect reflected by sample dispersion. These findings may reconcile the apparent conflict between the need for verb selection in order to determine the syntactic/semantic properties of pre-verbal arguments and the lack of verb-related SI effects before subject noun articulation in Schriefers et al. (1998). Based on this pattern of results, I argue that advance verb selection is selectively performed before object nouns, but not before subject nouns.

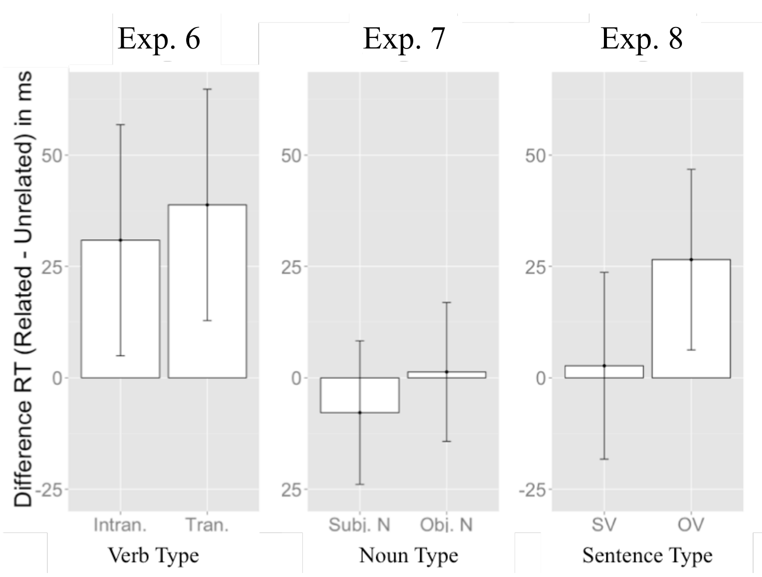


Figure 4-7. SI effect by Verb Type (Exp. 6), Noun Type (Exp. 7), and Sentence Type (Exp. 8), estimated from the statistical model. Error bars represent 95% confidence intervals.

Before I outline the potential mechanism underlying the selective nature of advance verb planning, it is worth noting that the current results at first seems to conflict with previous findings in the scope of planning literature. In particular, the results of Experiment 8 might be regarded as in conflict with a previous finding by Meyer (1996), where the lemma of the second noun in a sentence like "*the dog is next to the baby*" appeared to show evidence of advance planning. That is, it appeared that object nouns, positioned after the predicate, were planned before articulation of the subject noun phrase (*the dog*). If a word that appears after a predicate is already planned, this invites the conclusion that the predicate is also planned. However, this conclusion derives from the assumption that the scope of planning at the lemma level is linear. This assumption is, as far as I know, not warranted. In fact, the production model of Bock & Levelt (1994) assumes that linearization does not occur until the positional processing stage, which is subsequent to lemma selection. Also, it is important to note that speakers repeated the same structure (either *A and B* or *A is next to B*) throughout the experimental session in Meyer (1996), potentially expanding the scope of planning. The current experiment did not allow such formulaic production. Thus, it may not be appropriate to directly compare Meyer (1996) and the current experiments.

In addition, the absence of a semantic interference effect in the SV conditions in Experiment 8 appears to be at odds with the findings by Kempen & Huijibers (1983) and

Schnur et al. (2002), which suggest that a verb's lemma selection occurs even before subject noun articulation. One crucial difference between the current experiment and the previous experiments is that their target utterances had less variation in the pre-verbal noun phrases (4 variants in Kempen & Huijibers (1983), 2 variants in Schnur et al. (2002), and 24 variants in the current experiments). Also, Schnur and colleagues' preverbal noun was a pronoun (s*he or he*), which is particularly easy to process. Thus, although our results and Schriefers et al.'s (1998) results suggest that a verb's lemma is not necessarily retrieved in advance of the articulation of subject nouns, it might, under some circumstances, be retrieved in advance of that point, in situations where there is less response variability and hence less processing difficulty in the preceding words (cf. the retrieval fluency hypothesis; Griffin, 2003).

There are at least three classes of possible mechanisms that could explain the selective pattern of advance verb planning observed in Experiment 8. The first possibility, which, in a broad sense, falls under the lexicalist view of sentence production, is that encoding some structural dependency relations between verbs and their arguments require the advanced selection of verbs. One such potential dependency relation is the assignment of case to the verbs' arguments. Under this account, the selective pattern of advanced verb selection I observed in Experiment 8 can be explained by a difference in how grammatical case is assigned between subject nouns and object nouns. As noted in the section 4.2, accusative case is usually assigned by the lexical head V(erb), whereas nominative case is usually assigned by the inflectional head I(nflection). Thus, the necessity to assign accusative case via a verb head may require that the verb's lemma is selected before the encoding of the object noun is completed. Alternatively, it is possible that the building of structural representations itself is dependent on verbs (e.g., Ferreira, 2000). This is a stronger lexicalist position than the first alternative, in the respect that even the most elementary structural process of building or retrieving phrase structure is considered to be dependent upon verbs under this account. In other words, the current experiments may be interpreted as evidence of partial lexical guidance in phrase structure building: structure building where the subject noun position might not be dependent on verbs, whereas structure building for the object noun position would be crucially dependent on verbs.

The second possibility, which falls under a non-lexicalist view of sentence production, in which structure building can independently occur without lexical retrieval, is that the selective advanced selection of verbs observed here is due to syntactic constituency determining the scope of lemma selection. This possibility is broadly consistent with the proposal that syntactic phrases define the default scope of lemma planning (e.g., Smith & Wheeldon, 1999). This idea implies that phrase structure representations can be built prior to lemma selection (unlike the lexicalist accounts described above), and that such structural representations control the dynamics of lemma selection processes. Although I are not aware of specific claims that the verb phrase defines the scope of planning, our data are consistent with the idea that the syntactic phrase is a determinant of the scope of planning at the lemma level. In this view, the dynamics of lemma selection are controlled by such higher-level syntactic schemes.

A final possibility is that it is the structure of semantic/conceptual representations, rather than syntactic phrases, that is the source of the contrasting pattern that I observed in the current study. Specifically, it is possible that internal arguments and verbs constitute an integrated unit that is planned together at the lemma level or higher, e.g., the message level. This is consistent with the linguistic analysis of Kratzer (1996, 2002), although it is unclear whether semantic representations in linguistic theory should be equated to message level representations in production models. Kratzer argued that external arguments should be excluded from the argument structure of predicates, and suggested instead that inflectional heads introduce external arguments. If this is the case, then advanced verb selection may be due to the semantic necessity to compute a predicate and its internal argument in tandem, or at least temporally closely, either at the conceptual or the lemma level.

There are of course some limitations in the current study. For instance, aside from these theoretically motivated interpretations discussed above, I acknowledge that the SV and OV utterances might differ in how much thematic processing participants must perform before starting to speak. In particular, OV utterances might require deeper thematic processing because they might have forced participants to omit the agent. However, participants in our experiment were probably not actually *forced* to produce OV sentences. Given the structure of the experiments, it was more natural for Japanese speakers to omit the subject than to explicitly mention the subject in the OV conditions because the agent entities in most OV pictures were underspecified (and Japanese speakers normally prefer not to use overt pronouns) and hence hard to name without explicit instruction. This is reflected in the reaction time data: OV utterances had slightly *shorter* response latencies than the SV sentences, suggesting that they were, if anything, less costly than SV utterances.[25] This difference cannot be attributed to the lexical frequency of subject vs. object nouns, as the noun-only experiment showed closely matched average latencies for S nouns and O nouns. Thus, it is unlikely that the contrastive pattern of semantic interference in SV and OV utterances can entirely be attributed to a difference in thematic processing depth.

### *4.7 Experiment 9: English active and passive production*

Experiments 6-8 showed that advance verb planning occurs before an internal argument noun is uttered but not before an external argument noun is uttered in Japanese. Experiment 9 aims to (i) test whether such a generalization holds for English, and (ii) examine the underlying cause of the subject-object contrast in the timing of verb planning. Regarding (i), it is possible that advance verb planning is a strategy that head-final language speakers develop as they acquire language, or that it reflects a more universal characteristic that holds across languages. Regarding (ii), as discussed above, the cause of the timing contrast found in Experiments 6-8 is unclear. The contrast could

[25] One might think that, if the verb is planned in advance in the OV condition, then the reaction times in the OV sentences should generally be longer. However, it should be noted that this reaction time may largely reflect the speed of noun processing given the action picture, so naming a patient noun given a transitive action picture could be easier than naming an agent noun given an intransitive picture. Also, it should be noted that verb planning might not be necessarily costly. This is because the retrieved verb lemma might facilitate the planning of object nouns via semantic priming, or what one might call the 'resonance' between verbs and object nouns. This possible facilitative effect could counteract the cost of advance verb planning and could be additive to the semantic interference effect from related distractors.

reflect the difference in grammatical function or grammatical case between external and internal arguments. Alternatively, it could reflect the closer constituency relationship between internal arguments and verbs. Another possibility is that it could reflect a difference in the degree of thematic/conceptual dependency on verbs. Experiment 9 investigated the contrast between English active and passive sentences, to narrow down the space of possible explanations. In particular, if the external argument vs. internal argument contrast is due to the difference in grammatical function or case, the pattern should not transfer to the English active/passive contrast. This is because, in English passive sentences, subjects are internal arguments. This experiment hence might either support or eliminate the account based on grammatical functions or case.

### 4.7.1 Methods

*Participants*

A total of 71 undergraduate students from the University of Maryland participated in Experiment 9 for course credit. Two participants were replaced due to a high error rate in the active condition (> 40%), and six participants were replaced due to a high error rate in the passive condition (> 40%).

*Materials*

26 transitive action pictures were used as picture stimuli. In these depicted actions there were 13 uniquely identifiable characters that were the participants in those actions, e.g., doctor, boxer, swimmer, etc. Each character appeared four times in total, twice as an

agent of different actions and twice as a patient of other actions. This means that, as a set, the nouns preceding the verb in the active and passive utterances were exactly matched. This avoids the potential confounds of lexical frequency, codability, visual salience, etc. of nouns preceding verbs. The agent and patient entities appeared on the left side and right side 50% of the time. This prevented participants from relying on positional information to identify the agent and the patient in an event. For each action picture, semantically related distractor verbs were chosen according to native English speakers' intuitive judgment. 10 of the semantic distractors were selected from the verbs used as a targets in the other pictures. As in Experiments 6-8, the related distractor words were used as unrelated distractors for other pictures, so that the set of distractor words was identical in the related and unrelated conditions. Some example pictures are presented in Figure 4-8 below.

Figure 4-8: Four sample pictures used in Experiment 9. The target utterance for the top-left picture is *the swimmer is nudging the monk* or *the monk is being nudged by the swimmer*, for the top-right picture *the chef is sketching the cowboy* or *the cowboy is being sketched by the chef*, for the bottom-left picture *the cowboy is inspecting the boxer* or *the boxer is being inspected by the cowboy*, for the bottom-left picture *the cowboy is inspecting the boxer* or *the boxer is being inspected by the cowboy*, for the bottom-right picture *the monk is splashing the doctor* or *the doctor is being splashed by the monk*.

*Procedure*

Participants were tested in a sound-attenuated room with an experimenter present. Before Experiment 9, they experienced the picture naming task via a session where they named all the characters used in Experiment 9 with a different set of pictures, following a familiarization phase with target nouns written on top. A sample of the pictures used in this naming task is presented in Figure 4-9 below. In this familiarization phase, participants studied the pictures at their own pace until they felt comfortable with each picture and its corresponding target verb. Following the familiarization phase, participants also experienced the action naming task where they named all the actions with the verbs used in Experiment 9, with the same set of pictures and with the same set of distractors (with the same presentation parameters and the same number of repetitions). Following these practice sessions they then participated in Experiment 9. Since participants had already practiced naming the characters and actions in the preceding two tasks, there were no additional familiarization or practice trials. In the experimental session, participants were instructed to name the action depicted in the presented picture in sentential form, as if the action was happening right now, i.e., in progressive form. In addition, for the 33 participants who were assigned to the passive condition, they were instructed to first mention the character that corresponded to the target of action. For the other 38 participants who were in the active condition, they were instructed to first mention the character that corresponded to the doer of the action.

Figure 4-9: Three sample pictures used in the character naming task preceding Experiment 9. From the left, the target name is *boxer, chef* and *cowboy*.

Each trial was structured as follows. A fixation-cross appeared at the center of the screen for 750 ms, and then a distractor word was presented visually at the center of the screen. 150 milliseconds after the onset of the distractor's presentation, a picture stimulus was presented on the screen simultaneously with a brief click sound that served as a picture onset marker. The visual distractor disappeared after 500 milliseconds. The picture stimulus remained for 1500 milliseconds. A 3600 millisecond blank screen separated the trials. In each block, participants saw each picture two times with different types of distractors (related or unrelated). The ordering of pictures was pseudo-randomized for each participant, and the order of presentation of a picture with different types of distractors was counterbalanced across participants (except for one participant in the passive condition).

For each trial, the onset latency and the duration of noun production was manually measured using Praat (Boersma & Weenink, 2012). Any trials with audible hesitation (e.g., fillers) and utterances containing non-target words were excluded. Disfluencies were tolerated, as long as the utterance finished before the trial ended. 17% of the data points were excluded in the active condition due to errors, and 21% of the data points were excluded in the passive condition due to errors. For the onset analysis, for each participant trials with response times more than 3000ms and less than 500ms, or more than three standard deviation away from the mean for that participant were excluded. Of the remaining error-free trials, 1% of the data was removed due to this trimming procedure in each of the active and passive conditions. For the duration analysis, trials with durations of more than 2000ms, as well as trials with durations more than three standard deviation away from the mean for that participant were excluded. Of the remaining error-free trials, 2% of the data was removed due to this trimming procedure in each of the active and passive conditions. To combat non-normality, duration measurements were log-transformed before being submitted to the statistical analysis.

4.7.2 Results

The results are summarized in Table 4-4ab below.

| | Mean latency | | Onset interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |
| Active | 1169 [22] | 1170 [20] | -4 |
| Passive | 1288 [38] | 1240 [30] | 48 |

Table 4-4a: Mean onset latency by condition calculated over subject means in Experiment 9. Standard error of means in square brackets.

| | Mean latency | | Onset interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |
| Active | 571 [15] | 549 [14] | 22 |
| Passive | 708 [18] | 700 [19] | 8 |

Table 4-4b: Mean subject + auxiliary duration by condition calculated over subject means in Experiment 9. Standard error of means in square brackets.

*Onset analysis*

A mixed effects model with maximal random effects structure was constructed. However, the by-subjects random slope of Relatedness caused convergence failure and thus was removed from the model. In addition to the experimental factors, the duration of the subject noun phrase was also included as a predictor. The model revealed a marginal effect of Utterance Type ($\beta = 44$, $SE = 23.26$, $|t| = 1.86$, $p < 0.1$). The interaction between Relatedness and Utterance Type was significant ($\beta= 11$, $SE = 5.35$, $|t| = 2.07$, $p < 0.05$). The duration of the subject noun phrase, which was included as a predictor in the model, was also significant ($\beta= 0.19$, $SE = 0.03$, $|t| = 6.35$, $p < 0.001$). The effect of Relatedness was not significant ($p < 0.1$).

A planned comparison between the related and unrelated conditions in active and passive utterances was also conducted. Wilcoxon signed-rank tests were used for this purpose (Wilcoxon, 1945)[26]. This analysis revealed that the summed rank of the related condition was higher than the summed rank of the unrelated condition in passive utterances, both in subject averages ($Z = 2.95$, $p < 0.01$) and item averages ($Z = 2.15$, $p < 0.05$). The corresponding effect was not evident in active utterances ($ps > 0.7$).

*Duration analysis*

A mixed effects model with a maximal random effects structure was constructed. In addition to the experimental factors, onset latency was also included as a predictor. The model revealed marginally significant effects of Relatedness ($\beta = 5.74$; $SE = 2.52$, $|t| = 2.28$, $p < 0.05$) and Utterance Type ($\beta = 69.30$, $SE = 12.89$, $|t| = 5.35$, $p < 0.001$). The interaction between Relatedness and Utterance Type was not significant ($p =0.18$).

---

[26] Wilcoxon test was used here to combat normality, which is not achieved by log-transformation itself. Other experiments in this chapter used t-tests because the data in the other experiments was already analyzed for the purpose of a journal manuscript submission, by the time this analysis was conducted.

Planned comparisons between the related and unrelated conditions in active and passive utterances were also conducted. This analysis revealed that the summed ranks of the related condition was higher than the summed ranks of the unrelated condition in active utterances, both in subject averages ($Z = 3.87$, $p < 0.001$) and item averages ($Z = 2.48$, $p < 0.05$). The corresponding effect was not evident in active utterances ($ps > 0.1$).

4.7.3 Discussion

The results suggest that semantically related distractor verbs delayed the onset of subject nouns in passive sentences but not in active sentences. In contrast, semantically related distractor verbs elongated the time spent at the subject noun + auxiliary verbs in active but not in passive sentences. This pattern of results is shown in Figure 4-10. This suggests that verbs are planned before subject nouns are articulated in passive sentences, and during the articulation of subject nouns + auxiliaries in active sentences.

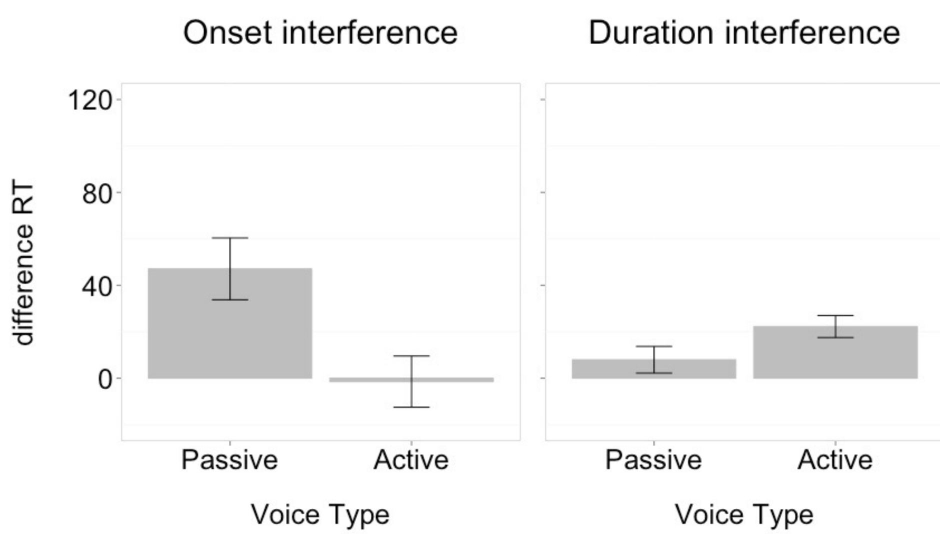



Figure 4-10: Experiment 9, Onset and duration interference (Related-Unrelated) in passive and active conditions: Error bars = Standard errors

One might argue that the current comparison between active and passive sentences is not straightforward, because passive sentences are more difficult to produce, i.e., cognitively taxing, and because the relative position of the verbs is different between active and passive sentences. For instance, there are claims in the literature that cognitive load determines whether upcoming elements are planned in advance (Wagner et al., 2010). Specifically, if there is more cognitive load, speakers cannot afford to plan upcoming parts of a sentence in advance. It is also a standard assumption that the linear position of a word determines whether the word is included in the initial scope of planning. However, notice that these factors should prevent rather than encourage advanced planning of verbs in passive sentences. This is because passive sentences are generally harder, i.e., more cognitively taxing, and the position of the verb is slightly later in passive than active sentences. Nevertheless, the current results suggest that advanced planning of verbs happens selectively in passive sentences. Thus, it is unlikely that the flexible scope account as described in the literature (e.g., Ferreira & Swets, 2002; Wagner et al., 2010) can capture the current pattern of results.

The results of the current experiment, together with Experiments 6-8, suggest that selective advance planning of the verb is likely not due to the case/grammatical function status of the internal arguments, or due to a language-specific strategy that speakers of head-final languages develop. Thus, some contrast between external and internal arguments that holds across languages must be responsible for the timing contrast with respect to verb retrieval.

### *4.8 Experiment 10: English unaccusative and unergative production[27]*

Experiments 6-9 showed that verbs are planned before the internal argument noun's articulation but not the external noun's articulation. However, the SV/OV and Active/Passive contrasts both involve surface differences in terms of syntactic structures. In Experiment 10, I test if the same contrast holds even when the surface structures are identical, using a linguistic contrast in intransitive SV sentences: unaccusatives and unergatives.

Linguistic analyses suggest that there are two types of intransitive verbs: unaccusatives, whose sole argument is a patient or theme (e.g., *fall*), and unergatives, whose sole argument is an agent (e.g., *jump*) [1]. In psycholinguistics, researchers have sought to find how this distinction modulates comprehension (Bever & Sanz, 1997, Friedmann et al., 2008) and production processes (Kegl, 1995; Kim, 2006; Lee & Thompson 2004; 2011; McAllister et al., 2009). However, especially in production, the processing consequences of this distinction are unknown, beyond the suggestion that unaccusatives somehow involve more complex processing than unergatives (cf. Lee & Thompson, 2011).

The unaccusative hypothesis claims that the subject of an unaccusative verb originates as the object of the verb (e.g., Burzio, 1986; Perlmutter, 1978). Supporting this hypothesis, a range of linguistic phenomena, including ne-cliticization and auxiliary selection in Italian (Burzio, 1986), English resultatives (Levin & Rappaport Hovav, 1995), and possessor datives in Hebrew (Borer & Grodzinsky, 1986), suggest that the

---

[27] This section is a lightly edited version of Momma, Slevc & Phillips (under review).

subjects of unaccusative verbs behave like objects. Reflecting this object-like nature of unaccusative subjects, in transformational theories such as Government-Binding Theory (Chomsky, 1981) unaccusative subjects are considered to be base-generated in the object position and move to the subject position (e.g., Burzio, 1986).

Given the above described findings that verbs are planned selectively in advance of internal argument nouns, the psycholinguistically translated version of the unaccusativity hypothesis makes an interesting prediction about the production of intransitive sentences. If unaccusative subjects are logical objects unlike unergative subjects, then unaccusative sentences but not unergative sentences should require the advanced planning of the verb before the subject noun is articulated. If this prediction is borne out, it would show that the subject of unaccusative sentences is processed like a deep/logical object in sentence production, and that split intransitivity directly impacts the time course of speaking. In the current study I used ePWI, just like in Experiment 6-9. In Experiment 10, I specifically examined the timing of verb planning in unaccusative and unergative sentences.

### 4.8.1 Methods

*Participants*

Twenty-four native speakers of English participated for either class credit or monetary compensation. They were all undergraduate students of University of Maryland.

*Materials and Design*

Twenty-four pictures of events were selected. Half corresponded to unergative verbs (e.g., *sleep*), and the rest corresponded to unaccusative verbs (e.g., *float*), as illustrated in Figure 4-14abc. An example picture with the distractor superimposed is shown in Figure 4-14c). The participants of the events corresponding to the unergative verbs were all animate. In contrast, half of the participants of the events corresponding to the unaccusative sentences were inanimate. This imbalance in the number of animate subjects in unergative vs. unaccusative conditions was due to the practical difficulty of drawing a picture in which animate participants undergo the action denoted by certain unaccusative verbs (e.g., *melt*). The 6 animate participants in the unaccusative pictures were exactly matched to the 6 animate participants in the unergative pictures. This identical subset of nouns was used to test whether any difference between unaccusative and unergative conditions could be be solely attributed to the difference in verbs. A full list of target sentences is available in the Appendix.

.

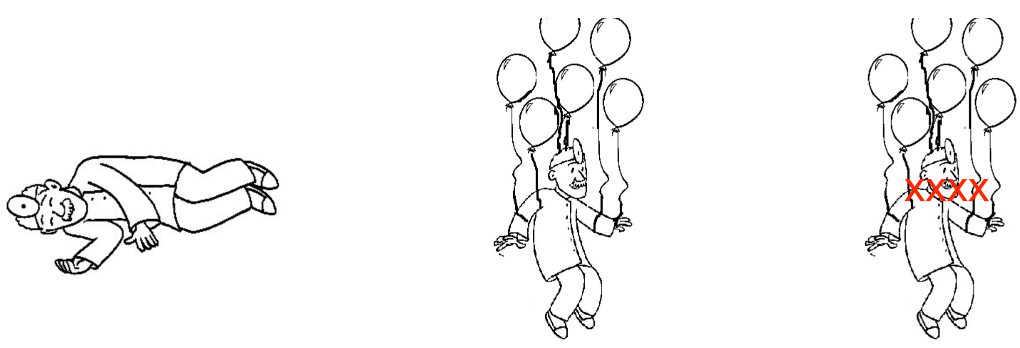


Figure 4-11abc: example pictures for unergative sentences (left; *the doctor is sleeping*) and unaccusative sentences (middle: *the doctor is floating*), and an example picture with a distractor (right: replace xxxx with a relevant distractor word) in Experiment 10.

For each picture, a semantically related distractor verb was selected from the set of target verbs for the other pictures. These distractors always corresponded to one of the other target verbs to maximize the chance of obtaining an interference effect (Roelofs, 1992). Semantic relatedness was estimated based on the cosine distance measure from latent semantic analysis (Landauer & Dumais 1997). This measure reflects how close two words are in a multi-dimensional semantic space in a numerical value ranging from 0 to 1. As a point of reference, the clearly related pair *cat* and *dog* receives a value of 0.36, while *cat* and *desk* receives a value of 0.01. The average cosine distance between each target verb and its related distractor verb was 0.31 for the unergative verbs, and 0.35 for the unaccusative verbs, with no significant difference between them ($p > 0.5$). Each of the related distractor words was re-paired with a picture from the same verb type to create the unrelated picture-distractor pairs. The average cosine distance between each target verb and its unrelated distractor verb was 0.08 for the unergative verbs, and 0.14 for the unaccusative verbs. Unsurprisingly, two-tailed t-tests revealed a significant difference in the cosine distance between related and unrelated pairs both with unergative verb pairs ($p < 0.001$) and with unaccusative verb pairs ($p < 0.001$). Importantly, the mean cosine distance between the related and related pairs differed by 0.23 for unergative verbs and by 0.21 for unaccusative verbs, so the relatedness manipulation was comparable for the two verb types.

In sum, the current study had a 2 x 2 design, with Verb Type (unergative vs. unaccusative) as a between-items factor and Relatedness (related vs. unrelated) as a within-items factor. Both of these factors were within-subjects factors. There were 24 filler trials where distractors were replaced with a string of four x's (i.e., xxxx). In total, there were 72 trials, and each participant saw the same picture three times, once with a related distractor, once with an unrelated distractor, and once as a filler with xxxx. Note that this number of repetitions is many fewer than in some previous picture-word interference studies (e.g., Schriefers et al., 1990).

*Procedure*

Participants first studied a booklet containing the pictures that they were going to see in the experimental session, with the target sentences corresponding to each picture, until they felt comfortable with each picture and sentence. This familiarization session was used in order to increase the accuracy and reaction time stability of their production and it is a standard procedure in picture naming studies (e.g., Schriefers et al., 1998). The experimental session directly followed these familiarization sessions. At the beginning of the experimental session participants were instructed to ignore the written distractor word (in red font) on top of the picture and to describe the picture in sentential form (in present progressive) as soon as they could, except when they saw xxxx as a distractor. When they saw xxxx, they were instructed to not describe the picture and instead press a space key. This step prevented participants from visually ignoring the distractor, thereby ensuring that the distractor words were processed at least to the extent that they could be distinguished from xxxx. Each experimental trial was structured as follows. First, the participant saw a fixation cross at the center of the screen for 750ms. Then, a distractor verb or xxxx appeared at the center of the screen for 500ms in red font. The distractor verb was related to the target verb in 50% of trials with word distractors, and it was unrelated to the target verb in the other 50% of relevant trials. 150ms following the appearance of the distractor a picture from the studied set appeared on the screen for 1500ms. A 2000ms blank screen separated the trials. The speech onset time from the picture onset to the utterance onset, as well as the duration of the subject noun head and the following auxiliary verb *is* were measured manually using Praat. These measures were log-transformed and then submitted to the statistical analysis.

#### 4.8.2 Results

The results are summarized in Table 4-5a (onset) and Table 4-5b (duration) below. A mixed effects model with a maximal random effects structure in the sense of Barr et al (2013) was constructed. For the model of subject noun duration, the number of syllables of the noun was included as a predictor.

| | Mean latency | | Onset interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |
| Unaccusative | 1154 [16] | 1079 [16] | 75 |
| Unergative | 1171 [44] | 1162 [38] | 9 |

Table 4-5a: mean speech onset latency for each condition, with standard error of means in square brackets in Experiment 10. Standard error of means in square brackets.

| | Mean latency | | Duration interference effect |
|---|---|---|---|
| | Related | Unrelated | (Related - Unrelated) |
| Unaccusative | 545 [7] | 549 [8] | -4 |
| Unergative | 580 [10] | 552 [8] | 28 |

Table 4-5b: mean subject noun + auxiliary verb articulation duration for each condition in Experiment 10. Standard error of means in square brackets.

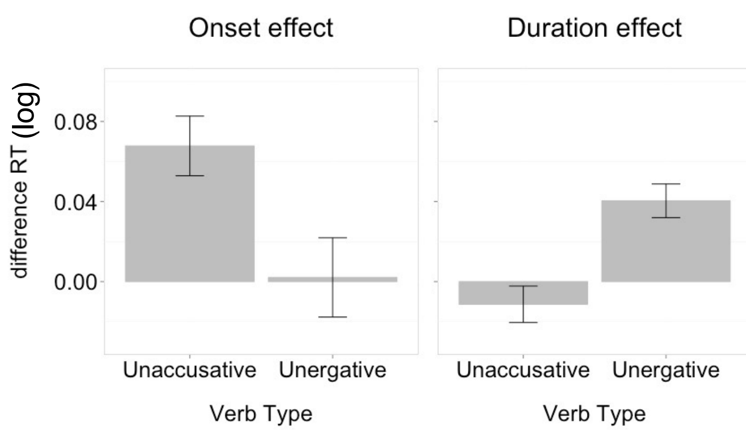



Figure 4-12: the effect of verb associates (Related-Unrelated) on subject NP onset latency (Left) and duration (Right) in Experiment 10.

The model of onset latency revealed a main effect of Relatedness ($\beta = -0.07$, SE = 0.02, $|t| = 3.42$, $p < 0.01$), but no main effect of Verb Type. The interaction between Verb Type and Relatedness was significant ($\beta = 0.06$, SE = 0.03, $|t| = 2.08$, $p < 0.05$). A following planned comparison revealed that distractor relatedness affected onset latency in unaccusative sentences ($t1(23) = 4.56$, adj. $p < 0.001$; $t2(11) = 2.92$, adj. $p < 0.05$) but not in unergative sentences ($t(23) = 0.11$; adj. $> p =0.9$; $t2(11) = 0.71$, adj. $p > 0.9$).

In contrast, the duration measure showed a significant interaction in the opposite direction ($\beta = 0.04$, $SE = 0.02$, $|t| = 2.14$, $p < 0.01$), with no main effect of Verb Type or Relatedness. The number of syllables in the pre-verbal noun was also significantly related to duration ($\beta = 0.12$, $SE = 0.03$, $|t| = 4.20$, $p < 0.001$), but did not interact with Verb Type. A planned comparison revealed that the participants lengthened the utterance of the subject in unergative sentences ($t1(23) = 3.09$, adj. $p < 0.01$; $t2(11) = 3.39$, adj. $p < 0.05$) but not in unaccusative sentences ($t1(23) = -0.6$, adj. $p > 0.9$; $t2(11) = -0.88$, adj. $p > 0.8$).

To ensure that these effects were not due to idiosyncratic differences between items, a secondary analysis examined the subset of 12 items that elicited exactly the same set of animate subject nouns in the unergative and unaccusative conditions. This yielded the same qualitative pattern of results, with a 64 ms onset interference effect for unaccusatives compared to a 21 ms onset effect for unergatives. Similarly, these items showed a greater duration interference effect for unergatives than unaccusatives (23 ms vs. -14 ms respectively).

### 4.8.3 Discussion

Based on evidence from Experiments 6-10 that verbs are planned before uttering a deep object but not before uttering a logical subject, Experiment 11 examined the processing consequences of the unaccusative hypothesis. As predicted, verb interference was found before subject onset in unaccusative sentences, but not during subject articulation in unergative sentences. This suggests that verbs are indeed planned before the utterance of subject nouns in unaccusative sentences, but not during the utterance of subject nouns in unergative sentences. Thus, I conclude that the unaccusative-unergative distinction is realized by the producer, and that this distinction is reflected in the selective advanced planning of verbs in unaccusative sentences.

The selective advanced verb planning in unaccusative sentences is naturally explained if the subject of unaccusative sentences is like an object at some level of representation. This is because advanced verb planning was selectively found before direct object nouns in Japanese sentences and before the subject of English passive sentences, but not before the subject in Japanese sentences or before the subject of English active sentences (Experiments 6-10).

Unaccusativity in sentence production research has been mainly studied in the context of agrammatic aphasia (Kegl 1995, Lee and Thompson 2004, 2011, McAllister, et al. 2009, Thompson 2003). These studies tested whether the increased complexity of the computation/representation involved in unaccusatives as compared to unergative sentences leads to increased difficulty (and hence increased error rates) in producing unaccusative sentences. In general, these results suggest that unaccusative sentences are more difficult than unergative sentences for agrammatic aphasics, both in natural speech (Kegl 1995) and elicited speech (Lee and Thompson 2004), although these results are not entirely consistent across different methodologies of eliciting speech (Lee and Thompson 2011). Notably, however, little work has investigated the production of unaccusative vs.

unergative sentences in normal speakers. One exception is Kim (2006), who studied unaccusative production in neurologically healthy participants and found that unaccusative sentences prime passive sentences. This study suggested that there is some shared representation or processes between unaccusative and passive sentences, which could reflect a shared movement operation in unaccusative and passive sentences. The current study goes beyond these earlier findings in that it tells us specifically how unaccusative and unergative sentences are processed differently in speaking.

Despite the difference in the goals of previous studies and the current studies, it is interesting to note that the majority of the errors that agrammatic aphasics made in Lee and Thompson (2004) in unaccusative sentences (e.g., the ball is bouncing) were sequencing errors (e.g., bouncing the ball). This could be understood as consistent with the current results, suggesting that verbs are planned before the articulation of the subject. From this pattern, I may infer that the lexical planning sequence in sentence production is guided by deep syntactic/semantic dependencies, only after which linearization occurs. On this view, the production deficit in agrammatic aphasics reported in Lee and Thompson (2004) might be attributed to a linearization problem. Under this account, this production deficit surfaces when the required linearization does not correspond to the relative timing of noun vs. verb planning, either due to a deficit in computation involved in linearization, or to deficits in the cognitive mechanisms needed to reliably perform such computations (e.g., working memory).

Also, it should be acknowledged that the current results do not necessarily suggest that verbs must be planned before uttering the unaccusative subject noun. Indeed, some studies suggest that perceptual accessibility modulates the structural choice (e.g., active/ passive choice depends on whether the agent or patient entity is perceptually more accessible) at least when the depicted events are difficult to name (van de Velde, Meyer and Konopka 2014), suggesting that the structural choice of the sentence is determined independently from the verb. This may lead to the claim that verb planning is not necessary to produce passive sentences, and by extension unaccusative sentences. Having said this, however, it should also be noted that Gleitman et al. also showed that passive utterances have generally longer speech onset latency than active utterances even when they are produced as a result of perceptually priming of the patient/theme entity. This suggests that speakers do not simply start articulating the patient nouns just because they are perceptually more available. After the first word of the sentence is determined based on its perceptual availability, some additional computations are performed before it is articulated. Such computations might well involve the advanced planning of the verb to encode some syntactic/semantic relation for their internal argument, as proposed in this article. Thus, the existing evidence is not incompatible with the strong interpretation of the current results, and until proven otherwise, I maintain the strong, more readily falsifiable, hypothesis that verbs must be planned before internal arguments can be uttered, at least when production contains no repair process (i.e., proceeds normally).

A question remains: why is the verb selectively planned before uttering a deep/ logical object? One reasonable explanation is that some computation needed to encode deep/logical object nouns depends on the lexical representation of the verb. Given that Experiment 6-10 found advanced verb planning before canonical objects in Japanese, this is unlikely to be due to computations involved in establishing a non-canonical mapping between thematic roles and grammatical functions, i.e., this is not due to movement operations per se. Certainly, this explanation based on non-canonical thematic-syntactic mapping cannot be convincingly ruled out, as it is not entirely clear whether it is fair to compare the different sentence structures in two different languages. However, under the assumption that the core property of the processing mechanism is universal, this explanation is disfavored. Also, given that advanced verb planning was found in unaccusative sentences in the current study and in English passive sentences (Momma et al. 2015b), the current effect is not likely due to computations related to accusative case assignment. The remaining candidates include (i) phrase structure building for the deep/object position and/or (ii) assignment of internal argument roles. Both candidates have to do with the role of the argument structure of verbs in sentence production. The first possibility relates to the claim that the phrase structure rules for VPs are dependent on the lexical properties of the verb (i.e., subcategorization) that are not deducible from the conceptual representation alone (Grimshaw 1990). Under this view, the conceptual representational system is sufficient to encode external arguments, but the lexical representation of verbs are necessary for encoding their internal arguments. The second possibility is based on the linguistic analysis of argument-predicate relationships by Kratzer (1996), in which she argued that only the internal arguments are true arguments of verbs. Under this view, the assignment of argument roles to the object might selectively require selecting a specific verb, while the assignment of agent argument roles to an external argument might be done independently from the verb head. In other words, the processing of the external argument is conceptually guided. Future studies should aim to distinguish between these possibilities, and this line of research will inform how theories of argument structure relate to the theories of sentence production.

### *4.9 General discussion*

Collectively, the current experiments showed that the timing of verb planning is such that verb planning precedes the articulation of internal, but not external arguments. This was obtained regardless of language (English/Japanese) or construction (active/ passive, unergative/unaccusative). Furthermore, Experiment 10 might even suggest that the timing of planning might precede the lexical planning of the internal argument in passive utterances. All of these suggest that, when it comes to verb planning, the linear position in a sentence is a poor predictor of whether verbs are included in the initial scope of planning.

I have explored several possible causes of the contrast between internal and external arguments in terms of the timing of verb planning. One of them was about the case and grammatical function contrasts, but the explanation was eliminated via Experiments 9-11, where the advanced verb planning effect was obtained in passive and unaccusative sentences. Furthermore, Experiment 10 found that the verb interference effect was not obtained when verbs were not included in a linguistic expression, even when similar conceptual processing of the picture was required by the task. This suggests that the semantic interference effect does not just reflect advanced planning at a non-linguistic, conceptual level.

One remaining possible explanation is a contrast between external and internal arguments regarding their thematic dependency on verbs. This is in accordance with Kratzer's (1996) argument that external but not internal arguments should be severed from the argument structures of verbs in syntax. That is, in her framework there is no such thing as theme/patient, and the identity of the predicate is required to define the thematic role that an internal argument takes. More specifically, in semantics, there are two major ways to associate predicates and their arguments: the traditional ordered argument method and the neo-Davidsonian method. The logical form representation of the verb *purchase* in each method is indicated in (4-1a) and (4-1b) respectively (taken from Kratzer, 2002).

(4-1) a. λxλyλe [purchase(x)(y)(e)]

'e is a purchase of x by y'

(4-1) b. λxλyλe[purchase(e) & theme(x)(e) & agent(y)(e)]

'e is a purchase with theme x and agent y.'

Only in the neo-Devidsonian representation, as in (4-1b), does the semantically primitive predicate corresponding to thematic roles like *theme* exist independently of the verb root. Kratzer (2002) specifically argued that internal arguments, but not external arguments, can be associated with verbs in the neo-Davidsonian argument association method. That is, Kratzer (2002) took an approach where both the ordered argument method and the neo-Davidsonian method are used to represent verb meanings, as in (4-1c) below.

(4-1) c. $\lambda x \lambda y \lambda e$[purchase(e, x) & agent(y)(e)]

'e is a purchase of x with agent y.'

In this representation, external but not internal arguments are represented independently of the verb root. If this hypothesis about the predicate representation is transparently reflected in sentence production, then speakers might have to have the verb root ready before assigning internal, but not external, argument roles. This nicely explains the current pattern of results that suggest that verbs are planned selectively before the articulation of internal argument nouns.

Another remaining possibility has to do with the contrast in the constituency relation between internal arguments and verbs and between external arguments and verbs. In a transitive sentence, the internal argument and the verb form a constituency relationship while the external argument and the verb by themselves do not. The current pattern may be explained if lexical planning is controlled by syntactic representation, as claimed in Smith & Wheeldon (1999) and in the syntactic gating account. I offer a specific version of this possibility, utilizing the mechanisms and findings that I have introduced throughout this thesis. First, in Chapter 3, Experiment 3, I showed that the creation of syntactic structure does not occur extensively before the subject articulation. In particular, I showed that the internal structure of the VP is not yet planned before the subject noun is articulated. This means that the verb phrase structure may not be available before the articulation of the external argument. Second, also in Experiment 3 I showed that the lexical retrieval of a VP internal argument is preceded by structural processes associated with the VP. This indicates the involvement of top-down structure building in the construction of VPs in production. If structure building for VPs generally involves top-down procedures,[28] the existence of a complement entails the existence of the verb position, as the first node counting form the top of the phrase is V. Third, in Chapter 3, Experiments 2-3, I argued for the existence of syntactic control over lexical retrieval (syntactic gating). When these three claims - about the size of structural units, the involvement of top-down structure building, and syntactic control over lexical retrieval - are combined, the following inference is possible. Planning internal arguments entails the construction of a verb phrase, and hence the structural position for the verb should be available before or as soon as internal arguments become available. This structural position for verbs is required for the verb to be retrieved while maintaining connectedness (i.e., incremental structure building). On the other hand, planning external arguments does not entail the construction of the verb phrase (as was shown in Experiment 5), and thus the structural position for the verb is not available by the earliest time that the external argument can be uttered. Since the structural position for the verb is not yet available, the verb is not guaranteed to be integratabtle to the existing structure. In other words, retrieving a verb before external arguments may violate incremental structure building defined in terms of connectedness. This explanation is of course not

[28] Admittedly, this might not be hold, because top-down structure building might be selectively triggered by verbs, but not by complements. Thus, it is an empirical question whether the precedence of syntactic processes over lexical processes shown in Experiment 3 transfers to utterances where verbs come later than their complements.

mutually incompatible with the first possible explanation in terms of a thematic dependency contrast between internal and external arguments. Indeed, since thematic roles and (deep) structural positions are not dissociable in some theories of syntax (e.g., Uniformity of Theta Alignment Hypothesis, Baker, 1988).

### *4.10 Implications for the single system view*

As argued in Chapter 1, the single system view demands that the notion of incrementality and how memory costs are evaluated should align between parsing and generation models. However, as I hinted in Section 4.1 and as I will elaborate more in Chapter 5, I argue that the notions of incrementality are rather distinct between the existing parsing and generation models. Incrementality is about reducing memory costs, so different notions of incrementality point to different cost notions. This misalignment poses a challenge for the single system view.

As I will see in more detail in Chapter 5, the notion of incrementality is defined in rather different ways in the parsing and generation literature. In parsing, incrementality is often defined in terms of what is called *connectedness* (e.g., Sturt & Lombardo, 2005). The idea is that the incremental structure builder analyzes each piece of input as it is received, and integrates it to a larger coherent structure without delay. This is called connectedness because the structure builder strives to maintain a connected memory representation of the sentence structure throughout the course of parsing. This strategy reduces the memory cost in parsing because, when successful, no more than one structural representation occupies the working memory space. If there is only one integrated item in memory, neither a capacity-based model of working memory nor an interference-based model of working memory predicts a substantial memory cost.

In generation, incrementality is often defined as what can be called *cascading incrementality* (e.g., DeSmedt, 1990; 1996). The idea is that the information at higher levels of representation cascades down to the next level of processing without delay (Levelt, 1989). Similar ideas are sometimes called *Wundt's principle* (Levelt, 1989), or the *principle of immediate mention* (Ferreira & Dell, 2010). In its strong version, concepts that come early in sentence planning are uttered early in a sentence, i.e., planning and articulation synchronize as much as possible. Cascading incrementality reduces memory costs because it allows speakers to 'release' a planned item without delay by articulating it, thereby preventing memory interference and memory overload. That is, cascading incrementality reduces memory cost not by integration, but by articulation.

The current findings that verbs are planned in advance of internal arguments are not easily compatible with a strong version of cascading incrementality, but they are compatible with a notion of incrementality defined in terms of connectedness. The central principle following from the cascading incrementality view is the synchronization between planning and articulation, and advance planning processes clearly violate this. Indeed, any evidence for the involvement of substantial advance planning in sentence production (e.g., Meyer, 1996; Smith & Wheeldon, 1999) is a violation of such a principle, and some researchers (e.g., F. Ferreira, 2000) take such evidence to be evidence against the strong version of incrementality, defined in terms of cascading.

On the other hand, the evidence for substantial advance planning is not in conflict with the view that the structure builder strives to maintain a connected structure throughout the course of processing. As long as the planned elements of a sentence are integrated into a larger structure immediately, generation can be conceived of as strongly incremental just like in incremental parsing. In the current case, the verb of a sentence is planned in advance before the articulation of internal arguments, creating a de-synchronization between planning and articulation. However, since the verb phrase presumably needs to be established before the pre-verbal internal arguments can be uttered, the planned verbs can be integrated into the structure as soon as they are retrieved. This is the sense in which advance verb planning can be entirely compatible with the strong version of incrementality, once the notion of incrementality in generation is aligned to parsing in terms of connectedness.

Indeed, the notion of incrementality defined in terms of connectedness is guaranteed to be satisfied if the syntactic gating mechanism is introduced in sentence production. This is because the syntactic gating mechanism ensures that memory retrieval operations do not initiate until there is a structural position available. On the other hand, the syntactic gating mechanism is not easily compatible with the notion of cascading incrementality. This is because the syntactic gating mechanism prevents synchronization between the conceptual preparation and articulation via syntactic representation. Thus, the syntactic gating mechanism is compatible with the notion of incrementality defined in terms of connectedness, but not the notion of cascading.

As there is a lot of evidence suggesting that speakers can (but need not) plan upcoming parts of a sentence in advance (Meyer, 1996; Garrett, 1980; Smith & Wheeldon, 1999; Konopka, 2010; Wagner et al., 2010; Schnur, 2002; among others), I argue that the cascading incrementality view has limited success in accounting for the majority of speech planning data. Instead, I argue that importing the notion of connectedness from parsing models is more successful in accounting for such data, and that the syntactic gating mechanism is the key mechanism that makes it possible for the same notion of incrementality to be plausibly satisfied in generation. More detailed discussion on this point can be found in Chapter 5.

# 5 Aligning parsing and generation

In Chapter 5, I revisit the five challenges for the single system view that I outlined in Chapter 1. For readers' convenience, they are repeated here: (i) the potential mismatch in the grain size of structural processing; (ii) the difference in the nature of the input; (iii) the potential mismatch in whether structure building operates serially or in parallel; (iv) the potential mismatch in the notion of incrementality and how memory costs are evaluated in parsing and generation models; and finally (v) the potential mismatch in the order of structure building. Note that these challenges were not chosen randomly; I chose these five aspects based on major issues widely discussed in parsing research. The experiments that I presented in the previous chapters clearly do not address all aspects of those five challenges. This is expected, because the single system view encompasses a collection of multiple hypotheses, so no individual experiment can test all aspects of it. As a result, some of the challenges remain unaddressed in the previous chapters, which focused more on presenting new experimental evidence than on presenting theoretical arguments. In this chapter, I thus aim to address the five challenges by synthesizing the experimental findings in this thesis and previous psycholinguistic research that I take to be relevant to the single system view. I then conclude by presenting a preliminary single system model of structure building.

## *5.1 Grain size of structural processing*

The single system view suggests that the size of the pre-compiled structural units in parsing and generation should be aligned. This section offers a discussion about previous and current experiments to assess whether this claim can plausibly be adopted as a tenable hypothesis.

In general, models of parsing tend to assume that the grain size of structural units is small. This reflects the widely held assumption that comprehenders interpret partial sentences without complete input, and update their analyses at every word. This is sometimes referred to as incremental interpretation or incremental comprehension, though here I prefer not to use the term incrementality because I use the term incrementality to specifically mean incremental structure building defined in terms of connectedness. Here, I will use the term *immediate interpretation* to refer to the idea that comprehenders immediately interpret partial input before the complete input arrives. The evidence for immediate interpretation is extensive, ranging from self-paced reading, ERPs, and eye-tracking. Under the assumption that (compositional) interpretation is parasitic upon structural parses (Stroud & Phillips, 2012; cf. Ferreira & Patson, 2007; Townsend & Bever, 2001), immediate interpretation suggests that structural and post-structural processes interleave frequently, possibly at every word or every phrase, and definitely before a whole clausal input arrives. This frequent interleaving between structural and post-structural processes suggests that the grain size of structural memory units in parsing is smaller than a complete clause, even in a language in which syntactically important parts of the sentence are positioned at the end of a clause.

In comparison to many parsing models that assume relatively small structural units, some early as well as recent models of sentence production assume that the grain size encompasses a clause (e.g., Bock & Cutting, 1992; Garrett, 1980, 1988; Ford, 1982;

Ford & Holmes, 1978; Griffin & Weinstein-Tull, 2003). If these models are correct, there is a mismatch between parsing and generation in terms of the size of structural units, and hence a problem for the single system view.

One way to proceed under the single system view is to adopt other production models like Kempen & Hoenkamp (1986), and V. Ferriera (1996), which assume more fine grained units of structural processing. In these models, words can be uttered as they are planned, without the clausal structure being specified. Thus, just like in incremental parsing models, the structural and post-structural processes (in this case articulation) interleave frequently, and the size of structural units is usually assumed to correspond to a unit much smaller than a clause, often a single word (Levelt, 1989; Kempen & Hoenkamp, 1986; De Smedt, 1990; 1996). Such models are often described as incremental models of sentence production. Although I argue that incremental generation models are not incremental in the same sense as incremental parsing models (see Section 5.4), I share with these models the view that the size of structural processing units in generation is small. Under such a position, the size of the structural units that the parser and the generator use can plausibly be aligned.

The view that the size of structural units is small in generation should not be confused with what is called the *linearly incremental* view of sentence production. Linear incrementality is the view in which no advanced planning beyond the immediate constituent structure needs to be conducted in advance of articulation (Bock, Irwin & Levelt, 2004). Although the linearly incremental view of sentence production does entail that the unit of structure building is small in generation, the view that the unit of structure building in generation is small does not automatically lead to the claim that sentence production is linearly incremental. This position can be taken as in conflict with the experimental evidence for advance verb planning presented in Chapter 4. If assuming a small grain size of structural processing for generation equates to adopting the linear incrementality view, I am presenting a internally incoherent argument in this thesis. I thus need to disentangle the position that the structural units in generation are small and the position that sentence production is linearly incremental.

An oft-cited piece of evidence for the linear incrementality view is Gleitman et al. (2007). In one of their experiments, Gleitman et al. showed that subliminally cuing the theme or patient entity in a picture stimulus increases the likelihood of passive utterances, suggesting that implicit perceptual attention influences structural choice. This is sometimes interpreted to mean that sentence production is linearly incremental, and thus involves little or no planning beyond the first content word of a sentence (Bock & Ferreira, 2014; Kuchinsky, Bock & Irwin, 2011; Van de Velde et al., 2014; Van de Velde, Kempen, & Harbusch, 2015). Kuchinsky, Bock & Irwin (2011) also replicated the effect, but only when the event depicted in the picture was hard to name and when one of the participants of the event was comparably easy to name. They argued that sentence production can be linearly incremental, though it need not be. This kind of evidence for perceptual cuing affecting structural choice, however, has little to do with whether the later parts of an utterance are planned before articulation onset. As long as perceptual attention somehow determines the starting point of the sentence, which in turn determines the following course of generation processes, the study like Gleitman et al. (2007) can be captured without assuming that sentence production is linearly incremental. For instance, it suffices to say that perceptual cuing affects which entity in the picture is assigned the highest structural function (i.e. the subject) in a sentence. Whether additional planning beyond the first constituent is conducted before speech onset is a completely independent question. Indeed, Gleitman et al.'s speech onset latency data showed that passive sentence utterances are initiated later than active sentences, even though speakers were not forced to produce passive sentences. This is unexpected under the linear incrementality view, and suggests that sentence parts beyond the first noun phrase could be processed well in advance of the articulation onset. That is, the fact that speakers choose to utter a perceptually attended item first does not constitute evidence that they do not plan beyond the first item. Hence, although I do argue that the structural units are small in generation as well as in parsing, I do not argue that sentence production is linearly incremental. This is because substantial planning still can happen before articulation onset.

Relatedly, it is worth noting that the scope of advance planning is often taken to be a direct reflection of the size of structural units in generation. The scope of advance planning refers to the extent to which speakers plan sentence parts before they initiate articulation. Studies on the scope of advance planing have yielded seemingly inconsistent results, some suggesting that the scope is clausal (Meyer, 1996), others suggesting that the scope is phrasal (Smith & Wheeldon, 1999), and yet others suggesting that the scope is sub-phrasal, i.e., a single word (Griffin, 2001). These studies are taken to be evidence that the size of processing units in production is clausal, phrasal, or lexical, respectively, and led to a claim that the scope of advance planning is flexible (Wagner et al., 2010; Konopka, 2010). Such an interpretation may be taken as in conflict with the current argument that the grain size of structural units is small, a claim that presupposes that the structural unit has a fixed size. However, evidence that speakers plan substantial portions of a sentence in advance of the onset of articulation does not entail that the grain size of structural units is large. This is because speakers do not have to start articulation immediately, even when they plan a sentence small bit by small bit. The decision to start articulating is likely to be determined on the basis of task demands and time pressure. Intuitively, speakers can plan an entire sentence before they start articulating, as in careful speech. Experimentally, some evidence indeed suggests that the decision to start articulation depends on task demands and experimental context (Ferreira & Swets, 2002; Wagner et al., 2010; Konopka, 2010). Importantly, however, such evidence for flexibility in when to start articulating (Ferreira & Swets, 2002; Wagner et al., 2010; Konopka, 2010) does not constitute evidence that the grain size of structural units is also flexible. Structure building processes can still take place small bit by small bit, and speakers can simply decide not to interleave structure building with articulation even when they, in principle, could do this. Thus, the evidence for a wide scope of advance planning is not inconsistent with the claim that the grain size of structural units is small. For the same reason, evidence that the scope of advance planning is flexible does not justify the claim that the size of structural units is flexible. The temporal interval between planning and articulation is only an indirect measure of the size of long-term stored structural units.

Of course, I do not claim that the advance planning experiments have no implications for the size of structural units. In particular, evidence for a narrow scope of advanced planning does suggest that the size of structural units is small. If it is assumed that articulation must be a post-structural process, as most if not all models of sentence production assume, the narrow scope of advanced planning then indicates that structural and post-structural processes can be interleaved within a certain window. This means that the structural unit must be small enough to allow such interleaving. Again, this does not mean that speakers must interleave planning and articulation. Thus, evidence that articulation can occur before substantial planning of structure does provide evidence that the grain size of structural units is small. For instance, in the simple sentence *John gave her flowers*, if one finds evidence that *John* can be uttered before the speaker decides whether to use the double object or prepositional object structure, it can be safely inferred that the size of the structural units is smaller than a clause, not including the internal structure of the verb phrase.

In Chapter 2, I presented one experiment that suggests that the internal structure of a verb phrase in ditransitive sentences is not specified in advance of the utterance onset. Indeed, the data may even suggest that the structure of verb phrases is not specified until after the verb itself starts to be uttered. This may suggest that the structural units in generation are even smaller than a phrase. This is unexpected under the view that the grain size of structural units is equal to or larger than a clause, and hence it supports generation models in which the size of structural units is assumed to be minimal. As the supported generation model is one that assumes small structural units, just like in widely accepted parsing models, this evidence increases the plausibility of aligning the size of structural units between parsing and generation.

### *5.2 Input difference*

As discussed in Chapter 1, a primary motivation to think that the parser and the generator are distinct mechanisms is that they take distinct kinds of input. To evaluate the validity of this inference, I first discuss the logic of this inference, and then review the relevant data points from the comprehension and production literatures.

As mentioned already, it is an undeniable fact that the starting point of comprehension and that of production are different. The starting point for comprehension is an auditory/visual signal, while the starting point for production is a message. Does this fact by itself qualify as evidence that the parser and the generator are two different mechanisms? Consider the lexical memory system. It clearly can take conceptual input, as in picture naming, and it clearly can take auditory and visual input, as in reading and listening. The system outputs the same representation: lexical representations. Does this show that there are two lexical systems, one specialized for conceptual input and another specialized for sound input? I suspect that the default response for most researchers is no. Indeed, some models of lexical retrieval like the TRACE model (McClelland & Elman, 1985) assume a single mechanism that can work on both conceptual and sound input. In this approach, memory retrievals rely on associative cues, and associations between memory items and external/internal cues can be arbitrary in type. This associative nature of the memory system seems to be universal across different memory systems. For instance, in retrieving a memory of a past event, virtually any kind of associative cues can be used to find the target memory. This includes visual, auditory, olfactory, and tactile cues (among many other arbitrary cue types) that happen to be associated with that memory. Of course, the effectiveness of each cue type may differ, and there may be biases that make some kinds of cues more effective than others. However, there does not seem to be any principled restriction that prohibits the formation of an association between a certain cue and a certain memory item. Despite the fact that the memory of a real world event can be retrieved via arbitrarily diverse kinds of cues, it is not fair to conclude that multiple memory systems of past events exist, one for each kind of cue. Thus, it is not a radical assumption that the linguistic memory system also exhibits this associative nature.

Indeed, in a task that requires the simultaneous processing of auditory and conceptual inputs (often presented in the form of picture), lexical recognition and lexical production performances are affected by both. In production, this can be seen, for example, in the picture word interference task, where conceptual information, sound information, and the relationship between them affects the naming latency (e.g., Schriefers et al., 1990). Such evidence that auditory information can affect the performance on production tasks in concert with conceptual information suggests that the lexical retrieval mechanism uses both sound and conceptual information in an individual act of producing a word. In this sense, the retrieval mechanism used in picture naming can interface with both sound and conceptual inputs if both are present in the environment. The same argument can be made for lexical recognition, although I am not aware of any evidence that simultaneous presentation of a conceptually relevant picture affects the speed of lexical recognition. Thus, to the extent that sound inputs affect production performance, and to the extent that conceptual inputs affect recognition performance, the retrieval mechanism can be a single mechanism that deals with both sound and conceptual inputs when they are available in the environment. So, as far as the retrieval mechanism is concerned, the difference in the kinds of inputs during comprehension and production tasks does not offer a convincing argument against the single system view.

The previous paragraph argued that, as far as the lexical mechanism is concerned, the fact that there are multiple types of input does not constitute evidence that there are multiple distinct retrieval mechanisms subserving parsing and generation. Now, the remaining question is whether this argument transfers to parsing and generation as a whole. If the parser and the generator use conceptual and sound information differentially, it poses a challenge to the single system view. This is because the single system view predicts that both conceptual and sound information can be efficiently used during parsing and generation alike, when such information is available in the environment.

Below, I will review the relevant literature in parsing and generation to assess whether they differentially rely on message and sound information during parsing and generation. I will argue that the parser can efficiently use conceptual information when available, although it remains unclear whether the generator can use phonological information as efficiently as the parser. As it is standardly assumed, I assume that the generator can use conceptual information efficiently. Thus, for conceptual information, I only discuss the relevant studies in parsing literature.

*Conceptual information in parsing*

Conceptual representations, broadly defined, encode both discourse and semantic information. This can include, for instance, information structure (topic, focus, contrast, etc.), thematic relations (agent, theme, etc.), and more broadly situation/mental models (cf. van Dijk & Kintsch, 1983; Johnson-Laird, 1983). There is a considerable body of evidence that conceptual information rapidly affects parsing decisions (e.g., Altmann, van Nice, Garnham & Henstra, 1998; Altmann, Garnham & Dennis, 1992; Perlmutter & MacDonald, 1992; Trueswell et al., 1994). To begin, consider the following examples, taken from Clifton & Ferreira (1986).

(5-1) a. *The defendant examined by the lawyer turned out to be unreliable.*

(5-1) b. *The defendant that was examined by the lawyer turned out to be unreliable.*

(5-1) c. *The evidence examined by the lawyer turned out to be unreliable.*

(5-1) d. *The evidence that was examined by the lawyer turned out to be unreliable.*

Although both of these sentences are locally ambiguous, the amount of disruption at the disambiguation region (*by the lawyer*) was reported to be smaller in (5-1d) than in the (5-1a) , possibly to the extent that the reanalysis cost is non-existent in (5-1d) (Trueswell, 1994; but see Clifton et al., 2003 for evidence otherwise). Specifically, Trueswell, Tannenhaus & Garnsey (1994), showed that the the ambiguous (5-1c) and unambiguous

(5-1d) sentences show little/no difference in processing difficulty. In contrast, they showed that the ambiguous sentence (5-1a) is clearly harder to process than unambiguous sentence (5-1b). Presumably, this difference is due to the fact that the subject noun is inanimate, and thus the likely patient of the verb *examined*, suggesting that *examined* is the verb of the reduced relative clause. This suggests that the plausibility rapidly constrains the initial structure building, such that the main clause analysis of *examined* is not considered.

Additionally, consider the following sentence from Tueswell et al. (1999)

(5-4) *Put the frog on the napkin into the box.*

Trueswell and colleagues showed that adult comprehenders look at a picture of a frog on the napkin over the picture of a frog not on the napkin upon hearing *on the napkin*, before reaching *into the box.* This indicates that adult comprehenders analyze the first prepositional phrase as a modifier of the frog when there is more than one frog in the visual scene. This is unexpected under the view that discourse information does not affect initial parsing decisions, because the syntactically simplest analysis is to analyze the first prepositional phrase as the argument of *put* rather than as the adjunct of *the frog*. These pieces of evidence at least suggest that discourse information is used efficiently in parsing.

*Prosodic representation*

Unlike conceptual representations, prosodic representations are a property of the end product, not the starting point, of the production system. As a result, it is unclear how to create a situation where prosodic information is available before the relevant generation operations start. Thus, I cannot offer any experimental evidence that the generator can use prosodic information efficiently when it is available in the environment. However, to the extent that prosodic information is used efficiently by the parser, the single system view predicts that the generator should be able to use prosodic information efficiently as well. Below I will review the evidence that the parser can use prosodic information efficiently.

There is a large body of research investigating whether prosodic information rapidly affects syntactic decisions (e.g., Marslen-Wilson, 1992; Snedecker & Trueswell, 2003; Nakamura, Arai & Mazuka, 2012; Pynte & Prieur, 1996; Kjelgaard & Speer, 1999; Steinhauer et al., 1999). An oft-cited study is Snedecker & Trueswell (2003). They examined the processing of ambiguous sentences like (5-5), given either of the visual context with two possible referents (Fig. 5-1a) or only one possible referent (Fig. 5-1b).

(5-5) *Tap the frog with the flower.*

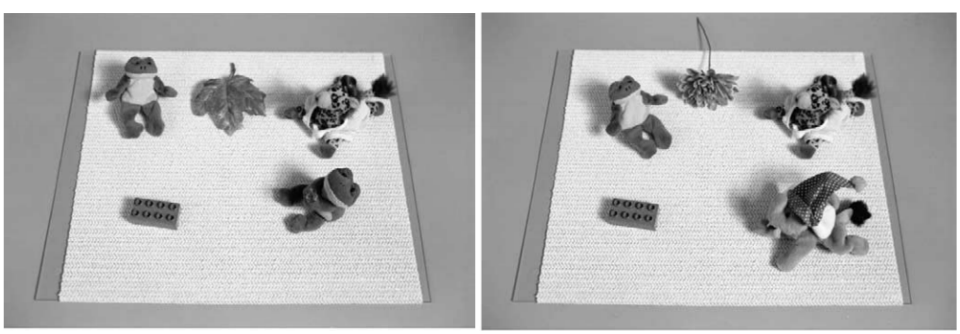


Figure 5-1ab: The visual stimuli used in Snedeker & Trueswell (2003).

They showed that comprehenders fixated more on the frog with flower given Fig. 5-6a with contrastive prosody than without contrastive prosody. Importantly, this effect was obtained even before the preposition was heard by participants, suggesting that they predicted the prepositional modification of the noun phrase even before the preposition was heard. They took this as evidence that prosodic information rapidly affects interpretation, and therefore, parsing.

In addition, Nakamura et al. (2012) showed a similar effect of prosody on parsing in Japanese, examining the processing of the following temporarily ambiguous sentence as in (5-6).

(5-6) *otokonoko-ga sanrinsha-ni notteita onnanoko-o mitsumeta.*

*boy-NOM tricycle-DAT was-riding girl-ACC looked at*

*'the boy looked at the girl who was riding on the tricycle'*

In Japanese, the first verb was-riding is temporally ambiguous between being the main clause verb and being the verb of a relative clause. Comprehenders are expected to parse it as the main verb according to the economy principle. Nakamura et al. reported that comprehenders, in the presence of contrastive prosody, moved their eyes toward the girl on the tricycle when there was another girl (who was not on the tricycle) before the relative clause head, suggesting that they predicted the relative clause is upcoming. This was not found when the visual stimuli did not contain two potential referents. This evidence suggests that, even before the disambiguating information appears in the input, prosody affects structural choice. Based on this, they argued that prosodic information affects *predictive* parsing, and hence the earliest moment of syntactic processing.

Both Trueswell et al. (2003) and Nakamura et al. (2012) showed that prosodic information affected parsing even before there is any ambiguity in the input. Although I am not aware of any evidence for or against the efficient use of prosodic information by the generator, the single system view predict that it should be used efficiently when available in the environment.

### *5.3 Seriality and parallelism*

In parsing, both the psycholinguistics and computer science literature identify two classes of structure building algorithms: serial and parallel. Serial structure building algorithms pursue only one structural analysis in parsing a sentence. Parallel structure building algorithms pursue multiple structural analyses. This serial/parallel distinction is also known as the distinction between depth-first and breadth-first parsing. If the parser and the generator are identical systems, it is expected that the parser and the generator converge in this respect. In this section, I will review the relevant evidence in parsing and generation on this serial/parallel distinction.

Before I begin, however, one might question whether it is necessary for the notion of parallel or serial structure building to be even relevant in generation. This is because, after all, there is no ‘ambiguity’ in generation (except in the speakers’ own already produced utterances). The serial/parallel distinction in parsing arises because there is structural ambiguity. However, I argue that this distinction is equally relevant in generation, because there are structural choice points in the course of sentence production, and ambiguous parts of a sentence are essentially structural choice points. In parsing, ambiguity occurs when the single input representation (sound) can be mapped onto more than one output representation (message). In generation, there are situations where the same input representation (message) can be mapped onto more than one output representation (sound). For instance, mostly the same message can be conveyed by double object dative sentences (e.g., *Bill gave him an apple*) or by prepositional object sentences (e.g., *Bill gave an apple to him*). That is, there is one-to-many correspondence between input and output, mediated by structural representations, both in parsing and generation (Fig. 5-2ab).

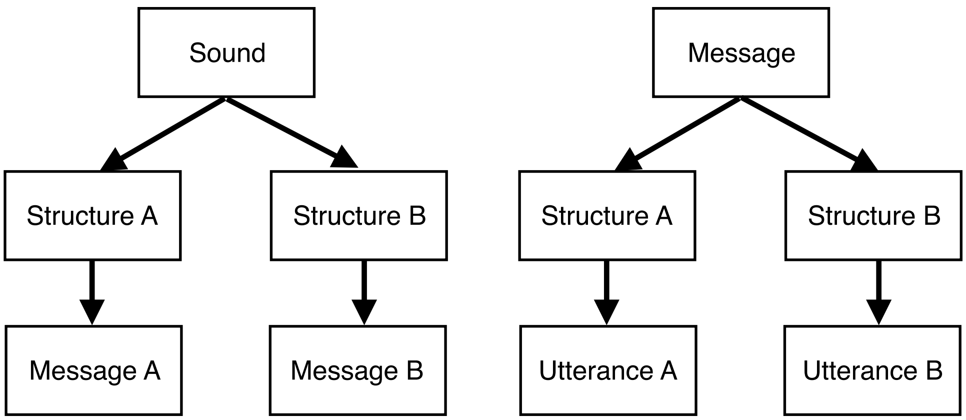



Figure 5-2ab: One-to-many input-output mapping in parsing (a) and generation (b), both mediated by structural representations.

Having argued that the question about whether the structure is built in a serial or in a parallel fashion can be relevant to both parsing and generation, I will first discuss the parsing literature.

The serial/parallel distinction in parsing can be illustrated using a classic garden-path sentence (Bever, 1970) as in (5-7ab)

(5-7) a. *The horse raced past the barn fell.*

(5-7) b. *The horse that was raced past the barn fell.*

The sentence (5-7a) is locally ambiguous unlike (5-7b), because the verb of the embedded clause *raced* could be analyzed either as the active form of the main verb that takes the horse as its subject (main clause analysis), or the passive form of the verb in a reduced relative clause that takes the horse as its semantic object (reduced relative clause analysis).

Serial and parallel models differ in how the sentence (5-7a) is analyzed, in at least two ways. First, at the ambiguous region *raced*, serial and parallel models differ in whether both the main clause analysis and the reduced relative analysis are considered. Serial models assume that only one of them is considered (based on some decision criteria), while parallel models assume that both of them are considered. In parallel models, it is often assumed that the multiple structures are ranked in some fashion, based on certain criteria, such as frequency of the structure. Under normal assumptions, both models prefer the main clause analysis, because it is simpler/more frequent, etc. Second, at the disambiguation region *fell*, they differ in how they deal with the incorrect commitment that they made. Serial models assume that the parser backtracks to a certain point and then recomputes an alternative analysis. Parallel models assume that the parser reranks the structural alternatives so that the analysis that is compatible with the disambiguation wins.

At the disambiguation region, both serial and parallel models predict a reading disruption, according to the numerous empirical findings reporting the disruption at the disambiguation for the unfavored analysis (e.g., Bever, 1970 among many others). Thus, with the proper assumptions, such disruption effects can be explained by either model and hence carry little force in distinguishing the types of models (though see Elman, 2004 for a modeling effect). Thus, as pointed out by Staub & Clifton (2008) the critical point is *before* the disambiguation region. Serial models predict that ambiguity should not be costly until the disambiguation region. In contrast, parallel models predict some cost of ambiguity before the disambiguation region, because carrying two competing analyses should be costly. As argued by Staub & Clifton (2008), there is little to no evidence (cf. Pickering and Traxler, 1998) for processing difficulty before the disambiguation region, providing no evidence for competitive models. However, this is evidence from a null effect and thus should not be taken as the positive evidence for serial models.

However, one potential positive piece of evidence for the serial nature of structure building comes from Van Gompel, Pickering & Traxler (2001, see also, Traxer, Pickering & Clifton, 2000). They compared the following types of sentences, using eye-tracking measures.

(5-8) a. *The hunter killed only the poacher with the rifle not long after sunset.*

(5-8) b. *The hunter killed only the leopard with the rifle not long after sunset.*

(5-8) c. *The hunter killed only the leopard with the scars not long after sunset*

(5-8a) is globally ambiguous, because the prepositional adjunct can plausibly be attached to either the VP *killed only the poacher* or the NP *the poacher*. In comparison, (5-8b) and (5-8c) are allowed only one attachment (either VP or NP attachment) due to a plausibility constraint. The surprising finding is that the critical word *rifle* in ambiguous sentences was read more rapidly compared to the counterpart in unambiguous sentences. This is unexpected for the models assuming competition between multiple structures (but see

Van Gompel et al.'s for more detailed discussion), because the globally ambiguous sentence (5-8a) should create the maximal amount of competition, and therefore produce slow-down effects. In other words, processing is easier when structural choices are available. This counts as evidence that multiple structural analyses are not kept in parsing, and thus supports the serial nature of structure building.

Based on this data, Van Gompel, Pickering & Traxler (2001) proposed the *unrestricted race model*, in which comprehenders attempt to be build multiple structures in the case of ambiguity but only one of them is adopted and kept for further analysis. This model, although it contains initial parallelism, is essentially a serial model in that only one of the structural alternatives survives. The critical difference between the unrestricted race model and the Garden-path model is whether the initial structural choice is fixed or varies across individual trials/items. The unrestricted race model adopts the latter position. That is, it is a serial, but *variable choice,* model. This explains Traxler et al. (1999)'s facilitative effect of ambiguity. In (5-8a), no matter which analysis a reader chooses, the modification is compatible with either. In comparison, in (5-8b) and (5-8c), only one analysis yields a plausible sentence interpretation, and hence when the reader makes a wrong commitment, it disrupts processing. This explains the slower reading time in (5-8b) and (5-8c) compared to (5-8a).

It seems fair to say that the debate is yet to be resolved, although there is interesting positive evidence for the serial nature of structure building (Van Gompel et al., 2001). Given that more than two decades of fairly intensive research on this topic has not resolved this issue, it is unclear whether this debate can ever resolvable with the existing kinds of data. Some new approach might be useful. Under the single system view, a whole new body of data will become relevant to this issue. As stated above and in Chapter 1, under the single system view, the parser and the generator should align in terms of how they build structure. Thus, the same question can be asked: when generating a sentence, do speakers pursue multiple structural alternatives? Now I will discuss the relevant part of the production literature.

An explicit discussion about the serial/parallel structure building distinction in the production literature is difficult to find, perhaps because the terminologies are different. However, a relevant discussion can be found in the issue of whether to incorporate a competitive principle when making a structural choice. For instance, Dell & O'Seaghdha (1994)'s model of sentence production adopts competition as the mechanism for choosing between alternative structures. The key mechanism in this model is *lateral inhibition*, by which the competing structures inhibit each other (Fig. 5-3).

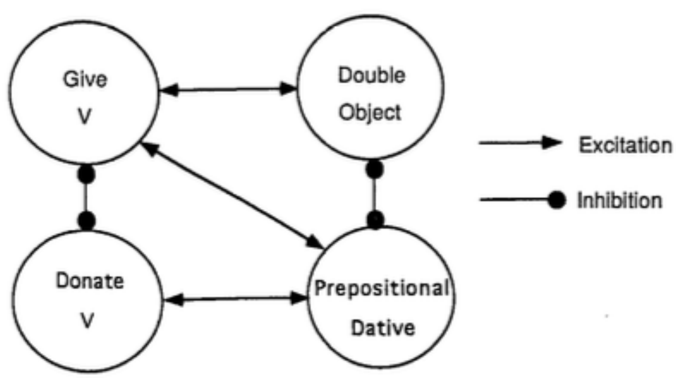



Figure 5-3: A competitive model adopted from Dell & O'Seaghdha (1994).

Although Dell & O'Seaghdha's model assumes competition between alternative structures, some experimental evidence suggests otherwise. For instance, V. Ferreira (1996) used a word rearranging task to have participants produce ditransitive sentences. The sentences sometimes contained a verb that allows both PO and DO (e.g., *give*), and sometimes contained a verb that allows only PO (e.g., *donate*). The competitive model predicts that the sentences with a flexible verb should be harder, because two available alternative structures should compete with each other, and therefore, incur the cost associated with competition resolution. On the other hand, the non-competitive model predicts that the sentences with a flexible verb should be easier, because speakers are free to produce whichever lexical item (either theme noun phrase or goal noun phrase) as it comes to mind. The results support the prediction of the non-competitive model: the error rate, as well as production latency, suggests that sentences with flexible verbs are easier to produce than the sentences with non-flexible verbs. This provides strong evidence that, upon facing a structural choice point, speakers consider only one structure, depending on the availability of a word. Notice that this evidence can be seen as in parallel with van Gompel et al. (2001) reviewed above: in both cases, the availability of structural choice facilitates processing. I argue that V. Ferreira (1996) and van Gompel et al. (2001) are tapping into the same property of the structure building mechanism, namely serial, but variable choice, structure building.

Certainly, there is some potential evidence for structural competition as well. For instance, Myachykov et al. (2013) found that the number of possible ways to describe a picture positively correlates with the speech onset latency, suggesting that structural competition does occur (e.g., Fig. 5-4). Comparing Russian and English speakers, they found that the  speech onset latency is generally longer in Russian speakers, and the onset measure was predictable from the number of alternative descriptions for a specific picture. They argued that this reflects the competition between alternative structures. These results, however, have an alternative explanation. Namely, it could reflect lexical, rather than syntactic, competition. In their experiments, the structural alternatives covary with the lexical alternatives, and both the agent and patient are animate. This is unlike in V. Ferreira (1996), in which the two competing nouns differed in animacy. This maximizes the possibility that two lexical items will compete with each other in Myachykov et al. (2013) but not in V. Ferriera (1996), according the principle of similarity based interference (e.g., Van Dyke & McElree, 2003). Thus, it is unclear whether the results of Myachykov et al. (2013) count as evidence for pure structural competition.

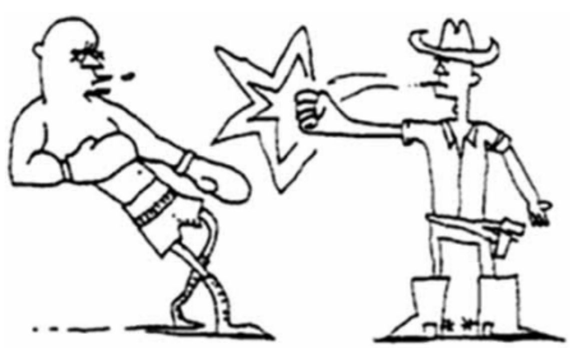


Figure 5-4: An example picture stimulus in Myachykov et al. (2013).

Another source of evidence for structural parallelism and competition comes from a type of speech error named *syntactic blends* (e.g., Garrett, 1980; Coppock, 2010). An example of this error is (5-9), taken from Fay (1982):

(5-9) *It doesn't make any matter.*

In this sentence, it seems like two expressions, *it doesn't matter* and *it doesn't make any sense,* are blended, resulting in a sentence that is likely not intended. However, it is not clear if this type of error is actually reflecting the competition of multiple syntactic frames, or the competition between lexical items. For instance, (5-9) can be explained as the parallel activation of the idiomatized lexical chunks *doesn't make any difference* and *doesn't matter*. Blending errors generally involve the substitution of lexical contents, and thus it is hard to assess whether speakers are keeping multiple structural alternatives active when faced with structural choice points in sentence generation. A more convincing case for purely structural competition should be blending between two lexically identical, but syntactically non-identical, utterance alternatives. For instance, when producing ditransitive sentences, parallel models should predict that both PO and DO structures are activated to some extent. When this competition does not resolve properly, speakers can in principle erroneously switch between PO and DO in the middle of an utterance, like in the imaginary error like (5-10).

(5-10) *Bob gave the man that was standing by the table to…*

In this imaginary error, the sentence first proceeded with a DO structure, and erroneously switched to PO structure, after the portion of DO structure was produced. In theory, this type of error seems to be predicted under models that assume parallel competition between structural representations during production. The question is whether this kind of error actually exists. I do not know if it does. Syntactic blending errors, as usually observed, to large extent can be explained by the competition between lexical contents, and it is not clear if pure syntactic blending errors actually can be observed.

The take-away from this discussion is the importance of separating lexical and structural competition. In order to support the existence of structural parallelism and competition, it needs to be shown that multiple alternative structural representations can compete with each other independently of lexical competitions. It is not trivial to separate syntactic and lexical competitions, as changing the syntactic structure almost necessarily involves changing the order of words, and changing the order of words can cause lexical competition. Given such intertwining between syntactic and lexical competitions, V. Ferriera (1996)'s evidence is especially important, because he observed the facilitation due to the availability of structural choice despite such a potential lexical competition. More research is clearly needed for this topic, but his evidence counts as fairly good positive evidence for the model without structural competition.

All in all, neither the serial/parallel debate in parsing nor the competitive/non-competitive debate in generation have been settled, and more research is clearly needed. However, what I would like to highlight here is that the single system view unifies two bodies of research, in such a way that each constrains the other. Moreover, when two relevant bodies of research are considered in tandem, there is interesting converging positive evidence for the serial nature of structure building: namely, facilitative effects due to the availability of structural choice.

*Interim summary*

In this section, I argued that the serial/parallel distinction is equally relevant in both parsing and generation. I argued that the ambiguity resolution literature in parsing and the structural flexibility literature in generation are tapping into the same aspect of structure building and hence should be considered in tandem. In parsing, the argument for a serial system is mostly based on negative evidence (i.e., null effects), and thus weak. However, when the data from both parsing and generation are considered, there seems to be some interesting converging positive evidence for the serial nature of structure building.

## *5.4 Incrementality and cost metric*

It is generally agreed upon in the modern psycholinguistics literature that parsing is incremental. Incrementality in parsing is usually motivated by consideration of (i) speed and (ii) memory limitations. Namely, it maximizes speed by allowing the parser to build structure (which is a necessary prerequisite for a systematic interpretation, see Chow & Phillips, 2013; Stroud & Phillips, 2012; for an overview, but see F. Ferreira & Patson, 2007 for an alternative view) without waiting for further input. This strategy avoids the accumulation of unanalyzed input, and thus minimizes the memory load by allowing the parser to integrate the incoming input immediately to form coherent structures. The critical question is whether the generator is also incremental, and if so, is it so in the same way and to the same degree.

To begin with, the term incrementality is extensively used both in parsing and generation literature (e.g., Stabler, 1994; Sturt & Lombardo, 2005; Demberg, 2010; Schneider, 1999; Phillips, 1996; Levelt, 1989; Kempen & Hoenkamp, 1996; De Smedt, 1990, 1996; F. Ferreira & Swets, 2002 among many others). However, on scrutiny, it is not clear if this shared terminology reflects the shared property of structure building, or simply reflects the surface similarity that both parsing and generation apparently proceed from 'left-to-right,' due the physical constraint that sound must unfold linearly from left-to-right. I argue for the latter. In particular, I argue that some conceptual reformulation of the notion of incrementality is necessary to align parsing and generation models. This conceptual reformulation helps organize the otherwise conflicting evidence in the production literature regarding 'scope of planning' (explained below). To do so, I first must introduce the memory limitation argument that is used to motivate incrementality in parsing and generation. This involves defining the 'cost metric' that is implicitly/explicitly ingrained in each incremental parsing and generation model.

In its original sense, incremental parsing simply means that structural analysis is performed as a new piece of input arrives, without waiting for further input. This point can be demonstrated by the classic garden path sentences (Bever, 1970) repeated below for a readers' convenience.

(5-11) *The horse raced past the barn fell.*

This sentence is locally ambiguous, and comprehenders have been shown to struggle with it at the end of the sentence when they encounter *fell*. One classic implication of this sentence is that structural analysis is incremental: the parser commits to the analysis that *raced* is the main verb of *the horse*. This structural commitment turns out to be incorrect at the point when *fell* is encountered. As there is no place in the structure where another verb can be accommodated grammatically, the parser needs to backtrack and reanalyze *raced* as the verb of the (reduced) relative clause modifying *the horse*. This explanation is not possible if the parser waits until the last bit of sentence before making a structural commitment.

It is worth emphasizing that the evidence from garden path reanalysis effects in online processing does not necessarily imply that sentence *interpretation* is incremental (cf. Crain & Steedman, 1985), at least under the assumption that restructuring a sentence structure itself is costly. It suggests that the structural analysis is incremental, and it is a further empirical question whether interpretation is also incremental (though there is evidence suggesting that interpretation is also incremental, e.g., Marslen-Wilson, 1975). Incremental structure building and incremental interpretation are two conceptually different, though tightly related (in that the latter presupposes the former), notions. For example, it has been shown that structure building processes of head-final languages are incremental (Aoshima, Yoshida & Phillips, 2006; Kamide, Altmann & Haywood, 2003), to the extent that a subtle constraint of syntax like binding restrictions is realized before the head of the sentence is heard. However, does this mean that comprehenders commit to a certain semantic role interpretation for each pre-verbal argument? It is not clear if it does. This is because, without verbs, it is not even clear whether semantic roles, especially those of internal arguments (theme/patient), can be definable (e.g., Kratzer, 1996; 2003). Furthermore, pre-verbal morphological information like case markers is reliably an indicator of structural position, but not semantic roles. As this thesis is specifically about structure building, I focus on the incrementality in structure building, without committing to whether the comprehension process as a whole is also incremental (cf. Crocker, 1996). Thus, the readers should note that the definition of incrementality may be narrower than how it is sometimes used in the comprehension literature generally, where incrementality is defined in terms of, for instance, maximal information use (e.g., Altmann & Mirković, 2009).

Any psychologically plausible model of language comprehension must achieve incrementality to some extent, and the consideration of the speed of comprehension and memory limitation leads some researchers to argue for a strong form of incrementality (e.g., Crocker, 1996; Konieczny, 1996; Schneider, 1999; Phillips, 1996). A strongly incremental parser satisfies the condition of *connectedness* (Frazier, 1979; Stabler, 1994; Sturt & Lombardo, 2005; Demberg, 2010; Schneider, 1999; Stabler, 1994, among others). If the parser satisfies connectedness, it establishes a single structure throughout the course of analyzing a sentence. Each lexical item is immediately incorporated into the main structure as soon as it is encountered. In other words, multiple sub-structures are not built for different parts of a single sentence. This ensures that no words/phrases are left unincorporated into the one coherent parse of a sentence, and thus also ensures that only a single structural representation occupies working memory at any point in time. This can be understood as being related to the notion of chunking in working/short-term memory models (e.g., Miller, 1956), and, just like chunking, it ensures that the parser will not be overloaded with multiple unintegrated items in the face of rapid streams of input as in real time comprehension. As reviewed in Chapter 2, strong evidence for incremental parsing can be found in Aoshima, Yoshida & Phillips (2009), for example.

Like in comprehension, the term incrementality has often been used in the production literature (Kempen & Hoenkamp, 1987; De Smedt, 1990, 1996; Levelt, 1989; F. Ferreira, 2000; Iwasaki, 2011). Indeed, it is sometimes claimed that incrementality is the shared property of comprehension and production that points to the same underlying system (Kempen, 2000). However, the use of the same terminology is misleading, because, on scrutiny, the underlying notion is often different. In production, the term incrementality often describes the degree of synchronization between planning and articulation (e.g., Levelt, 1989; Griffin & Bock, 2000; Kempen & Hoenkamp, 1987; F. Ferreira, 2000; De Smedt, 1990, 1996 among others). For instance, De Smedt (1990) defined incrementality as *cascading,* as schematized in Fig. 5-5.

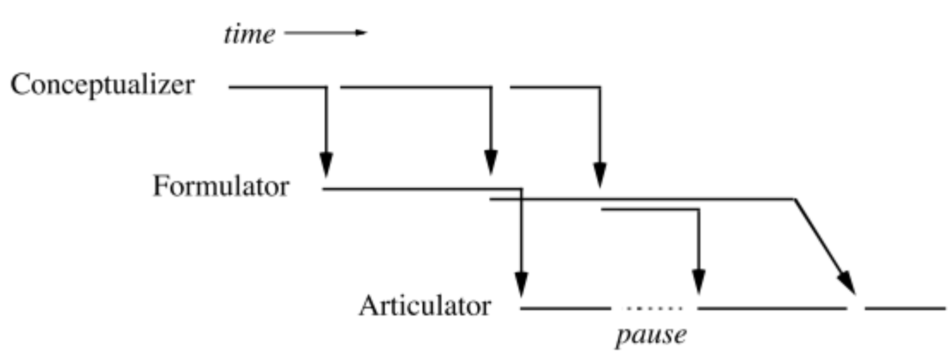



Figure 5-5: An illustration of cascading incrementality, adopted from De Smedt (1990).

Under this view, each piece of the sentence (usually a word) is passed on to the next level without delay (when the grammar allows it, at least). Conceptualized elements of the message are immediately sent to the 'formulator', by which a corresponding lexical representation (lemma and lexeme) is retrieved and assigned its structural status. The retrieved lexical representation is then immediately sent to the articulator. As a result, planning synchronizes as much as possible with the articulation. I will call this notion *cascading incrementality.*

Cascading incrementality is not referring to the same property of structure building as the notion of incrementality in parsing, despite the shared terminology. As discussed above, parsing is incremental if it satisfies the connectedness condition. In principle, the same notion of incrementality can be applied to production. If generation is incremental in the same way as in parsing, it requires that the every lexical item must be integrated into the single structural representation as soon as it is retrieved. I will call this notion of incrementality *incrementality in structure generation* to distinguish it from cascading incrementality.

To see the difference between cascading incrementality and incrementality in structure generation, it is useful to discuss one concrete experiment that has been taken to support the non-incrementality of sentence production. Meyer (1996) found that, using what is called extended Picture-Word Interference (ePWI) task, distractor words that are semantically related to *B* in the utterance like *the A is next to B* and *the A and the B* delay the onset of picture descriptions (relative to unrelated distractor words). This suggests that, before A is starting to be uttered, the lemma representation of B is already planned. This has been taken as the evidence that the sentence formulation is not incremental (e.g., F. Ferriera, 2000). Certainly, this evidence shows that production processes do not have to be cascading incrementally, but it does not show that it is not incremental in structure building. Structure building can still be perfectly incremental, maintaining connected structure in the entire course of processing. Speakers may simply choose not to immediately start articulating the sentence that they built incrementally.

This difference between cascading incrementality and incrementality in structure generation reflects the difference in how the memory load is conceived. That is, it reflects the difference in how the cost metric of structure building is defined. In incrementality in parsing and incrementality in structure generation, what incurs memory load are the floating substructures that are unintegrated. Incrementality in parsing and in structure generation is assumed to mitigate memory load because it prevents newly retrieved substructures from floating around in memory. As mentioned above, this definition of memory load is motivated by the idea of chunking in short-term/working memory (Miller, 1956). In comparison, in cascading incrementality, what incurs memory load is anything that has to be held in memory at all. Cascading incrementality is assumed to mitigate the memory load by 'releasing' a unit piece by piece by sending it to the next processing level. This is justified by the consideration of memory interference, where items of a similar kind (i.e., items within the same representational level) interfere with each other. The idea behind cascading incrementality is that it minimizes simultaneous processing of items at a particular level of processing (i.e., intra-level parallelism, De Smedt, 1996) so that it minimizes the similarity-based interference (e.g., Gillund & Shiffrin, 1984) that occurs within a level. Notice, however, that this assumes items can be forgotten/cleared from memory as soon as they are sent to the next processing level. It is not clear whether this assumption holds. At the very least, speakers must keep track of what they already said vs. what needs to be said in a sentence. In order for speakers to do this necessary task, some traces of memory need to be maintained, or speakers would constantly lose track of where they are in a sentence. Moreover, there is evidence suggesting that interference from already spoken bits of sentence occurs in production (*perseveration*; e.g., Cohan, 1980; cf. Nooteboom, 1969). Thus, it is not obvious if cascading incrementality is an effective strategy that minimizes memory costs. Of course, I do not deny that similarity-based interference is an important determinant of memory load in sentence production. It clearly is. The point that I emphasize here is simply that it is not obvious that cascading incrementality mitigates similarity-based interference.

Furthermore, cascading incrementality cannot plausibly be an architectural constraint of the production system. When a speaker is not under conversational time-pressure, for example, he or she can certainly elaborate on a whole sentence (or perhaps even two) as much as he or she wants before starting to articulating it. Of course, this is merely a phenomenological observation. However, experimental evidence also suggests that incremental production is not a requirement that has to be satisfied (e.g., Meyer, 1996; F. Ferreira & Swets, 2002; Wagner et al., 2010). Nevertheless, the debate about incrementality is often based on the evidence for or against the synchronization between planning and articulation, or equivalently the debate about the *scope of advanced planning*. The scope of advanced planning refers to the extent to which parts of an utterance are planned before articulation onset. When the scope of advanced planning is set to the minimum, planning and articulation synchronize as much as possible, i.e., production is cascading incremental. This scope of advanced planning literature has not produced converging results. Some studies suggest clausal scope (e.g., Meyer, 1996; Schnur, Costa & Caramazza, 2006[29]), some suggest phrasal scope (e.g., Smith & Wheeldon, 1999; Martin et al., 2010), and the others suggest a sub-phrasal scope (e.g., Griffin, 2000). These divergent results led some researchers to conclude that scope of advance planning is flexible (Ferriera & Swets, 2002; Konopka, 2012; Wagner, 2010), essentially giving up the idea that cascading incrementality is an architectural constraint of the production mechanism. However, such a flexibility view is compatible with incrementality in structure generation, because nothing in incrementality in structure generation requires planned parts of an utterance to be articulated immediately, as long as planned substructures are connected.[30] Thus, sentence production can be flexible in its scope of advance planning and be strictly incremental at the same time.

Moreover, cascading incrementality can prevent the incrementality in structure generation, under some circumstances with certain assumptions about what forms a grammatical constituent. For example, imagine that a speaker of English uttered the subject noun phrase, and then the conceptual element corresponding to object noun comes to mind. The cascading incrementality requires that the corresponding object noun

---

[29] Schnur, Cost & Caramazza (2006) tested the scope of planning at the phonological, rather than the lemma level of processing. However, under the assumption that the scope of planning at the higher level must be larger than the lower level of processing (e.g., Garrett, 1980), their evidence suggests that both conceptual and lemma level processing are clausal or more.

[30] Of course, I am not arguing that the advance planning of data is irrelevant to the question about incrementality in structure generation, because the timing of structure building can reliably modulate the timing of articulation. The only argument that I am making here is that the attested variability in the scope of advance planning can come from a factor that has little to do with structure building *per se*.

be retrieved. However, with the standard phrase structure grammar where the subject noun and object noun cannot by themselves form a constituent, the speaker cannot maintain a connected structure until he or she elaborates on a verb phrase. Thus, not only are cascading incrementality and incrementality in structure generation conceptually independent, the former could make the latter harder to satisfy under certain assumptions about grammatical formalisms and certain assumptions about the order of conceptualization. This does not mean that they are mutually exclusive, but extreme versions of cascading incrementality can be hard to reconcile with connectedness.

In Chapter 3, I have provided some arguments that sentence production is not cascading incremental, but is incremental in structure generation. In particular, I argued that if there is some mechanism that I described as syntactic gating mechanism, then cascading lexical retrievals are prevented. The evidence for such a mechanism was that lexical retrieval is strongly constrained by syntactic category, as shown in Experiments 2 & 3. I showed that similarity-based interference in a lexical retrieval task can be eliminated when the syntactic category of interfering items is manipulated to mismatch with the category of the target word. This demonstrates the existence of syntactic gating in lexical retrieval. Notice that this mechanism makes it easier to maintain connected structure throughout the course of production processes. The reason is that, because retrievals of lexical items are conditioned by the availability of a syntactic category slot, the retrieved lexical items are guaranteed to have a syntactic position in a sentence. In this sense, the syntactic gating mechanism is a mechanism for maximizing incremental structure generation by preventing cascading.

In Chapter 4, I showed that the advance planning of the verb occurs selectively before the articulation of the internal argument noun, but not before the external argument noun. I argue that this follows from the syntactic gating mechanism, based on the following reasoning. Articulating internal argument nouns requires computing the structure of verb phrase. Thus, the verb's structural position must be available before the internal argument can be articulated. On the other hand, articulating external arguments does not require computing the structure of verb phrase, and hence the verb's structural position does not have to be available before external argument can be articulated. The availability of verbs' structural slots makes a verb 'ready' to be retrieved while maintaining connected structure. Collectively, these studies provide some initial evidence for a syntactic gating mechanism that prevents cascading and encourages connectedness.

It may be worth emphasizing that I am not arguing that cascading does not happen at all throughout the course of production. Namely, I am not denying the empirical facts that whatever lexical items that are conceptually more prominent tend to occupy the earlier position in a sentence, according to the prediction of cascading incrementality. This is known as the availability effect on word order (e.g., Bock, 1987; Christianson & F. Ferreira, 2005; Slevc, 2011; MacDonald, Bock & Kelly, 1993). I do not claim that syntactic gating is active in the entire course of generation, and allows some parts of the sentence, especially the left-edge of constituents, to be constructed by a cascading, word-first, 'race-like' process. As we will see below, I will argue that structure generation is a mix of 'bottom-up' structure building (enumeration of syntactic structure from leaves to roots) and 'top-down' (enumeration of syntactic structure from roots to leaves) structure building processes, just like in some parsing algorithms (e.g., left-corner parsing). Importantly, syntactic gating presupposes some top-down structure building, and top-down structure building is usually preceded and constrained by bottom-up structure building. Until the top-down structure building mechanism creates some structural slots, the syntactic gating mechanism cannot be active. When the syntactic gating is inactive, cascading determines what is to be processed, triggering bottom-up structure building. See the next section for more detailed discussion on how structures are built during generation.

Finally, I argue that the same syntactic gating mechanism may be needed in order to maintain connectedness in parsing as well. In particular, comprehenders have been shown to predict upcoming lexical items (e.g., DeLong, Urbach & Kutas, 2005). However, there is a seeming tension between such a lexical prediction mechanism and incrementality. In particular, informative elements of a sentence for making a certain lexical prediction can be arbitrarily far away from where the predicted item can occur. If the retrieval occurs without syntactic gating, this makes it impossible for the parser to maintain connectedness, as lexical items will be retrieved in advance, arbitrarily before the syntactic position becomes available, and hence must be kept in memory in an unintegrated fashion. Thus, I argue that the same syntactic gating mechanism as in generation operates in (predictive) parsing. I offered some experimental evidence for this claim in Chapter 3. Namely, I showed that the syntactic category strongly limits the effects thought to reflect lexical retrieval processing difficulty.

*Interim summary*

In this section, I argued that the notion of incrementality in parsing and generation is not aligned well. I suggested an alternative conceptualization of incrementality in generation, which maximally aligns with the concept of incrementality in parsing. This alignment is based on the consideration of cost metrics in parsing and generation. I argued that, once the same notion of incrementality is adopted, the extant conflict in production with respect to the scope of advance planning dissolves. I furthermore argued for the existence of a 'syntactic gating' mechanism, which prevents cascading and encourages connectedness, in both generation and (predictive) parsing.

## *5.5 Order identity*

The discussion on incrementality sets the stage for tacking the last problem: the question of whether parsing and generation can be order identical. To recapitulate, the order identity hypothesis claims that the order in which structures are built in parsing and generation are identical. This is a much harder hypothesis to support than the mechanistic identity defended above, because it requires examining the internal order of parsing and generation operations and comparing them in detail. The internal order of generation operations has not been studied extensively, and it is not clear how it can be studied (though some interesting studies in the eye-tracking literature exist, e.g., Griffin & Bock, 2000). At the end of this chapter, I will offer some preliminary discussion about how the syntactic gating mechanism that I introduced above might plausibly align the order of structure building in parsing and generation.

5.5.1 Vertical order

In order to properly discuss the order of structure building, a discussion about the manner of structure building is necessary. This is because there are two dimensions in which order can be identical: vertical and horizontal. The vertical dimension determines whether structure building proceeds from-top-to-bottom (top-down) or from-bottom-to-top (bottom-up). The horizontal dimension determines the degree to which lexical input is analyzed from left-to-right. These two dimensions cannot be separated completely, because, the vertical dimension critically constrains the horizontal dimension. Thus, it is necessary to first discuss the vertical dimension - the manner of structure building.

Structure building can be described as top-down, bottom-up, or a mix of both. This distinction, although usually used in parsing, could also be relevant to generation. In order for the order identity to be true, the vertical dimension of structure building must align. Parsing models vary in how much top-down structure building is involved. To clarify, top-down structure building does not mean knowledge-driven processing, as the terminology is typically used in cognitive psychology. Rather, top-down structure building is structure building in which the process starts from the leaves rather than from the root, as in bottom-up structure building.

There is ample evidence that parsing involves top-down structure building. Such evidence includes the studies I reviewed in Chapter 2 (Lau et al., (2006), and Staub & Clifton (2006)) and these studies suggest that the parser predicts the syntactic category of the upcoming input both at the phrasal and lexical levels. In addition to these, many other studies show that such top-down structure building occurs not only locally, but also in a long-distance fashion. Much evidence for long-distance top-down structure building in parsing comes from the investigations on the processing wh-dependencies. For instance, Crain & Fodor (1985) studied the following types of sentences (5-12a and 5-12b):

(5-12) a. *Who had the little girl expected* ***us*** *to sing those stupid French songs for __ at Christmas?*

(5-12) b. *The little girl had expected us to sing those stupid French songs for Cheryl at Christmas.*

Using self-paced reading task, they found the reading time slowed down when the participant reached *us* in the wh-sentence (5-12a) relative to that in the declarative sentence (5-12b). Similarly, Stowe (1986) found a slow down effect due to the 'filled-gap,' at the region where bolded pronoun *us* is encountered instead of the phonetically null pronoun, the gap.

(5-13a) *My brother wanted to know who Ruth will bring* ***us*** *home to __ at Christmas.*

(5-13b) *My brother wanted to know if Ruth will bring* ***us*** *home to Mom at Christmas.*

In these studies, the slow-down effect at the pronoun region suggests that participants are 'surprised' to see the pronouns in (5-13a). One interpretation of this 'filled-gap effect' is that the human parser actively constructs a gap site before encountering the gap in the input. This is the case for top-down structure building, because the gap position is built from the top, before the gap site is recognized by the parser. These studies, and other studies that demonstrate the active nature of gap-filling (e.g., Omaki, et al., 2016) point to the involvement of non-local top-down structure building.

In Chapters 3 and 4, I presented some arguments for the involvement of top-down structure building (syntactic category constraints on lexical retrievals). Specifically, I claimed that the syntactic processing of the verb phrase structure precedes the lexical processing of the verb internal arguments in Chapter 3. Also, I presented evidence that the lexical retrieval is strongly constrained by the syntactic category of the word to be retrieved. Combined with previous evidence that available words tend to occupy the left-edge of sentences, I argued that generation involves both top-down and bottom-up processes in a single sentence utterance. Now, provided that both bottom-up and top-down structure building exist in generation, the question is whether there is any consistent 'division of labor' in building the structure of a sentence, and whether such a division of labor is the same between parsing and generation. Some models of production suggest that speakers do not have a single consistent way of building a sentence, but switch between two 'modes' of structure generation, two modes corresponding to radically bottom-up and radically top-down structure building procedures (Bock & V. Ferreira, 2014; Kuchinsky, Bock & Irwin, 2011; Van de Velde et al., 2014; Van de Velde, Kempen, & Harbusch, 2015). This essentially gives up on the idea that there is any constancy in the order in which structures are built, and hence is problematic to the vertical order identity view. However, it is worth noting that the primary evidence cited for such a flexible view (e.g., Kuchinsky, Bock & Irwin, 2011; van de Velde et al., 2014) can be explained in terms of the flexibility in the first lexical item that enters the generation process. What lexical item enters the generation first determines the following course of structure building, and the following course of structure building can still be consistent across utterances.

Of course, answering the current question about order identity to any theorist's satisfaction requires knowing all of the procedural details of structure building in both parsing and generation. Thus, psycholinguistics cannot currently offer a satisfactory answer. However, I argued that both bottom-up and top-down structure building are involved in parsing and generation, in a way that both can be used in constructing a structural representation of a single sentence. This is a necessary condition for the vertical order identity to hold true.

#### 5.5.2 Horizontal order

Horizontal order defines the order in which the lexical leaves of a structure are realized, i.e., the order in which structural positions are filled with lexical items. If one assumes purely bottom-up structure building, specifying vertical order automatically specifies horizontal order. This is because, in purely bottom-up structure building, every node grows out of a lexical item, and so there is no lexical filling to begin with. A problem arises, however, as soon as top-down structure building is involved. In many cases, a structural unit contains more than one terminal node, and there is, in principle, no constraint specifying which position gets filled first. In parsing, it is plausible that lexical filling is left-to-right. To the extent that lexical retrieval is triggered by sound input[31], the

[31] This is actually not so straightforward, as soon as one incorporates predictive lexical retrieval. This is because predictive lexical retrieval makes it possible for later-coming lexical items of the sentence to be retrieved before an earlier-coming lexical item.

order in which lexical items become available can plausibly be left-to-right, and hence lexical filling can plausibly be from left-to-right as well (I temporally ignore the non-linearity caused by predictive lexical retrieval). In contrast, in generation, one cannot simply assume that the concept arises in the mind in the same order that it is uttered. Of course, in some languages with little word-order constraints, the word order can be adjusted such that it perfectly corresponds to the order of lexical retrieval. But such an explanation requires that, everything else being equal, the processing load associated with sentence production dramatically varies across languages. Thus, if one wants to develop a cross-linguistically adequate model under the view that the order of lexical filling is identical between parsing and generation, there must be some mechanism that enforces left-to-right order in generation.

One potentially relevant body of evidence for left-to-right filling of lexical items comes from the eye-tracking literature in production. For instance, using eye-tracking during speaking, Griffin & Bock (2000) found the order of fixations on particular objects to be expressed in a visual scene corresponded strongly to the order of mention in a sentence. To the extent that the eye-fixation must precede the initiation of lexical retrieval, this pattern suggests that lexical retrievals are performed accordingly to the order of mention ('seeing for saying'). This pattern, however, does not tell us whether all kinds of lexical items are filled from left-to-right. The evidence is strictly about the ordering between different nouns, and there is no guarantee that lexical items of other syntactic categories (e.g., verbs, determiners, prepositions, etc.) follow this pattern. Even ignoring the concern that eye-fixations may not isolate lexical retrieval processes, the eye-tracking method may not be suitable to investigate the internal order of structure generation. The concern is that lexical items that are not nouns do not have a discrete corresponding region in the visual scene.

When looking at lexical items other than nouns, there are not many data points that we can make inferences about regarding the order of lexical retrieval. If anything, there seems to be counter-evidence for the left-to-right order of lexical retrieval. This comes from a study that investigated the relative time-course of determiner vs. noun head retrieval in Dutch, by Schriefers (1993). He used a picture-word interference task, and compared the time-course of interference with respect to adjectives and the head noun, varying the relative timing at which distractors and target pictures were presented. He interfered with the adjective's retrieval by presenting an adjective distractor that was semantically related to the target adjective, and interfered with the noun's retrieval by presenting a noun distractor that was semantically related to the target noun head. He also varied the type of utterance between determiner + adjective + noun and adjective + noun. He found that the adjective distractor delayed the speech onset in the determiner-less utterances only, while the noun distractor delayed the speech onset in both types of utterances. In other words, the effect of noun interference occurred at an earlier point than the effect of adjective interference. This pattern can be interpreted as evidence that the retrieval of the noun head, which linearly comes later than the adjective, is performed at an earlier point than the retrieval of the adjective, perhaps because the form of the determiner depends on the noun head but not the adjective. This pattern may suggest that generation is not entirely from left-to-right, at least when the phrase or sentence involves adjuncts.

*Interim summary*

I argued that the order identity view comes in two varieties, one concerning the vertical aspects of phrase structure building and the other concerning the horizontal aspects of lexical insertions. I reviewed the evidence that parsing involves top-down structure building, and provided a corresponding argument in generation, meeting a prerequisite for the alignment in terms of vertical order. In terms of the horizontal order, I argued that the horizontal order is harder to be aligned between the parser and generator.

## *5.6 A preliminary sketch of a single system model*

Admittedly, so far I have made only weak commitments to the specifics of the resulting model of the single structure builder. Thus, as the conclusion of this thesis, I will attempt to make some more specific commitments about the properties of the resulting single-system model, based on the evidence presented so far. Of course, some of these commitments will not be well supported empirically or not explicit enough to be implementable computationally, but they can be tested and elaborated upon in future research. I provide a sketch of a model in order to help readers criticize the current proposal more constructively and specifically, and at the same time in order to encourage them to come up with their own. I also would like to acknowledge that many aspects of this model are inspired by existing models, including, but not limited to, the ones by Vosse & Kempen (2002; 2009); Marcus (2001;2013); and Lewis & Vasishth (2005).

In the resulting single system model, structure building is an interleaved chain of retrieval and combinatorial operations between small-sized structural units and the lexical items that fill them. They form a chain because each operation triggers and conditions the next operation. Each structural unit does not have to be anchored to an overt lexical item. This commitment is motivated by the evidence that the identity of the lexical head need not be known to build a structure (e.g., Aoshima et al., 2009; Kamide et al., 2003). This means that some structural units may not have to be anchored to any overt lexical node; in this respect, the current model differs from, for example, lexicalized TAG (Joshi & Schabes 1997) and Vosse & Kempen's (2009) lexicalized models, among other lexicalized models of sentence processing. The size of each structural unit is small, as demonstrated by Chapter 2, Experiment 1. Specifically, it does not contain the subject noun phrases and verb phrases at the same time. Thus, a psycholinguistic analog to TAG's elementary tree corresponding to a clausal structure does not exist in this model.

Lexical items and structural units can be retrieved either based on conceptual or sound features associated with them, so the system can operate over both message inputs and sound inputs. However, those cues are used together with syntactic category features, which strongly guide the use of conceptual or phonological inputs. This is motivated by the evidence for the strong influence of syntactic category on lexical retrieval in (predictive) parsing and generation, as presented in Chapter 3. Retrieval operations do not generally happen in an opportunistic fashion but are triggered by the presence of category features in the structural representation. However, at the beginning of a sentence or a constituent, where no structural representation is built yet, lexical retrieval can happen opportunistically. This left-edge opportunism may explain the availability effect on word order in sentence production (e.g., McDonald et al., 1993). Note that the availability effect on the order of arguments in the dative alternation (e.g., Bock & Warren, 1985) may seem to cause a problem, because this ordering effect does not appear at the left-edge of the sentence. However, it does not necessarily present a challenge. This is because regardless of the structure a speaker picks between PO and DO, the syntactic category immediately following the verb is a noun phrase or a noun. Thus, the syntactic gating mechanism allows either the indirect or direct object to enter the structure, whichever comes first.

Once a lexical item is retrieved at the beginning of sentence, its category feature triggers the retrieval of a structural unit that contains the same category feature in its lower left edge. Of course, other kinds of thematic, semantic, or prosodic features also play a role in which lexical items and structural units should be retrieved, as retrieval of structural units does not have to be based purely on syntactic features. The retrieved structural unit can have an open structural position, in which case it provides the syntactic category cue for subsequent retrieval operations for lexical items that fill them, or structural unit that combines. Once the subsequent retrieval operation is completed, the newly retrieved lexical items or structural units combine with the existing structure. Since this combinatorial operation is generally preceded by a syntactically gated retrieval operation, combinatorial operations are guaranteed to be successful syntactically. This is because lexical items and structural units that are not syntactically combinable with the existing partial structure will not be retrieved to begin with, due to the syntactic gating mechanism. However, combinatorial operations are not guaranteed to be successful semantically. This is because the wrong lexical items or structural units of the same syntactic category can pass the syntactic gate, resulting in speech errors that are semantic in nature, like exchanges and substitutions (e.g., Fromkin, 1984; Garrett, 1980). However, production generally is successful semantically as well, because, retrieval of lexical items is also controlled by conceptual and sound features, and potential monitoring can catch the errors before the articulation (e.g., Levelt, 1983). On the side of comprehension, the exact same mechanism can explain the contrast between Experiments 4 and 5 in Chapter 3, where it was shown that prediction of lexical items is only slowly constrained by thematic/grammatical roles but rapidly constrained by syntactic category.

The resulting model is strongly incremental, maintaining a connected structure throughout the parsing and generation processes. This condition for connectedness is satisfied via the syntactic category gating mechanism, which ensures that any retrieved substructures are immediately combinable with the existing structure. For instance, even when a message element that corresponds to the verb is available early in verb final language production, the structure builder will not initiate a retrieval of the verb until the structural position for the verb is ready. This may capture the relative timing of verb retrieval in English and Japanese as discussed in Chapter 4, because presumably the verb position is available before the internal argument noun can be uttered, while it may not be available before the external argument can be said. Thus, the model is not cascading incremental (De Smedt, 1990; Iwasaki, 2011), but incremental in the respect that it strives to maintain a single connected structure.

The model exhibits initial parallelism in retrieving structural units, because retrieval is based on content-addressable cue based retrieval, which is parallel in nature (e.g., McElree, Foraker & Dyer, 2003). This content-addressability holds for both lexical and structural memory retrieval. However the model is essentially serial in that it maintains only one structural alternative in the course of parsing and generation. In particular, in Chapter 2, I argued that van Gompel, et al.'s (2001) and V. Ferreira's (1996) evidence suggests that the parser and the generator both speed up their processing when a structural choice is available.

It is hypothesized that the current model exhibits consistent order in which structure is built vertically due to the syntactic gating mechanism. This means that whether a particular part of a particular sentence is built in a top-down or bottom-up fashion is consistent within and across parsing and generation, but not necessarily the order in which the lexical items fill the structural slots, especially in generation. For instance, if a structural unit contains more than one lexical slot, two lexical items may start filling the syntactic positions in whichever order is determined by the availability in the sound/message input. This is partly motivated by the result of Experiment 10, where it was found that the planning of the verb root might precede the planning of the subject noun head in passive utterances. Such an order of planning is surely not consistent with parsing, in which the processing of the subject of a passive sentence precedes the processing of the verb of that sentence. Thus, at the lexical level of the structure, the order in which structural slots are filled with lexical items may not align between parsing and generation.

In sum, the model posits relatively small-sized structural memory units and lexical items as a basic unit of retrieval, which are retrieved by conceptual and sound cues, in combination with syntactic category cues. It assumes that structure building is a chain of retrieval and combinatory operations, in which the first move has a cascading effect on the subsequent moves. It builds structure in a race-like, serial fashion, and it is strongly incremental in that it maintains connected structures. The key mechanism underlying this incremental nature is the syntactic gating mechanism. The model exhibits a consistent vertical order of structure building in which structures are elaborated either in top-down or in bottom-up fashion, but not necessarily horizontal order in which lexical items fill the structural slots. Of course, this thesis did not address all the challenges that have been presented, and did not provide complete evidence for the alignment in terms of the core properties of the parser and the generator. However, I hope that the experimental evidence and the arguments presented throughout this thesis show that aligning the model of parsing and the the model of generation is more plausible than previously thought.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3), 233-250.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 791.

Allum, P. H., & Wheeldon, L. R. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1240.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.

Altmann, G. T., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, 31(5), 685-712.

Altmann, G. T., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, 38(4), 459-484.

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive science*, 33(4), 583-609.

Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event−related brain potential study. *Psychophysiology*, 32(2), 177-190.

Anderson, J. R. (2013). *The architecture of cognition.* Psychology Press.

Anderson, J. R., Budiu, R., & Reder, L. M. (2001). Theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language*, 45(3), 337-367.

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax*, 12(2), 93-134.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Badecker, W. and Lewis, R. L. (2007) A New theory and Computational Model of Working Memory in Sentence Production: Agreement Errors as Failures of Cue-based Retrievals. In *Proceedings of the CUNY Sentence Processing Conference*, San Diego, CA.

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing.* Doctoral dissertation. Massachusetts Institute of Technology.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–9.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language,* 68: 255-278.

Barner, D., & Bale, A. (2002). No nouns, no verbs: psycholinguistic arguments in favor of lexical underspecification. *Lingua*, 112(10), 771-791.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models*. Unpublished manuscript.

Bever, T.G (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Bever, T. G., & Sanz, M. (1997). Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguistic Inquiry*, 28, 69-91.

Bicknell, K., Elman, J., Hare, M., McRae, K., & Kutas, M. (2011). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4), 489–505.

Bloom, P., A. Barss, J. Nicol, & L. Conway (1994). Children's Knowledge of Binding and Coreference: Evidence from Spontaneous Speech, *Language,* 70, 53-71.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(4), 575.

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

Bock J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language,* 31, 99–127.

Bock, J. K. & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds), *The Oxford Handbook of Language Production* (pp. 21-46). New York: Oxford University Press.

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

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

Bock, J.K., Levelt, W.J.M., (1994). Language production: grammatical encoding. In: Gernsbacher, M. (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

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

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

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, 23(1), 45-93.

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437-458.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. (2003). Minding the clock. *Journal of Memory and Language*, 48(4), 653-685.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer* (Computer program). Available at http://www.praat.org/.

Borer, H., & Grodzinsky, Y. (1986). Syntactic cliticization and lexical cliticization: the case of Hebrew dative clitics, In Borer, H. (Ed.), *The syntax of pronominal clitics, syntax and semantics,* 19 (pp. 175–215). New York, NY: Academic Press.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59, 55-73.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1), 34-44.

Burzio, L. (1986). *Italian syntax.* D. Reidel: Dordrecht.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4), 572-582.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review*, 113(2), 234.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

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

Chomsky, N. (2008). On phases. *Current Studies in Linguistics Series*, 45, 133.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain research*, 1506, 76-93.

Chow, W. Y., Wang, S., Lau, E. & Phillips, C. (Under revision). *Electrical brain potentials revealed temporal dynamics of word prediction during language comprehension.*

Chien, Y.-C. & Wexler, K. (1990). Children's Knowledge of Locality Conditions on Binding as Evidence for the Modularity of Syntax and Pragmatics, *Language Acquisition,* 1, 225-295.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a 2free word order language (Odawa). *Cognition*, 98(2), 105-135.

Clifton, C., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, 2(2), 234-250.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49(3), 317-334.

Cohen (1980). Correcting of speech errors in a shadowing task. In V A Fromkin (ed.). *Errors in Linguistic Performance: Slips of the Tongue, Ear. Pen, and Hand*. London: Academic Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.

Coppock, E. (2010). Parallel grammatical encoding in sentence production: Evidence from syntactic blends. *Language and Cognitive Processes*, 25(1), 38-49.

Cowan, N. (2005). *Working memory capacity*. Psychology press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. *Natural language parsing*, 320-358.

Crocker, M. (1996). *Computational Psycholinguistics*. Dordrecht: Kluwer.

De Smedt, K. (1990). IPF: An incremental parallel formulator. *Current research in natural language generation* (pp. 167-192).

De Smedt, K. (1996). Computional models of incremental grammatical encoding. In A. Dijkstra & K. de Smedt (Eds.). *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 279-307). London: Taylor & Francis, 1996.

Dell, G. S. , O'Seaghdha, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In Dagenbach, D. (ed.). *Using smart source parsing. Inhibitory processes in attention, memory, and language*. (pp. 409-453). San Diego, CA: Academic Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, 93(3), 283.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1634), 20120394.

DeLong, K. a, Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–21.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, 8(8), 1117-1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.

Demberg, V. (2010). *Broad-coverage model of prediction in human sentence processing*. Doctoral dissertation, University of Edinburgh.

Dikker, S., Rabagliati, H., & Pylkkanen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, 110(3), 293–321.

Eberhard K., Cutting J., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review,* 112, 531–559.

Fay, D. (1982). Substitutions and splices: A study of sentence blends. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 163-195). Amsterdam: Mouton.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, 41(4), 469-495.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation time for prepared utterances. *Journal of Memory and Language,* 30, 210-233.

Ferreira, F. (2000). Syntax in language production: an approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of memory and language*, 25(3), 348-368.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1−2), 71-83.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46(1), 57-84.

Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause“ island” contexts. In A. Cutler (ed.), *Twenty-first Century psycholinguistics: Four cornerstones* (pp. 263-278). Mahway, NJ: Lawrence Erlbaum Associates.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35(5), 724-755.

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

Ferreira, V. S., & Humphreys, K. R. (2001). Syntactic influences on lexical and morphological processing in language production. *Journal of Memory and Language*, 44(1), 52-80.

Ferreira, V.S., & Slevc, L.R. (2007). Grammatical encoding. In Gaskell, M.G. (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford University Press.

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32, 669 – 692.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, 44(4), 516–547.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognitio*n, 28(1-2), 3-71.

Ford, M. & Holmes, M. (1978). Planning units and syntax in sentence production. *Cognition*. 1978;6:35–53.

Ford M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (ed.). *The mental representation of grammatical relations* (pp. 797–827). Cambridge, MA: MIT Press.

Frazier, L. and Fodor, J.D. (1978) The sausage machine: A new two-stage parsing model. *Cognition* 6, 291-325.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research*, 1, 183–192.

Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic inquiry*, 39(3), 355-377.

Fromkin, V. (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Garrett, M.F. (1975) The analysis of sentence production. In G.H. Bower (Ed.) The psychology of learning and motivation (Vol. 9). New York: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. L. Butterworth (Ed.), *Language production, Volume I: Speech and talk* (pp. 177-220). London: Academic Press.

Garrett. M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), Linguistics: The Cambridge survey, HI: Language: Psychological and biological aspects (pp. 69-96). Cambridge, UK: Cambridge University Press.

Garrett, M. F. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain* (p. 31). San Diego, CA: Academic Press.

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

Geschwind, N. (1970). The Organization of Language and the Brain. *Science*, 170, 940–944.

Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject-verb agreement production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 39(2), 515–28.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, 91(1), 1.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4), 274-279.

Griffin, Z. M., & Weinstein-Tull, J. (2003). Conceptual structure modulates structural priming in the production of complex sentences. *Journal of Memory and Language*, 49(4), 537-555.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: the MIT Press.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*. Oxford, UK: Wiley-Blackwell.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438-441.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 193–204.

Halle , M. , & Marantz , A . (1993). Distributed morphology and the pieces of inflection . In K. Hale& S. J. Keyser (Eds.), *The view from building 20* (pp.111–176). Cambridge, MA : MIT Press.

Harley, H. & Noyer, R. (1998). Licensing in the non-lexicalist lexicon: nominalizations, Vocabulary Items and the Encyclopaedia, in H. Harley, ed., *Proceedings of the Workshop on Argument Structure and Aspect*, MIT Working Papers in Linguistics, 32.

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive psychology*, 42(2), 113-157.

Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research*, 22(2), 1993.

Hoeks, J. C. J., Stowe, L. a, & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1), 59–73.

Humboldt, W. V. (1836). *Uber die Verschiedenheit des menschlichen Sprachbaues*. Berlin: Claasen & Roether.

Hwang, H., & Kaiser, E. (2014). The role of the verb in grammatical function assignment in English and Korean. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1363.

Jackendoff, R. (1977). X-bar-Syntax: A Study of Phrase Structure. *Linguistic Inquiry Monograph 2*. Cambridge, MA: MIT Press.

Janssen, N., Melinger, A., Mahon, B. Z., Finkbeiner, M., & Caramazza, A. (2010). The word class effect in the picture–word interference paradigm. *The Quarterly Journal of Experimental Psychology*, 63(6), 1233-1246.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Joshi, A. K., & Schabes, Y. (1997). Tree-adjoining grammars. In Handbook of formal languages (pp. 69-123). Springer Berlin Heidelberg.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159–201.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, 49(1), 133-156.

Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network, in S. I. Small, M. K. Tanenhaus and G. W.

Cottrell (eds), *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics*, *Neuropsychology & Artificial Intelligence*, Morgan Kaufman.

Kegl, J. (1995). Levels of representation and units of access relevant to agrammatism. *Brain and Language*, 50: 151-200.

Kempen, G. (2000). Could grammatical encoding and grammatical decoding be subserved by the same processing module?. *Behavioral and Brain Sciences*, 23(01), 38-39.

Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics*, 12(1), 111-142.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive science*, 11(2), 201-258.

Kempen G., & Huijbers P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14, 185–209.

Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3), 345-380.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* . Prentice-Hall, Inc.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225.

Kim, C. (2006). Structural and thematic information in sentence production. In *Proceedings of the 37th Annual Meeting of the North East Linguistic Society*. University of Illinois, Urbana-Champaign.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1* (pp. 423-430). Association for Computational Linguistics.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85(1), 1–36.

Konieczny, L. (1996). *Human sentence processing: A semantics-oriented parsing approach.* Doctoral Dissertation. University of Freiburg, Institut für Informatik und Gesellschaft.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1), 143-162.

Kratzer, A. (1996). Severing the external argument from its verb. In: Johan Rooryck, Lauri Zaring (eds.): *Phrase Structure and the Lexicon*, 109 – 137. Dordrecht: Kluwer.

Kratzer, A. (2003): The event argument and the semantics of verbs. Unpublished Manuscript. University of Massachusetts at Amherst. [Downloadable at http://semanticsarchive.net/Archive/GU1NWM4Z/ ]

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 748.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, 119(1), 23-42.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, 1146, 23–49.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and language*, 98(1), 74-88.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502.

Lee, J., & Thompson, C. K. (2011). Real-time production of unergative and unaccusative sentences in normal and agrammatic speakers: An eyetracking study. *Aphasiology*, 25(6-7), 813-825.

Lee, M., & Thompson, C. K. (2004). Agrammatic aphasic production and comprehension of unaccusative verbs in sentence contexts. *Journal of Neurolinguistics*, 17(4), 315-330.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.

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

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Levin, B., & Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT press.

Lewis, R. L., & Vasishth, S. (2005). An activation‐based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3), 375-419.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, 44(1), 27-46.

Lindsley J. R. (1975). Producing simple utterances: How far ahead do I plan? *Cognitive Psychology*, 7(1), 1–19.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7(6), 485-495.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. Frontier in Psychology, 4(226), 10-33, 89.

MacDonald, M. C., & Thornton, R. (2009). When language comprehension reflects production constraints: resolving ambiguities with the help of past experience. Memory & Cognition, 37(8), 1177–86. doi:10.3758/MC.37.8.1177

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological Review, 101, 676-703.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. Psychological Review, 89(5), 483.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(3), 503.

Marcus, G. F. (2001). The algebraic mind: reflections on connectionism and cognitive science. Cambridge, MA: MIT Press.

Marcus, G. F. (2013). Evolution, memory, and the nature of syntactic representation. Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain, 27.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. Nature, 244, 522-3.

Marslen-Wilson, W., & Tyler, L. K. (1975). Processing structure of sentence perception.

Marslen-Wilson, W., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. The Quarterly Journal of experimental psychology, 45(1), 73-87.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C. L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. Cognition, 116(2), 177-192.

Mazuka, R., & Itoh, K. (1995). Can Japanese speakers be led down the garden path. Japanese sentence processing, 295-329.

McAllister, Tara, Asaf Bachrach, Gloria Waters, Jennifer Michaud, and David Caplan. 2009. Production and comprehension of unaccusatives in aphasia. Aphasiology 23: 989-1004.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive psychology, 18(1), 1-86.

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. Cognitive Psychology, 25(2), 188-230.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. Journal of Memory and Language, 48(1), 67-91.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology: General, 126(2), 99–130.

McRae, K., Hare, M., Elman, J., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. Cognition, 33(7), 1174–1184.

Melinger, A., & Dobel, C. (2005). Lexically-driven syntactic priming. Cognition, 98(1), B11-B20.

Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. Language, Cognition and Neuroscience, 29(6), 663–683.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. Journal of memory and language, 66(4), 545-567.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. Journal of memory and Language, 35, 477-496.

Miller, G. & Chomsky, N. (1963) Finitary models of language users. In R.D. Luce, R.R. Bush, and E. Galanter (eds)., Handbook of Mathematical Psychology, volume 2, pages 419–491. Wiley, New York, NY.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review 63 (2): 81–97. doi: 10.1037/h0043158. PMID 13310704.

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. The Quarterly Journal of Experimental Psychology, 66(8), 1601-1619.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. Cognition, 125(2), 317-323.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of experimental psychology: general, 106(3), 226.

Neville, H., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. I. (1991). Syntactically-based sentence processing classes: evidence from event-related brain potentials. Journal of Cognitive Neuroscience, 3, 151–165.

Nieuwland, M. S., & Van Berkum, J. J. a. (2006). When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience, 18(7), 1098–111.

Nooteboom, S, G. (1969). The tongue slips into patterns. In G, Sciarone, van Essen, A & Van Raad, A, (eds.). *Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Mouton.

Novick, J. M., Kim, A., & Trueswell, J. C. (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. Journal of Psycholinguistic Research, 32(1), 57-75.

Oishi, H., & Sakamoto, T., (2010). Immediate interaction between syntactic and semantic outputs: evidence from event-related potentials in Japanese sentence processing.

Poster presented at the 22nd annual CUNY Human Sentence Processing Conference, Davis, CA.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of memory and language, 31(6), 785-806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. Journal of Memory and Language, 34, 739-773.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. Journal of Memory and language, 41(3), 427-456.

Pechmann T, Garrett MF, Zerbst D. The time course of recovery for grammatical class information during lexical processing for syntactic construction. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004;30:723–728.

Pechmann T, Zerbst D. The activation of word class information during speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002;28:233–243.

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

Perlmutter, D.M. (1978). Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society* (pp. 157–190). Berkeley, CA: University of California, Berkeley Linguistics Society.

Phillips, C. (1996). Order and structure. Doctoral dissertation, Massachusetts Institute of Technology.

Phillips, C. (2013). Parser-grammar relations: We don't understand everything twice. Language Down the Garden Path: *The Cognitive and Biological Basis for Linguistic Structures*, 294-315.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. Studies in Linguistics, 6, 11-47.

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

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

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. Trends in cognitive sciences, 11(3), 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences, 36(04), 329-347.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 940.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of memory and language*, 43(3), 447-475.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165–192.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review,* 88(5), 454–462.

Resnik P. (1992). "Left-Corner parsing and psychological plausibility," in *Proceedings of the 14th Conference on Computational Linguistics,* Stroudsburg, PA.

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

Roelofs, Ardi. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition,* 42, 107-142.

Schneider, D. A. (1999). *Parsing and incrementality*. Doctoral dissertation, University of Delaware.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of psycholinguistic research*, 35(2), 189-213.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 841.

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1368.

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, 39(4), 609-632.

Schriefers, Herbert, Antje S. Meyer, and Willem J Levelt. 1990. Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29, 86-102.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2011). Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex*, 249.

Shieber, S. M. (1985). *Evidence against the context-freeness of natural language* (pp. 320-334). Springer Netherlands.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, 111(43), E4687-E4696.

Sitnikova, T.; Kuperberg, G.; Holcomb, P.J. (2003). "Semantic integration in videos of real0world events: an electrophysiological investigation". *Psychophysiology* 40: 160–164.

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. Journal of Experimental Psychology: Learning, Memory, and *Cognition*, 37(6), 1503.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73(3), 205-246.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 675.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, 48(1), 103-130.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. *Perspectives on sentence processing*, 303-336.

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of experimental psychology: Learning, memory, and cognition*, 32(2), 425.

Steedman, M., & Baldridge, J. (2011). *Combinatory categorial grammar. Non-Transformational Syntax: Formal and Explicit Models of Grammar.* Wiley-Blackwell.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience*, 2(2), 191-196.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American scientist*, 57(4), 421-457.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643.

Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and language*, 120(2), 108-126.

Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2), 291-305.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... & Federmeier, K. (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language*, 51(2), 247-250.

Tabossi, P., & Collina, S. (2002). The picture-word interference paradigm: conceptual effects in the production of verbs. *Italian Journal of Linguistics*, 14, 27-42.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes*, 11(6), 583-588.

Townsend, D., J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., & Tooley, K. M. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes*, 23(5), 609-645.

Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558–592.

Trueswell, J. C., & Kim, A. E. (1998). How to Prune a Garden Path by Nipping It in the Bud: Fast Priming of Verb Argument Structure. *Journal of Memory and Language*, 39(1), 102–123. doi:10.1006/jmla.1998.2565

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89-134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of memory and language*, 33(3), 285-318.

Van Berkum J. J., Brown C. M., Zwitserlood P., Kooijman V., Hagoort P. (2005). Anticipating upcoming words in discourse: evidence from erps and reading times. *Journal of Experimenal Psychology: Learning, Memory and Cognition*. 31 443–467.

Van Berkum, J. J. a, van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580–591.

Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behavior*, 23(1), 67-83.

Van de Velde, M., Kempen, G., & Harbusch, K. (2015). Dative alternation and planning scope in spoken language: A corpus study on effects of verb bias in VO and OV clauses of Dutch. *Lingua*, 165, 92-108.

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, 71(1), 124–144.

Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, NY: Academic Press.

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

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157-166.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45(2), 225-258.

Van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18(7), 1181–1197.

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–255.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280(5363), 572-574.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8(4), 314-317.

Vigliocco, G., Vinson, P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*. 94, 91–100.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75(2), 105-143.

Vosse, T., & Kempen, G. (2009). The Unification Space implemented as a localist neural net: Predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, 3(4), 331-346.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206-237.

Wheeldon, L. R., Smith, M. C., & Apperley, I. (2011). Repeating words in sentences: effects of sentence structure. *Journal of Experimental Psychology: Learning Memory and Cognition,* 37(5), 1051–1064.

Wheeldon, L., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *The Quarterly Journal of Experimental Psychology*, 66(8), 1653-1673.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6), 80-83.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, 1(6), 209-216.

Wright B, Garrett M. Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition*. 1984;12:31–45.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, 123(5).