Grammatical Encoding

Victor S. Ferreira, Adam Morgan, and L. Robert Slevc

University of California San Diego, University of Maryland

Please address correspondence to:

Victor Ferreira

University of California San Diego

Department of Psychology

9500 Gilman Drive # 0109

La Jolla, CA 92093-0109

U. S. A.

Email: vferreira@ ucsd.edu

Voice: +1-858-534-6303

Fax: +1-858-534-7190

At the heart of the faculty of language are the processes of grammatical encoding. Grammatical encoding has the task of selecting and retrieving the syntactic and lexical forms that can convey nonlinguistic thoughts (Vigliocco & Meteyard, this volume), and then determining the morphological forms and their constituent ordering in preparation for their phonological spell-out (Wheeldon & Konopka, this volume) and eventual externalization by the oral (Hickok & Walker, this volume) or manual (Corina & Lawyer, this volume) articulators. As such, grammatical encoding processes most directly determine the gross characteristics of our individual utterances. Therefore, it is only a minor indulgence to claim that to understand why and how grammatical encoding carries out its duties is to understand a significant part of the why and how of language itself.

In this chapter, we describe the state of the field by describing the major debates that current research on grammatical encoding addresses. To situate these debates, Section 1 broadly describes a consensus view of the general architecture of grammatical encoding (illustrated in Figure 1). This consensus holds that grammatical encoding consists of two component sets of subprocesses, one that deals with content and the other that deals with structure. Each set of subprocesses proceeds through two phases or stages, the first involving selection and the second involving retrieval. Section 2 then describes ongoing debates that operate within (or question aspects of) this consensus view, beginning with debates over this content-and-structure and selection-then-

retrieval character of grammatical encoding. Next, Section 3 describes two debates that have maintained a relatively high level of visibility, attesting to their fundamental status in the grammatical encoding literature: The first concerns the *incrementality* or *scope* of grammatical encoding (how far do we plan ahead in an utterance before beginning it?), and the second concerns the factors that influence syntactic choice (given an idea to express, why do we say what we say?). In Section 4, we look forward to emerging debates in the field that are likely to receive increased attention in the coming years, largely due to the confluence of their central questions with other prominent and topical issues in cognitive science. These debates concern rational or optimal production, effects of ongoing learning, and dialogue. With all of this discussion of debate, it is easy to lose sight of the insights that the field has made and how our knowledge of the way grammatical encoding works has accumulated, and so Section 5 closes on a constructive note, by highlighting two fundamental insights that we have gained as a field along the way.

1. Grammatical encoding: A consensus model

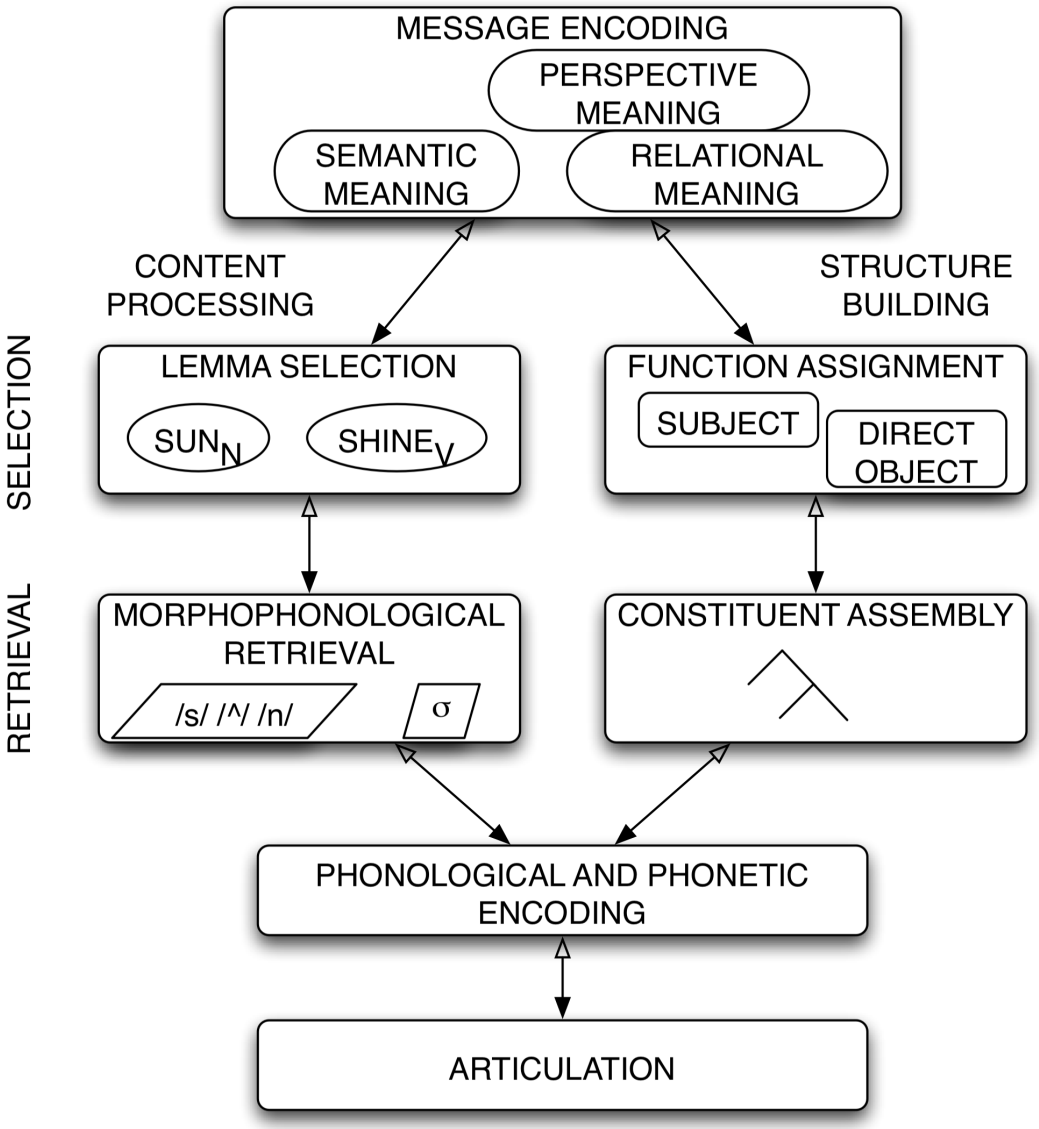

Figure 1. Schematic of consensus model of grammatical encoding. Filled arrowheads mark direction of primary information flow, open arrowheads mark possible feedback information flow.

Producing linguistic expressions involves encoding nonlinguistic meanings, termed *preverbal messages* (Levelt, 1989) or *interfacing representations* (Bock, 1982), into a set of linguistic representations that can ultimately be phonologically encoded. We will start our description of

grammatical encoding at the beginning, by describing what theories of grammatical encoding generally require preverbal messages to represent.

1.1 Message encoding.

Preverbal messages are derived from conceptual representations – that is, they are not *thought*, but the collection of semantic and pragmatic information that the language system needs so that it can encode the meaning that the speaker aims to express. The information in preverbal messages will typically be a superset of the information that a speaker wishes to convey, a necessary consequence of the fact that in order to make a sentence, languages often require a speaker to specify features like tense, aspect, number, perspective, animacy, and definiteness, any number of which may be irrelevant to the communicative goal.

The processes that formulate preverbal messages have been termed *message encoding* processes or the *message component* (Bock, 1995b; V. S. Ferreira, in press), *conceptualization* processes (Levelt, 1989), or *referential arena* processes (Bock, 1982). The first step of message encoding is to specify the goal to be achieved by producing an utterance -- whether to request something, provide information, guide action, and so forth. Any such *speech act* can only be carried out by expressing a meaning, and so the next step is to encode this meaning into the preverbal message. Meanings are encoded into what are here termed *events*, which represent the “who did what to whom” that is to be expressed by linguistic utterances. Events include three specific aspects of

meaning: *Semantic meaning* represents the *who*, *what*, and *whom* themselves -- the semantic features of the expressed entities, states, actions, events, etc. *Relational meaning* represents how the who, what, and whom in the event relate to one another -- who/what is performing which action or is in which state, who/what is having which action performed on it, and so forth. *Perspective meaning* represents which semantic or relational aspects of the event are more or less important -- for example, which aspects are foreground or background, which are topic or comment, and which aspects have meaning added to them and which aspects are the added meaning.

For example, take the utterance “In San Diego, it’s always sunny.” Someone might express this utterance to inform his or her interlocutor of a relevant observation (say, if they were discussing cities in southern California). This observation corresponds to an event in which the semantic meaning includes a representation of a geographical location that has the proper name “San Diego” and a state that is a weather condition brought on by unobstructed exposure to the sun. The relational meaning in the event involves attributing the state corresponding to the weather condition to the geographical location. The perspective meaning encodes that it is the geographical location that is the topic and the weather-state that is the added information (compared to, say, the meaning expressed by “It’s always sunny in San Diego”).

1.2 Grammatical encoding.

Once semantic, relational, and perspective meanings are encoded into at least part of a preverbal message, grammatical encoding can begin. Grammatical encoding consists of separable subprocesses that deal with formulating the content and specifying the structure of the eventual utterance. In turn, the content and structure subprocesses each proceed through two phases or stages (sometimes termed functional and positional processing; Bock, 1995b; Garrett, 1975, 1982, 1988), where linguistic features are selected from a candidate set, and then the properties of those selected linguistic features are retrieved.

1.2.1 Content subprocesses. Though possibly an oversimplification, the function of the content subprocesses of grammatical encoding is to select and then retrieve the details of the meaning-carrying or *content* words in an utterance. The first step of content processing is *lexical selection* (Bock, 1995b) -- picking a set of words that "covers" (i.e., conveys a sufficient extent of) the semantic meaning represented in the preverbal message. A counterintuitive finding is that there are at least two separate types of representation for every word: one, termed a *lemma* (Kempen & Huijbers, 1983; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999) or a *lexical entry* (Bock, 1995b), which encodes grammatical properties of the word like its form class (e.g., noun, verb, etc.) and grammatical gender, and another encoding metrical and phonological information.

Lemma selection is the first step in content processing. Lemmas in turn point to morphological, segmental, and metrical information necessary for spell-out of the word. The retrieval of this morphophonological knowledge forms the second step of content processing, namely *lexical retrieval* (Bock, 1995b). Most theories of grammatical encoding claim that access to this metrical and phonological knowledge is mediated by distinct whole-word representations sometimes termed *lexemes* (Kempen & Huijbers, 1983; Levelt, 1989; Levelt et al., 1999) or *word forms* (see also Garrett, 1975).

For example, take the utterance, “the sun shines.” Content subprocesses select and then retrieve the syntactic, morphological, and phonological details of the words “sun” and “shine,” but not the word “the” nor the suffix “–s” (which are not primary meaning-carrying words and so are termed *function* words). Lexical selection must determine the word that expresses the semantic meaning of the large gaseous body that warms the earth, and thereby select the lemma for “sun” (and not, say, “moon”), as constrained by the grammatical category membership of candidate lemmas (which in turn comes from structure subprocesses). As the lemma for “sun” is accessed, lexical retrieval retrieves the singular count-noun morpheme for “sun” as well as its segmental content (/s/, /^/, and /n/) and metrical specification (a single stressed syllable). A similar lexical-selection-then-lexical-retrieval process occurs for the word “shine.”

1.2.2 *Structure subprocesses*. While content subprocesses select and retrieve content words that convey semantic meaning, structure subprocesses

select and retrieve the syntactic representations necessary to convey relational and perspective meaning. Like content processing, this occurs through two stages, first selecting *grammatical functions* (Garrett, 1975) and then retrieving *positional* (Garrett, 1975) or *constituent* (Bock, 1995b) structures necessary to realize those grammatical functions. For ease of description, we focus on the expression of relational meaning, returning to the role of perspective meaning below.

The first step in structure building has been termed *function assignment* (Bock, 1995b). It involves selecting grammatical functions -- representations that relate one aspect of a linguistically encoded event to another -- in accordance with the relational meaning represented in the preverbal message. For example, the *subject, direct object*, and *indirect object* grammatical functions relate entities in events expressed by nouns to actions or states in events expressed by verbs. Different *modifier* functions might relate a simple property in an event expressed by an adjective or a complex property expressed by a full clause to an entity expressed by a noun. Thus, function assignment involves consulting the relational meaning in a preverbal message, and determining which grammatical functions must be selected to cover that relational meaning.

Once selected, grammatical functions proceed through the second step of structure building, which has been termed *constituent assembly* (Bock, 1995b). This involves the retrieval of constituent structures that can express the grammatical functions selected at function assignment. In *fixed word-order*

languages like English, this primarily involves arriving at a sequential ordering of words that convey the represented relational meanings. For example, in most English sentences, one of the nouns before a verb is the subject of that verb, and so a constituent structure must be assembled that specifies that sequential ordering. That is, to communicate "the sun shines," constituent-assembly processes must make sure "sun" is mentioned before "shines". This can be complicated because most sentences simultaneously express several aspects of relational meaning. For example, in "the bright sun that warms the earth shines," both "bright" and "warms the earth" bear a relational meaning to "sun," which in turn bears a relational meaning to "shines" (not to mention the relational meanings encoded in "that warms the earth"). In order for these multiple relations to be conveyed by a single sequence, constituent-assembly structures must appeal to hierarchical principles that determine how simultaneously expressed relational meanings can be embedded in one another so that an addressee can recover relational meaning from the resulting linear sequence. *Free word-order* languages like Japanese place less burden on sequential ordering, instead relying more heavily on affixes or *case-markers* that use phonological content to convey relational meaning (e.g., the suffix "–o" applied to a noun indicates that it's the object of a verb, relatively independently of the positions of that noun and verb). The role of such affixes in structure building implies that function words and affixes (e.g., "the" and "–s" from "the sun shines") are included as parts of constituent frames, rather than retrieved by content

subprocesses (e.g., most especially, Garrett, 1975). In short, relational meaning is expressed by an appropriate set of grammatical functions in an appropriate affixed sequence, as selected in the function assignment and constituent assembly stages of structure building, respectively.

This description of structure building has so far considered exclusively relational meaning, but perspective meaning also influences structure building in important ways that are often described as the *information structure* of an utterance (e.g., Lambrecht, 1994). This can be seen from the fact that a given relational meaning can be expressed with more than one combination of function assignments and constituent structures. Consider "In San Diego, it's always sunny" versus "It's always sunny in San Diego." Though these sentences express the same relational meanings, the different sequential orderings convey different perspective meanings, in that the first sentence is about San Diego and the second sentence is about always being sunny. Thus, perspective differences influence structure-building mechanisms by affecting both function assignment and constituent assembly processes. An example of perspective meaning primarily affecting function assignment is the difference between a sentence in the active and passive voice (which, for a verb like "kick," respectively assign the kicker or "kickee" to the subject function), whereas an example of perspective meaning primarily affecting constituent assembly is the difference between the examples above ("In San Diego, it's always sunny" and "It's always sunny in San Diego"). Of course, perspective meaning could influence both stages of structure

building, resulting in the production of complex perspective-communicating structures (e.g., clefts like “It is San Diego that is always sunny”).

1.3 On dividing and uniting.

Together, the operations of the content and structure subprocesses of grammatical encoding, proceeding through their respective selection and retrieval stages, determine the gross-level characteristics of speakers’ linguistic utterances. But the assumption that sentences are produced by independent content and structure subprocesses raises two thorny issues. First, why process content and structure separately? Second, if content and structure are processed separately, how are they brought together again?

The answer to the first question has to do with the sheer expressive power of language, in that we can describe with language almost any thought we can conceive, at least at some level of coarseness. This implies that the devices that create linguistic expressions must be systematic -- they must be able to cover a comprehensive range of possible meanings -- and they must be productive -- they must be able to create, in principle, an infinity of possible linguistic expressions. This systematicity and productivity of language derives directly from the separation of content and structure. That is, because a linguistic expression is a combination of two relatively independent devices (content and structure), each of which expresses relatively independent aspects of meaning (semantic and relational/perspective meaning), these devices can be freely combined to express any semantic meaning arranged in any relational or

perspective manner. If structure and content were not independent, then even if the separate semantic, relational, and perspective properties of an event were known, any previously unexperienced combination of those properties would not necessarily have a linguistic device for its expression. In short, it is the independent combination of content and structure that allows two bounded systems to systematically express boundless meanings (for a similar argument with respect to thought processes, see Fodor & Pylyshyn, 1988).

But, if structure and content are processed separately, they must be brought together again if speakers are to produce a single utterance that combines content and structure. And of course, the combination of content and structure has to be the *right* combination -- the word that is used to express the subject of the verb better express the semantic features of the subject rather than say, the object (a process that sometimes goes awry in speech errors; for seminal observations, see Garrett, 1975). This has been dubbed the *coordination problem* (Bock, 1987a), and is a version of the more general *binding problem* in information science. (Another well-known example of the binding problem is that, because the identity and location of visually perceived objects are processed separately [e.g., Ungerleider & Haxby, 1994], perception processes must somehow keep track of which location corresponds to which identity.) At the moment, no complete solution to the coordination problem exists. Grammatical category labels (noun, verb, etc.) are likely to be critical, as are something akin to event- or thematic-role representations (agents, themes,

goals, etc.). A different solution is to avoid the coordination problem altogether by claiming that content and structure are *not* represented separately, a point of view approaching that expressed by lexically-based theories of syntax (e.g., F. Ferreira, 2000; Levelt, 1989; see section 2.2 below for further discussion on this debate). Progress on this issue awaits not only further research, but likely, revolutionary insights into the way that cognitive mechanisms work.

2. Fundamental debates: On stages and structures

"Consensus" should not be confused with "unanimous." This section describes some of the challenges to the view just characterized that make it a consensus view and not a unanimous one. Along the way, we hope to show that, with some flexibility, enough semblance of the consensus framework can usually be maintained that its spirit survives (which has directed and structured research on grammatical encoding for over 40 years). Both major divisions in the consensus model have been questioned. We first discuss challenges to the idea that grammatical encoding consists of two separable stages (both with regard to content and to structure subprocesses), and then discuss challenges to the idea that content and structure are processed relatively separately.

2.1 Does grammatical encoding proceed through two stages?

The debate over the extent to which grammatical encoding can be characterized as involving staged mechanisms has played out mostly separately with respect to content and structure subprocesses, and so each is discussed in turn.

2.1.1 *Content processing.* The original modern-day view of lexical production (Garrett, 1975) and views that followed up on it (Levelt, 1989; Levelt et al., 1999) claimed that word production consults two stages of lexical representation. The first stage includes lexically specific representations that are critically syntactic in nature (i.e., one representation per word in the language, defined especially with respect to its form-class membership). This became the lemma level in current theoretical discourse. The second stage includes full word-form representations whose critical characteristic is a morphophonological nature. That this level was at least partly phonological in nature led to the claim that these representations are sound-form specific (e.g., *sun* and *son*, in all of their meanings, share one of these representations), so that they became the *lexeme* level in current theoretical discourse. The syntactic-versus-phonological distinction between these levels naturally leads to the view that the lemma level is modality general (so that the same level is consulted in speaking or writing, hearing or reading), whereas the lexeme level is modality specific.

A long-standing debate about these lexical representations concerns a detail of processing: Is access to these stages *discrete* (Levelt, 1989; Levelt et al., 1999), in that lemmas must be fully selected before lexemes begin to be retrieved? Or is access *interactive*, so that lexemes begin to be retrieved even before a lemma is fully selected (*cascading*), possibly even allowing partially retrieved lexemes to influence lemma selection (*feedback*; Cutting & Ferreira, 1999; Dell, 1986; Dell et al., 1997b; Rapp & Goldrick, 2000)? This debate has

maintained prominence due to its association with the modularity debate (Fodor, 1983) in psycholinguistics and cognitive science generally. It is worth noting, however, that even if processing is interactive, this leaves intact a fundamentally staged character to lexical production that Dell and O'Seaghdha (1991) termed 'globally modular but locally interactive' -- processing still proceeds through stages of lexical selection and lexical retrieval; it's simply a debate concerning whether the dynamics of retrieval are influenced by the intermediate products of selection (*cascading*), and whether the dynamics of retrieval influence the timing or nature of selection (*feedback*).

Another challenge to the original view poses that only one lexical stage operates during lexical production (for initial volleys, see Caramazza, 1997; Roelofs, Meyer, & Levelt, 1998). Some previous models have assumed only one lexical level, either for reasons of substance or convenience (see, e.g., Dell et al., 1997b); however, these models typically assumed the sole level to be the lemma level. The new challenge suggests that the sole level of lexical representation has properties that cross-cut the distinctions between lemma and lexeme: On the one hand, like lemmas, the representations are lexically specific (so "sun" and "son" have different lexical representations), but on the other hand, like lexemes, they are modality specific (with different representations for speaking, writing, hearing, or reading). Implications of this different organization has led to tests of alternative predictions, leading to the current disagreement (Caramazza,

Bi, Costa, & Miozzo, 2004; Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak, Meyer, & Levelt, 2003).

It is important, however, to view this debate for what it is and what it is not. This challenge does not hold that production has anything other than a selection-then-retrieval character. Indeed, the one-level view is compatible with the claim that processing is discrete rather than interactive, and so in some ways is *more* staged than some versions of the two-lexical-level view. The entire debate can be summarized with the question of whether lemmas are modality specific -- an important question that does not undermine the fundamental character of the above-described consensus model.

An interestingly different approach to lexical production comes from a series of experiments and theoretical proposals from Strijkers and colleagues (see especially Strijkers & Costa, 2016). This approach stems from electrophysiological evidence (i.e., changes in the electrical field given off by brain activity), showing that remarkably early changes in the electrophysiological record can be observed as a function of differences among words that, according to traditional theoretical approaches, should not arise until later in the timecourse of processing. Most striking is evidence showing that within 200 ms of seeing a to-be-named picture, systematic differences can be observed in the physiological record when bilingual speakers name pictures of objects that have *cognate* names (i.e., names that sound similar in the bilingual speakers' two languages) versus non-cognate names (Strijkers, Costa, & Thierry, 2010). The difference

between cognates and non-cognates is entirely phonological, and according to standard theories, phonology is not accessed until the last step of lexical selection (note that according to some estimates within traditional frameworks, phonological access does not even begin until 250 ms after picture onset, unfolding over the subsequent 350 ms; Indefrey & Levelt, 2004). Evidence like this has led Strijkers and colleagues to propose a quite different time course of information access, whereby an initial *ignition* stage involves parallel activation of neurons coding any information that is relevant to producing a word, followed by *reverberations*, or sequential activation of components of words' representations, ordered so that each component is available as it is needed. On the one hand, this alternative framework for lexical production represents a significant departure from the sequential, selection-then-retrieval nature of modal models of word production. On the other, especially given the reverbatory phase's similarity to sequential staged production, it is unclear whether the departure is a significant one at the level of information-processing (as opposed to the level of neural implementation, for which the implications are clearly profound).

This does raise the question of what kind of architecture *would* undermine the fundamental character of the consensus model. Two aspects of the consensus view are central to its selection-then-retrieval character. One is that it involves lexically specific representations -- symbols of the content words in a speaker's vocabulary. The second is that these lexically specific representations mediate meaning and form. As long as lexical production requires moving through lexically

specific representations to access phonological and eventually articulatory knowledge, then a selection-then-retrieval character is preserved (for a challenge to these kinds of assumptions in the domain of word reading, see Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989).

2.1.2 *Structure building*. Beginning three decades ago, the discreteness-versus-interactivity debate flared with respect to structure building, just as it had with lexical production. The question was whether function assignment could be influenced by the dynamics of constituent assembly. Some evidence suggested not (Bock, 1986a; Bock & Warren, 1985), whereas other evidence suggested so, at least indirectly (Bock, 1987b; Levelt & Maassen, 1981). Like the corresponding debate in lexical production, however, this question concerns an important processing detail of the consensus model that is relevant to its staged character, but not a fundamental challenge to that staged character.

A more recent challenge to this staged characterization comes from evidence gleaned from an especially powerful methodology for investigating structure building, namely *syntactic persistence* (which has also been termed *syntactic priming*, *structural persistence*, or *structural priming* -- a neat 2 x 2 nomenclature design). Briefly, speakers tend to persist in the use of previously processed structures. This is typically investigated by assessing the effect of the structure of a *prime* sentence upon the subsequent production of a *target* sentence. For example, speakers who hear or say passive prime structures are likely to describe a subsequent picture with passive target descriptions, relative

to if they had heard or said active prime descriptions (for a recent review, see Pickering & Ferreira, 2008).

One research thread assessing syntactic persistence has explored whether constituent assembly has its own staged nature. That is, after the functional structure of an impending utterance is specified, do speakers first determine dominance relations in a constituent structure, and only afterward *linearize* that structure into a specific word order? Evidence for this possibility is that mere word-order appears to exhibit syntactic persistence (e.g., the difference in Dutch sentences like “On the table is a ball” and “A ball is on the table”; Hartsuiker, Kolk, & Huiskamp, 1999; Hartsuiker & Westenberg, 2000; see also Vigliocco & Nicol, 1998). Alternatively, constituent structure may be undifferentiated, specifying dominance relations and linear order with a single integrated process. Evidence for this is that dominance relations by themselves do not exhibit persistence (e.g., structures like “The driver showed to the mechanic the overalls” does not cause the persistence of structures like “The patient showed the injury to the doctor”; Pickering, Branigan, & McLean, 2002).

A more fundamental challenge to the staged nature of structure building comes from recent modeling work by Chang (Chang, 2002; Chang, Dell, & Bock, 2006; Chang, Dell, Bock, & Griffin, 2000), demonstrating that much of the evidence from syntactic persistence can be simulated by architecturally complex computational models that learn to generate sequences of words. These architectures do not work through a straightforward sequence of function

assignment then constituent assembly. Instead, they work by developing sequencing representations akin to syntactic constructions that are triggered by a combination of lexical and event-semantic knowledge (as well as previous learning, leading to persistence). To the extent that these models can describe grammatical encoding successfully, they represent a very different way of construing the structure-building process.

2.2 Where is the line between content and structure?

In cognitive science broadly, it is not uncontroversial to claim that structure and content are separately processed (see Fodor & Pylyshyn, 1988; Rumelhart & McClelland, 1986). Interestingly, this distinction is more firmly established within the subfield of language production. Instead, the controversy with respect to theories of grammatical encoding has concerned the nature of the content and structure systems' representations.

The above-described consensus model proposes a very neat line between content and structure: Content subprocesses select and retrieve content words, structure subprocesses assign functions and build constituent structures. This view is often termed *frame-based*, because it assumes structures that are strictly independent of content-word content (although structure building must be influenced by selected content words; for an initial proposal, see Garrett, 1975). A recurring challenge to this neat division comes from *lexically based* models of grammatical encoding (see especially Levelt, 1989), which claim that content

words belong in the structure system, indeed forming the fundamental basis of the process of structure building.

One explicitly developed lexically-based account of structure building comes from F. Ferreira (2000). This approach grounds structure building in a linguistic formalism called *lexicalized tree-adjoining grammar* (Joshi, Levy, & Takahashi, 1975; Schabes, Abeille, & Joshi, 1988). In a nutshell, the approach argues that grammatical encoding builds structure by retrieving content words that include elementary trees -- component bits of syntactic structure that are unified by operations including *substitution*, *adjoining*, and more recently, *overlay* (F. Ferreira, Lau, & Bailey, 2004). This kind of approach has the obvious advantage that it can straightforwardly account for how structure-building processes operate so that only certain grammatical options are used with certain content words. In a frame-based view, such lexical dependencies must either derive from distinctions represented in the preverbal message (which is not ideal, given that many such dependencies have little or no basis in meaning), or they must arise during the coordination process described above. On the other hand, lexically based accounts require additional processing machinery to explain lexically independent structure-building effects, most especially evidence that syntactic persistence occurs completely independently of lexical content (e.g., Pickering & Branigan, 1998).

A different view of the content-structure distinction comes from Pickering and Branigan (1998). This approach represents structure with *combinatorial*

nodes that specify how content words can combine into constituent structures. Such nodes are viewed as like the traditional lemma nodes that are selected by content subprocesses; in fact, combinatorial nodes and lemma nodes form a kind of seamless network of grammatical-encoding knowledge. Thus, this approach allows for the representation of the just-described lexical dependencies as well as the independence of constituent structure that is implied by patterns of syntactic persistence. From the perspective of the distinction between structure and content, this approach is mixed. Unlike the consensus view presented above, structure and content freely intermingle, but unlike lexically-based approaches, structure knowledge (as represented by combinatorial nodes) and content knowledge (as represented by traditional lemmas) are fully distinct. The combinatorial-node-based approach is in principle compatible with a view of structure building that separates function assignment and constituent assembly, so long as function assignment can influence baseline activations of combinatorial nodes directly and constituent assembly is informed by the grammatical knowledge embodied by combinatorial nodes.

3. Perennial debates: Incrementality and syntactic choice

Two particularly persistent debates in the literature on grammatical encoding concern the scope of grammatical planning and the factors influencing syntactic choice.

3.1 Incrementality and the scope of planning

Production is at least to some extent *incremental* (a property termed Wundt's principle by Levelt, 1989). This implies grammatical encoding has two critical characteristics: It creates structure piecemeal and it does so unidirectionally. Basic evidence for incrementality is the influence of *accessibility* on grammatical encoding. Specifically, grammatical encoding processes tend to build sentence structures such that more accessible content words (e.g., ones that have been semantically primed; Bock, 1986a) are mentioned earlier than less accessible content words (more on this below). Given that more accessible words can be processed sooner, it makes sense that they would be mentioned earlier under two assumptions: First, grammatical encoding must create structures piecemeal, otherwise, grammatical encoding processes would have to wait for all parts of the sentence to become accessible anyway. Second, grammatical encoding must operate unidirectionally, otherwise, an accessible content word could be processed sooner even by assigning it to a later sentence position.

Whereas little evidence disputes the unidirectional nature of grammatical encoding (but see Momma, Slevc, and Phillips, 2016, discussed below), evidence for the piecemeal nature of grammatical encoding is less coherent. This is closely tied to the question of the *scope* of grammatical encoding -- with respect to the eventually spoken utterance, how far ahead does the grammatical encoding process specify structure before production begins? The answer to this question is complicated by two factors. First, the scope of grammatical encoding

likely varies by level of encoding, so that as production proceeds from 'higher' levels (e.g., message encoding) to 'lower' ones (e.g., phonological encoding), the scope of encoding narrows (for an elegant demonstration, see Dell, 1986). This can be seen in Garrett's (1975) original model, where the scope of planning at function assignment is a full clause and at constituent assembly is more phrase-like. Additional evidence supporting the idea that the highest levels of grammatical encoding involve a clause-sized scope of planning include patterns of elicited errors of subject-verb agreement (Bock & Cutting, 1992) and of pauses and hesitation during speech (Ford, 1982; Ford & Holmes, 1978). Other evidence shows that speakers can detect upcoming difficulty in a sentence surprisingly early, again suggesting substantial advance planning (F. Ferreira & Swets, 2005). There is also evidence that the scope of planning narrows at later stages of encoding. For example, Meyer (1996) showed that when speakers produce short sentences, semantic distractors related to either subject or object nouns affected initiation times, suggesting that both nouns were semantically planned to some extent. However, phonological distractors only affected initiation times when related to subject nouns, suggesting that only the subject noun was phonologically planned (see also Wheeldon & Lahiri, 1997).

All of that said, other evidence suggests that the scope of planning at early levels of grammatical encoding can sometimes be narrower than the clause. In a pictures-description task, Smith and Wheeldon (1999) showed that speakers began utterances more slowly when the subject noun phrase was complex and

the object noun-phrase simple rather than vice versa. This suggests that more planning occurs for the subject noun phrase than the object noun phrase before utterance onset. Similarly, Griffin (2001) measured speech-onset times and eye-movement patterns during a pictures-description task, and showed that lemma-level properties of names produced in direct object phrases did not affect performance; this suggests that those lemmas were not accessed prior to speech onset. Brown-Schmidt and Konopka (2008) contrasted English and Spanish production, providing evidence suggesting that even production within a phrase can show evidence of incrementality. These different degrees of planning scope, especially for earlier stages of grammatical encoding, probably occur because the degree to which speakers produce sentences incrementally appears to be strategically sensitive. This is illustrated directly by evidence from F. Ferreira and Swets (2002), who showed that speakers produced sentences more incrementally when under a production deadline.

A potentially interesting new wrinkle concerning incrementality comes from Momma et al. (2016) and related work. Momma et al. asked Japanese speakers to describe pictures using either subject-verb sentences (e.g., the Japanese translation of “The dog howls”) or object-verb sentences (e.g., the Japanese translation of “pets the cat,” which is ordered “cat pet” in Japanese; note that in Japanese, arguments such as subjects can be omitted). The pictures had superimposed upon them distractor words that were related in meaning to the verb (e.g., “rub” for “pets the cat”). Results showed that distractors related to the

verb slowed the initiation of picture description – an interference effect regularly observed in production – but only for object-verb sentences and not for subject-verb sentences. Related evidence from English points to analogous results, whereby distractor words semantically related to verbs slow the initiation of passive sentences (i.e., sentences where the thematic object begins the sentence) but not active sentences (i.e., sentences where the thematic subject begins the sentence), and for sentences with unaccusative verbs (ones that begin with a thematic object, like “the ship sank”) but not unergative verbs (ones that begin with a thematic subject, like “the man ran”). Together, this is compelling evidence that speakers do not begin to articulate the thematic objects of sentences until they have selected the lemma for the verb, but they can begin to articulate the thematic subjects of sentences before they select the lemma for the verb. In turn, this may be because thematic objects form an integral part of the predicate that is expressed by a sentence in a way that thematic subjects do not (as illustrated by the fact that aspects of verb meaning are determined by thematic objects, as in the difference between “hit the ball” and “hit the road”).

3.2 Syntactic choice

The way that language works requires sentences that differ in meaning to also differ in form (setting aside ambiguity). Interestingly, the opposite claim is not so -- sentences that differ in form do not always differ in meaning, at least not obviously. For example, the sentences “I know that San Diego is always sunny” and “I know San Diego is always sunny” differ, yet the difference in meaning

between the two is extremely difficult to discover (as illustrated by the fact that papers appear every few years purporting to have done so; see, e.g., Bolinger, 1972; Dor, 2005; Thompson & Mulac, 1991; Yaguchi, 2001). This raises an important question about grammatical encoding: When meaning does not guide speakers to produce one sentence form versus another, what does?

The answer is that many factors seem to affect grammatical encoding relatively independently of meaning. Based on current research, we restrict the present analysis to just three. The first is the above-described syntactic persistence: Given a choice between two roughly meaning-equal syntactic structures, speakers tend to produce structures they have just experienced. Syntactic persistence is evident both in laboratory settings (Bock, 1986b) and in naturalistic production (Szmrecsanyi, 2004), in spoken as well as in written production (Pickering & Branigan, 1998), in isolated production as well as in dialogue (Branigan, Pickering, & Cleland, 2000; Levelt & Kelter, 1982), in English, Dutch (Hartsuiker & Kolk, 1998), German (Scheepers, 2003), American Sign Language (Hall, Ferreira, and Mayberry, 2015) and even from one language to another (Hartsuiker, Pickering, & Veltkamp, 2004; Loebell & Bock, 2003). The reason why speakers persist in their production of syntactic structure is a matter of active debate (see V. S. Ferreira & Bock, 2006, Pickering & Ferreira, 2008), but it likely is motivated by reasons of efficiency (Bock & Loebell, 1990; Smith & Wheeldon, 2001), communication (see below; Pickering & Garrod, 2004), or is a signal of learning (Chang et al., 2006).

The second set of factors relate to the above-described accessibility effects: Given a choice between two roughly meaning-equal structures, speakers tend to produce the one that allows for the earlier mention of more accessible sentence material. The range of factors that condition accessibility effects is impressively broad, including semantic priming (Bock, 1986a), semantic interference (V. S. Ferreira & Firato, 2002), phonological interference (Bock, 1987b), imageability (Bock & Warren, 1985; James, Thompson, & Baldwin, 1973), prototypicality (Kelly, Bock, & Keil, 1986), coreference (V. S. Ferreira & Dell, 2000), and salience or prominence (Prat-Sala & Branigan, 2000). (For additional review, see Bock, 1982; McDonald, Bock, & Kelly, 1993.)

One explanation for accessibility effects is that they make grammatical encoding proceed more efficiently. The idea is that if grammatical encoding is incremental, then producing accessible content sooner allows speakers to dispatch it sooner. This presumably circumvents the need to buffer that accessible content and buys time to access the remaining less accessible content. V. S. Ferreira (1996) provided evidence consistent with this possibility, by showing that speakers produced sentences more efficiently when grammatical encoding had more structural options available.

Under this explanation, accessibility effects might infringe on the influence of perspective meaning on grammatical encoding. That is, recall that speakers will produce different structures depending on what perspective they take on a situation, as represented by the perspective meaning represented in their preverbal message (“San Diego is always sunny” vs. “It’s always sunny in San

Diego"). At least some such effects might be due to raw accessibility rather than to perspective meaning per se. An illustration of this distinction comes from Cowles and Ferreira (2012), who suggested that speakers mentioned one kind of argument earlier in sentences (*given* arguments) because of accessibility, whereas they mentioned another kind of argument earlier in sentences (*topic* arguments) independent of accessibility.

A third set of factors that can influence grammatical encoding relatively independently of meaning are *audience-design* factors. Here, given a choice between two roughly meaning-equal structures, speakers might choose the one that would be easier for their addressee to understand (for review, see V. S. Ferreira & Dell, 2000). The most heavily investigated factor in this set has been ambiguity: All things equal, does grammatical encoding select a less ambiguous rather than a more ambiguous syntactic form?

Evidence concerning the effect of ambiguity on grammatical encoding has been mixed (for review, see Ferreira, 2008). Some evidence has shown that grammatical encoding processes do not preferentially select unambiguous structures, neither in spoken (V. S. Ferreira & Dell, 2000) nor written (Elsness, 1984) production, nor in dialogue (Kraljic & Brennan, 2005), nor with different kinds of structures (Arnold, Wasow, Asudeh, & Alrenga, 2004), nor with prosody (Allbritton, McKoon, & Ratcliff, 1996; Schafer, Speer, Warren, & White, 2000). Other evidence suggests that grammatical encoding might select unambiguous structures in highly interactive dialogue (Haywood, Pickering, & Branigan, 2005),

or in written form (Temperley, 2003), or with prosody (Snedeker & Trueswell, 2003). Keys to sorting out these mixed results likely include taking into account the effects of ambiguity-independent factors on syntactic choice (see V. S. Ferreira & Dell, 2000), and separating nonlinguistic-ambiguity avoidance and linguistic-ambiguity avoidance (V. S. Ferreira, Slevc, & Rogers, 2005).

4. Emerging debates: Rationality, learning, and dialogue

Within grammatical encoding, certain current research threads have special promise for progress, due largely to their tight relationships to areas of active investigation in psycholinguistics or cognitive science. This includes work investigating rationality, learning, and dialogue.

4.1 Rational models of sentence production

Across the behavioral sciences, a new focus of research has been on *rational* or *optimal* models of behavior. Many such accounts are Bayesian in nature. The general idea behind rational accounts of behavior is that when the various contextual factors that are relevant to a behavioral domain are taken into account, individuals' behaviors can be seen as maximizing or optimizing some set of desirable outcomes.

In the domain of sentence production, the most prominent rational account of behavior is likely *Uniform Information Density* (Jaeger, 2010). According to Uniform Information Density, speakers aim to smoothly express an optimal amount of information across an utterance. To do so, they use whatever flexibility is available to them to dilute stretches of language that may be too

informative, or enrich stretches of language that may be too uninformative. In this way, a speaker can avoid overwhelming their listeners' comprehension system by providing too much information in too little time, or wasting time by producing a sequence that could have been more informative.

A straightforward example of such information smoothing arises with the mention of the optional *that* in certain sentence structures. For example, in a *verb-* or *sentence-complement structure*, the main verb in a sentence takes a clausal complement that (in English) can optionally be introduced with the complementizer *that*, as in *The proud mother announced (that) the wedding would be a big event*. As noted above, the complementizer carries little (if any) semantic or relational meaning, but it can convey a valuable bit of information, specifically, that a clausal complement is upcoming (rather than, say, a direct object as in *The proud mother announced the wedding yesterday*). According to Uniform Information Density, such an optional element can be used to smooth the rate of information delivery. In particular, if a comprehender anticipates the arrival of an upcoming clausal complement, then the *that* only contributes redundant information. But if the upcoming clause is less expected, including *that* will reduce the spike in information density that would otherwise arrive when the clausal complement is eventually inferred. Consistent with such a prediction, speakers are in fact less likely to mention the optional *that* after verbs that usually are followed by clausal complements (as in *The proud mother announced the wedding would be a big event*), but are more likely to mention it after verbs that

are rarely followed by clausal complements (as in *The talented photographer accepted the money could not be spent yet*; Ferreira & Schotter, 2013; Jaeger, 2010).

4.2 Ongoing learning effects

With the exception of the Chang et al. (2006) model described above, most investigations of language production up through the 2000s assumed that our linguistic knowledge is relatively static. However, starting in the 2000s, a series of experimental programs have investigated whether ongoing experience has enduring effects on our representation of linguistic knowledge – that is, whether we are constantly learning about the language we experience in a way that affects speakers' production.

One such program has already been described (twice): Syntactic persistence. Although syntactic persistence has been viewed as due to residual accessibility of structural information (Pickering & Branigan, 1998) or as a communicative effect (Pickering & Garrod, 2004), another well-developed approach views it as a learning effect. As mentioned, Chang et al. (2006) presents a comprehensive model of sentence production that explains syntactic persistence as due to learning, an account that has been extended into the rational domain by Jaeger and Snider (2013).

Another form of learning that has been heavily investigated can be termed *phonotactic learning*. First reported by Dell, Reed, Adams, and Meyer (2000), phonotactic learning refers to the fact that producers seem to pick up on the fact

that certain speech sounds can be restricted to certain syllable positions. Such restrictions occur naturally in language. For example, the “h” sound in English can only happen at the beginnings of syllables, and the “ng” sound can only happen at the end. This is not due to physical or biological constraints, as other languages do not have such restrictions (e.g., Vietnamese permits the “ng” sound to begin syllables and [h] appears in syllable-final position in many dialects of Brazilian Portuguese). Dell et al. showed that if speech sounds in English that are normally unrestricted are instead restricted to specific syllable positions, people’s productions will reflect this.

To show this, Dell et al. exploited a speech-error phenomenon called the *syllable-position constraint*: When a speaker slips in their production of a speech sound, the erroneously placed sound will usually be positioned in the intended syllable position (but in the wrong syllable). For the English sounds “h” and “ng,” the syllable-position constraint is observed 100% of the time (so, a speaker intending to say, “the man sang” might say, “the mang…,” but will never say, “the ngan”). For unrestricted English sounds like “s” and “f,” the syllable-position constraint is observed about 70% of the time (so, a speaker intending to say, “the man sang” might say, “the san…” 70% of the time, but “the mas…” 30% of the time). Dell et al. discovered that if speakers are asked to produce sequences of seemingly random syllables, but where normally unrestricted sounds like “s” and “f” are in fact restricted to either the beginnings of ends of syllables – like “h” and “ng” are in English – then like “h” and “ng” in English, when speakers made

speech errors with “s” and “f,” they observed the syllable-position constraint almost 100% of the time. That is, speakers learned a brand new phonotactic restriction as a function of their recent linguistic experience.

4.3 Dialogue

For decades, two lines of work in psycholinguistics have proceeded mostly separately. One, in which the above-described consensus view is situated, views psycholinguistic theorizing as a branch of cognitive psychology, where the nature of general mechanisms is inferred from summary measures of performance during highly controlled tasks. The other line views psycholinguistic performance as language use (Clark, 1996) -- as a set of tools that people use to accomplish goals in socially coordinated fashion. This work relies more on the logic and techniques used in the philosophy of language and linguistic pragmatics, observing and cataloguing performance to analyze language as a system of strategies. Study within this line focuses on *dialogue* contexts in which more than one interlocutor interact, usually in the performance of some game or task. The separation between these lines emerged partly from the heavy emphasis in the 1970s and 1980s on the study of reading -- a socially impoverished setting for language use, to say the least. But valuably, in the last two decades, the increasing prominence of research on language production and on spoken-language comprehension has encouraged a synthesis of these heretofore more independent lines.

Research on grammatical encoding has figured prominently in this synthesis. One relevant angle is the above-described debate concerning the effects of audience design on syntactic choice (see also Brennan & Clark, 1996; Horton & Keysar, 1996; Schober & Brennan, 2003). Another angle that has become relevant to controlled research using dialogue is syntactic persistence. Branigan, Pickering, and Cleland (2000) reported robust syntactic persistence in a laboratory-based dialogue task, and the numerical size of these persistence effects was larger than that observed in previous, monologue-based demonstrations (e.g., Bock, 1986b). Pickering and Garrod (2004) brought this dialogue-based persistence effect together with research on similar semantic coordination effects (Garrod & Anderson, 1987) to propose a broad view of language use as *alignment driven*. The idea is that in dialogue, interlocutors aim to coordinate their use of linguistic devices at all possible levels, so they use corresponding pronunciations, locutions (e.g., Clark & Wilkes-Gibbs, 1986), framing (Garrod & Anderson, 1987), and most ground-breakingly, syntactic structures (Branigan et al., 2000) during conversation. The function of such alignment is to ultimately achieve corresponding *situation models*, which can be considered analogous to preverbal messages in production theories, thereby achieving successful communication. In turn, this alignment approach to linguistic performance has come together with work in cognitive science more broadly on imitation (e.g., Iacoboni et al., 1999), embodiment (where cognitive representation is seen as critically 'external' in nature; e.g., Barsalou, 1999), and

'mirror-neuron' systems (whereby perception and action involve the same neural substrates; e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) to form a distinct but prominent subfield within psycholinguistics. The resulting promise for cross-disciplinary interaction and unification is an extremely valuable strength of this view.

More recently, frameworks that aim to advance understanding of language production in dialogue have adopted insights from theories of action control in cognitive science and cognitive neuroscience (Grush, 2004; Wolpert, 1997). In particular, according to a number of accounts (Pickering & Garrod, 2007, 2013), successful dialogue comes from the coordination of the mechanisms typically thought to underlie language comprehension and production (e.g., semantic and lexical representations). Such coordination has been argued to underlie predictions as to the features of language that will be heard or produced during dialogue, and by comparing the predicted language against what is actually heard or produced, interlocutors in dialogue can monitor whether language was accurately produced, and learn about linguistic features so as to inform future acts of production and comprehension.

5. Fundamental insights

Like any area of active inquiry, research on grammatical encoding is more easily characterized in terms of debates and disagreement than in terms of consensus and agreement. Nonetheless, the field has come a long way in the short forty (or so) years of its current incarnation. Below, we briefly mention two

specific points for which little debate exists in mainstream theories of grammatical encoding, but for which there was at least uncertainty (if not outright rancor) in other areas or in times past.

5.1 Linguistic knowledge and nonlinguistic knowledge are different

Every current approach to grammatical encoding postulates distinct nonlinguistic and linguistic representational systems. Indeed, this separation was vital for the initial growth of the field, so that theories of grammatical encoding could develop without the burden of accounting for the nature of thought more generally. The assumption of linguistic-nonlinguistic separation is not trivial. For language production, even Fodor (1983) once rejected it. In other areas, there are a number of well-known incursions on this assumption that have not managed to get a foothold in accounts of grammatical encoding. For example, the popularity of the Whorf-Sapir hypothesis has ebbed and flowed in the broader study of language over the twentieth century (see Boroditsky, 2001; Lucy, 1992; Whorf, 1956). According to this class of views, the nature of the linguistic devices offered by a language critically determine the thought patterns of those who use that language. Yet, approaches to grammatical encoding have generally found it useful to postulate distinct representational systems for conceptual constructs versus linguistic constructs (although a valuable middle ground comes from Slobin's [1996] "thinking-for-speaking" approach and related work). Similarly, views of psychological performance deriving from the behaviorist perspective (Skinner, 1957) aimed to reduce grammatical patterns to

patterns of instrumental responses ingrained by reinforcement and punishment contingencies. Some connectionist and parallel-distributed-processing frameworks (Rumelhart & McClelland, 1986) could be viewed as neobehaviorist in nature, yet it is notable that connectionist accounts of grammatical encoding of any comprehensiveness (e.g., Chang, 2002; Chang et al., 2000, 2006; Dell, 1986) involve a much richer and structured cognitive architecture than comparably comprehensive accounts of, say, single-word reading (Plaut et al., 1996; Seidenberg & McClelland, 1989). Finally, the above-mentioned embodied approaches to cognition (e.g., Barsalou, 1999) promise a different way to blur the distinction between language and thought, namely by driving at least the perceptual characteristics of language into thought. Nonetheless, the account of grammatical encoding that is most embodied in nature (Pickering & Garrod, 2004) still includes independent and distinct representational systems for thought and for language. In short, among students of grammatical encoding, it is almost universally held that thinking and talking are different, and so are based on distinct systems of representation.

5.2 Syntax is in there somewhere

A constant tension in approaches to language acquisition and language comprehension is the status of syntactic representations. Some approaches (e.g., Frazier, 1988; Pinker, 1989) view syntactic knowledge as the irreducible basis of our grammatical knowledge (even if, of course, non-syntactic knowledge can be bootstrapped to acquire it). Others (e.g., MacDonald, Pearlmutter, &

Seidenberg, 1994; Tomasello, 2000) view syntactic knowledge as derived from or reducible to other forms of knowledge, including conceptual and perceptual knowledge.

Among approaches to grammatical encoding, this tension is far less prominent, largely because some form of syntactic knowledge is seen as fundamental to how grammatical encoding works (as represented by the above described consensus model). Three lines of empirical work have led to this standpoint. The first comes from the speech-error observations that pioneered research on language production (Fromkin, 1971, 1973; Garrett, 1975). Specifically, it is notable that most speech-error investigations explore the fact that errant productions maintain their syntactic integrity, even when semantic integrity is compromised (for discussion, see Bock, 1990). For example, because about 85% of word-exchange errors involve exchanging words that belong to the same grammatical categories (Garrett, 1975; Stemberger, 1985), the syntactic structures of errant utterances will conform to speakers' intentions (and will be well formed), even when their meanings do not (e.g., "that *log* could use another *fire*," V. S. Ferreira & Humphreys, 2001; "she *sings* everything she *writes*," Garrett, 1975). The second line is syntactic persistence: Most early work on syntactic persistence (see especially Bock, 1986b, 1989; Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992) determined that syntactic contributions to persistence are separate from conceptual, semantic, lexical, or phonological contributions. Thirty years later, the research landscape suggests

that nonsyntactic factors seem to influence syntactic persistence either independently of syntactic factors (e.g., Bock et al., 1992; Pickering & Branigan, 1998), or only when syntactic factors are neutralized (see especially Chang, 2002; Griffin & Weinstein-Tull, 2003). The third line is work on the production of agreement (e.g., in English, verbs agree with the grammatical number of their subject). Specifically, patterns of agreement errors show that performance is heavily influenced by grammatical features (see Bock, 1995a) and hierarchical representation (e.g., Franck, Vigliocco, & Nicol, 2002), with nonsyntactic influences (see Haskell & MacDonald, 2003; Thornton & MacDonald, 2003) of limited scope (Eberhard, Cutting, & Bock, 2005). Together, observations like these suggest that syntactic structures form the foundation of spoken utterances, in accordance with the approach described above.

6. Summary

How and why do speakers say what they say? The consensus model that opened this chapter provides a sketch of how: Independent but mutually influential component systems that process structure and content proceed through stages of selecting linguistic features and then retrieving their details. How staged these processes are and where the line should be drawn between structure and content are subjects of active debate. The remaining debates outlined in this chapter provide a sketch of why speakers say what they say: In addition to the expression of meaning, speakers' utterances are influenced by incrementality of processing, the accessibility or persistence of linguistic features,

audience design, rational behavioral goals, ongoing learning, and influences during dialogue. Ongoing research will play out these debates, resolving some and spawning others. Through all of this, this research trajectory is providing fundamental insights into the way that language works.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 714-735.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory & Language, 51*, 55-70.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences, 22*(4), 577-660.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1-47.

Bock, J. K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575-586.

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

Bock, J. K. (1987a). Coordinating words and syntax in speech plans. In A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 3, pp. 337-390). London: Erlbaum.

Bock, J. K. (1987b). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*, 119-137.

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

Bock, J. K. (1990). Structure in language: Creating form in talk. *American Psychologist, 45*, 1221-1236.

Bock, J. K. (1995a). Producing agreement. *Current Directions in Psychological Science, 8*, 56-61.

Bock, J. K. (1995b). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition. Vol 11: Speech, language, and communication* (pp. 181-216). Orlando, FL: Academic Press.

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

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150-171.

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

Bolinger, D. (1972). *That's that*. The Hague: Mouton.

Boroditsky, L. (2001). Does language shape thought? Mandarin and english speakers' conceptions of time. *Cognitive Psychology, 43*(1), 1-22.

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482-1493.

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*, 274-280.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177-208.

Caramazza, A., Bi, Y. C., Costa, A., & Miozzo, M. (2004). What determines the speed of lexical access: Homophone or specific-word frequency? A reply to jescheniak et al. (2003). *Journal of Experimental Psychology-Learning Memory and Cognition, 30*(1), 278-282.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*(6), 1430-1450.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26*, 609-651.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming Syntactic. *Psychological Review*, *113*, 234-272.

Chang, F., Dell, G. S., Bock, J. K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research, 29*(2), 217-229.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Cowles, H. W., & Ferreira, V. S. (2012). The influence of topic status on written and spoken sentence production. *Discourse processes*, *49*, 1-28.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*(2), 318-344.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. *Psychological Review, 98*(4), 604-614.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech Errors, Phonotactic Constraints, and Implicit Learning: A Study of the Role of Experience in Language Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1355-1367.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997b). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801-838.

Dor, D. (2005). Toward a semantic account of that-deletion in english. *Linguistics, 43*(2), 345-382.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review, 112*(3), 531-559.

Elsness, J. (1984). *That* or zero? A look at the choice of object clause connective in a corpus of american english. *English Studies., 65*, 519-533.

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, US: Psychology Press/Taylor & Francis.

Ferreira, F., Lau, E., & Bailey, K. (2004). *A model of disfluency processing based on tree-adjoining grammar.* Paper presented at the Seventeenth Annual CUNY Conference on Human Sentence Processing, College Park, MD.

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. Hove, UK: Erlbaum.

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. (in press). How are speakers' linguistic choices affected by ambiguity? In A. S. Meyer, A. Krott & L. R. Wheeldon (Eds.), *Language processes and executive function.* London: Psychology Press.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation*, *49*, 209-246.

Ferreira, V. S., & Bock, J. K. (2006). The functions of structural priming. *Language and Cognitive Processes*, *21*, 1011-1029.

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., & Firato, C. E. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review, 9*, 795-800.

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

Ferreira, V. S. & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *Quarterly Journal of Experimental Psychology*, *66*, 1548-1571.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263-284.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

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

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*. Cambridge, MA: MIT Press.

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

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in french and english: The role of syntactic hierarchy. *Language & Cognitive Processes, 17*(4), 371-404.

Frazier, L. (1988). Grammar and language processing. In F. J. Newmeyer (Ed.), *Linguistics: The cambridge survey (iii) linguistic theory: Extensions and implications* (pp. 15-34). Cambridge: Cambridge University Press.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Fromkin, V. A. (Ed.). (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, pp. 133-177). New York: Academic Press.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The cambridge survey, iii: Language: Psychological and biological aspects* (Vol. 3: Language: Psychological and biological aspects, pp. 69-96). Cambridge: Cambridge University Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in a conceptual and semantic co-ordination. *Cognition, 27*, 181-218.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1-B14.

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

Grush, R. (2004) The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences 27*, 377–396.

Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2015). Syntactic Priming in American Sign Language. *PloS one*, *10*(3), e0119611.

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic persistence in dutch. *Language & Speech, 41*(2), 143-184.

Hartsuiker, R. J., Kolk, H. H. J., & Huiskamp, P. (1999). Priming word order in sentence production. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52*(1), 129-147.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in spanish-english bilinguals. *Psychological Science, 15*(6), 409-414.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*(2), B27-B39.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal Of Memory And Language, 48*(4), 760-778.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguity during dialogue? *Psychological Science, 16*, 362-366.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528.

Indefrey, P. & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101-144.

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

Jaeger, T. F. & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57-83.

James, C. T., Thompson, J. G., & Baldwin, J. M. (1973). The reconstructive process in sentence memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 51-63.

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on caramazza, costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology-Learning Memory and Cognition, 29*(3), 432-438.

Joshi, A. K., Levy, L., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences, 10*, 136-163.

Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language, 25*, 59-74.

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

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

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

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*, 78-106.

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. Levelt (Eds.), *Crossing the boundaries in linguistics* (pp. 221-252). Dordrecht: Reidel.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences, 22*(1), 1-75.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics, 41*(5), 791-824.

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge, England: Cambridge University Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703.

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

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 813-824.

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

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory & Language, 46*(3), 586-605.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, *134*, 427-459.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-226.

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

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

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in english and spanish. *Journal of Memory & Language, 42*(2), 168-182.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460-499.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131-141.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on caramazza and miozzo (1997). *Cognition, 69*, 219-230.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of english verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel*

distributed processing (vol. 2): Psychological and biological models (pp. 216-271). Cambridge, MA: MIT Press.

Schabes, Y., Abeille, A., & Joshi, A. K. (1988). *New parsing strategies for tree adjoining grammars.* Paper presented at the 12th International Conference on Computational Linguistics.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*(2), 169-182.

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

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser & M. A. Gernsbacher (Eds.), *Handbook of discourse processes* (pp. 123-164). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Skinner, B. F. (1957). *Verbal behavior*. New York, NY: McGraw-Hill.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge, England: Cambridge University Press.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*(3), 205-246.

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production: An online study. *Cognition, 78*(2), 123-164.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory & Language, 48*(1), 103-130.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). London: Erlbaum.

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 484-503.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex*, *20*, 912-928.

Szmrecsanyi, B. (2004). *Persistence phenomena in the grammar of spoken english.* Unpublished Ph. D. dissertation, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.

Temperley, D. (2003). Ambiguity avoidance in english relative clauses. *Language, 79*, 464-484.

Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer *that* in conversational english. *Journal of Pragmatics, 15*, 237-251.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal Of Memory And Language, 48*(4), 740-759.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Ungerleider, L. G., & Haxby, J. V. (1994). 'what' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*, 157-165.

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition, 68*(1), B13-B29.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37*, 356-381.

Whorf, B. L. (1956). The relation of habitual thought and behavior to language. In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of benjamin lee whorf*. Cambridge, MA: MIT Press.

Wolpert, D. M. (1997) Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*, 209–216.

Yaguchi, M. (2001). The function of the non-deictic that in English. *Journal of Pragmatics, 33*(7), 1125-1155.