The Relationship between Parsing and Generation

Shota Momma[1] and Colin Phillips[2]

shotam2@gmail.com, colin@umd.edu

[1]University of California, San Diego
[2]University of Maryland

ABSTRACT

Humans use their linguistic knowledge in at least two ways. On one hand, they use their linguistic knowledge to convey what they mean to others or to themselves. On the other hand, they use their linguistic knowledge to understand what others say or they themselves say. In either case, they must assemble the syntactic structures of sentences in a systematic fashion, in accordance with the grammar of their language. In this article, we advance the view that a single mechanism for building sentence structure may be sufficient for structure building in comprehension and production. We argue that differing behaviors reduce to differences in the available information in either task. This view has broad implications for the architecture of the human language system, and provides a useful framework for integrating largely independent research programs on comprehension and production by both constraining the models and uncovering new questions that can drive further research.

KEYWORDS
Parsing, generation, syntax, sentence processing

SUMMARY POINTS

1. Aligning the parser and the generator is a prerequisite for aligning theories of grammar and language processing.
2. We synthesize existing proposals and challenges regarding the relationship between parsing and generation.
3. We clarify what it means for the parser and the generator to be the same, and we identify a series of properties of structure building mechanisms that could match or mismatch in parsing and generation.
4. We argue that fundamental properties of parsing and generation mechanisms are the same, and that differing behaviors reduce to differences in the available information in either task.

1. Introduction

Humans use their linguistic knowledge in at least two ways. On one hand, they use their linguistic knowledge to convey what they mean to others or to themselves. On the other hand, they use their linguistic knowledge to understand what others say or they themselves say. In either case, they must assemble the syntactic structures of sentences in a systematic fashion, in accordance with the grammar of their language. In this review, we discuss the benefits of a single model of structure building for comprehension and production (Section 2), we synthesize previous proposals (Section 3) and challenges (Section 5) for relating comprehension and production mechanisms. We argue that it is plausible to pursue a single model (Section 4)

To avoid confusion, we should first clarify what we do not intend to argue for in this review. We do not claim that the cognitive systems for comprehension and production are identical in their entirety. Such a view would require that the cognitive mechanisms for moving our lips and tongues are involved in comprehension, and that the cognitive mechanisms for analyzing sound waves or visual patterns are critically involved in production. Although related claims are sometimes advanced under the heading of *embodied cognition* (e.g., Pickering & Garrod 2007; 2013 among others), we remain agnostic about such a view. Thus, despite superficial similarities, our proposal should not be confused with that view, and we will not review the literature on that topic, e.g., evidence for/against the involvement of motor control mechanisms in sentence comprehension. Our primary concern is whether the cognitive mechanisms syntactic structure building mechanism is shared between comprehension and production. Thus, we henceforth use the terms *parsing* and *generation* to refer specifically to the structural processes involved in comprehension and production, respectively.

Although our claims are restricted to identity of structure building mechanisms in comprehension and production, we maintain that this has far-reaching consequences for the architecture of the human language system, as we discuss next.

2. Why the problem is important

Understanding the relation between models of parsing and generation is central to the larger project of connecting high-level linguistic theories with psycholinguistic theories. In relating psycholinguistic and linguistic theories it is important to distinguish three dimensions: tasks, levels of analysis, and mechanisms.

<u>Tasks</u>: It is clear that we carry out different tasks using our linguistic knowledge. When we speak we start with an intended meaning and try to find an external form that conveys that meaning. When we understand we start with an external form and try to identify the intended meaning. When we make acceptability judgments we take a form-meaning pairing and try to identify whether they are connected. All three tasks have a shared subgoal of building a

structure that links the form and the meaning. They differ only in what information is provided as a starting point.

Levels of analysis: The same cognitive system can be described at multiple levels of analysis. Traditional linguistic theories are theories of mental representations, but they abstract away from details of timing, how those representations are encoded in memory, neural connectivity, etc. It is common to appeal to the three levels proposed by Marr (1982) for analyses of visual systems, but Marr's levels do not correspond closely to standard practice in linguistics and psycholinguistics. It is probably more accurate to view levels of analysis on a continuum, with as many different levels as there are details that can be included or abstracted away from.

Mechanisms: The distinctions among tasks and among levels of analysis are necessary, and accounts of speaking, understanding, and acceptability judgment tend to focus on different tasks and on different levels of analysis. But this does not amount to the claim that distinct cognitive mechanisms build the same representations for different tasks, i.e., a parser that builds structures in comprehension, a generator that builds structures in production, and a grammar that is shared between these, and also plays a central role in acceptability judgments. One can distinguish tasks and levels of analysis without committing to distinct mechanisms.

In other work we have argued that is beneficial and feasible to treat the structure building system used in comprehension and the grammar as the same cognitive system (Phillips 1996; Phillips & S. Lewis 2013). Evidence that the distinction is unnecessary comes from two broad domains. First, rapid structure building processes during comprehension are highly grammatically sensitive (Phillips et al. 2011; S. Lewis & Phillips 2015). Second, there may be benefits for standard grammatical analysis from assuming a structure building system that proceeds in the same order as comprehension and production processes (Bianchi & Chesi 2014; Kempson et al. 2001; Phillips 2003; Shan & Barker 2006; Steedman 2000). However, there is little benefit in conflating the parser and the grammar if the parser and the generator cannot also be conflated. In fact, if the parser and the generator are distinct cognitive mechanisms that assemble the same representations, then it is necessary to have a third cognitive mechanism that defines the possible representations, i.e., a grammar. In other words, the choice is between one cognitive mechanism or three. Two is not an option.

3. Previous proposals

We are certainly not the first to claim that parsing and generation share the same mechanisms. Here we review previous claims in psycholinguistics and neurolinguistics about the relationship between comprehension and production. We situate our current account in relation to two broad classes of claims.

3.1 Interactionist view

Many previous claims about the comprehension-production relationship can be classified as what we call the *interactionist view*. This view claims that comprehension mechanisms and production mechanisms are distinct, but they interact heavily during a single act of understanding or speaking.

Levelt (1983) raised the possibility that self-monitoring during speaking is carried out by comprehension mechanisms. This contrasts with the claim that self-monitoring is carried out by a separate mechanism internal to the production system (see Postma 2000 for review). This is an interactionist claim in the respect that some sub-tasks of production are performed by a generator-independent parser. The empirical status of this claim is beyond the scope of this review. See Hartsuiker & Kolk (2001) and Huettig & Hartsuiker (2010) for an overview.

The interactionist hypothesis regarding self-monitoring concerns the use of the parser during production tasks. Some have discussed possibilities in the other direction, i.e., the use of generation mechanisms during comprehension tasks. For instance, Federmeier (2007) argued, based on evidence that the left-hemisphere is critical in generating lexically specific prediction, that linguistically fine-tuned prediction is the product of production mechanisms, which arguably reside in the left-hemisphere. Similarly, Pickering & Garrod (2007; 2013) proposed an interactive model of language processing in which parser-independent "production implementers" generate an internal representation of yet-to-be-heard elements of the sentence. Dell & Chang (2014) also claimed that their simple recurrent neural network model predicts upcoming words in sentence comprehension, in much the same way as this system produces sentences.[1] Aside from the claim that prediction is a function of production mechanisms, Garrett (2000) discussed the possibility that the generator plays an important role in structural ambiguity resolution in sentence comprehension. It does this by taking lexical elements identified by the comprehension system and combining them with discourse knowledge to generate structures that are plausible in the current context. Those candidates can be used to filter the candidate structures proposed by the parser. All of these claims are interactionist in the sense that they all assume a parser-independent generator.

There is a different kind of interactionist claim that we call *external interactionism*. This is the claim that comprehension and production interact with each other via the external world, namely via the corpus that speakers generate and that comprehenders experience (Gennari & MacDonald 2009; MacDonald 2013). The idea is that speakers have certain biases in producing certain sentences, e.g., animate entities tend to be mentioned first, which skews the distributional patterns in speech corpora. This pattern is learned by comprehenders, in such a way that rarer patterns are more difficult to comprehend than more frequent patterns. We agree that something like this plausibly happens, but this account establishes the link between comprehension and production via the external world, not within an individual's mind.

[1] Here we highlight this parallel between Dell and Chang's model and other interactionist proposals. However, in other respects their proposal is like a single mechanism model.

3.2. Single mechanism view

Another class of claims is that the cognitive mechanisms that are responsible for building the structural representation of sentences are identical. We call this view the *single mechanism* view. The possibility of shared mechanisms has been repeatedly mentioned in the literature.

Kempen (2000) argued that comprehension and production processes have much in common. For example, both speakers and comprehenders are susceptible to agreement attraction (see Section 4.3.2 below). Based on such similarities, he suggested that the mechanisms for grammatical encoding (i.e., generation) and decoding (i.e., parsing) are shared. Kempen et al. (2012) also offered empirical evidence that speakers override comprehended sentence structures with produced sentence structures when paraphrasing. They compared the response latency of anaphoric pronouns in a paraphrasing task, in which participants were asked to paraphrase direct speech (e.g., *The angry headmaster complained: "I have seen a nasty cartoon of myself in the hall*) into indirect speech (e.g., *The angry headmaster complained that he has seen a nasty cartoon of himself in the hall*). They found that participants did not slow down at the ungrammatical reflexive *himself* when the direct speech sentences contained a mismatching reflexive (e.g., *The angry headmaster complained: "I have seen a nasty cartoon of himself in the hall*), despite readily detecting the same anomaly in a self-paced reading task. They suggested that the syntactic representation of what participants understood uses the same workspace as the syntactic representation of what they said. Kempen (2014) subsequently proposed a model that handles structure building processes in comprehension and production with a single processor.

Finally, it is worth noting similar attempts to integrate the parser and the generator in computer science. The general framework is termed *bi-directional grammar* (e.g., Appelt 1987). Although this approach only commits to the position that the grammar but not necessarily the parser-generator can be integrated, it has also been suggested that the architecture of the processing mechanism can also be integrated (Shieber 1988).

4. Current proposal

The view advanced here is a version of the single mechanism view. However, as it currently stands, the claim that parsing and generation mechanisms are identical remains somewhat vague. We must first clarify what it means for the parser and the generator to be identical. To this end, we classify identity claims into three levels.

4.1. Levels of identity

There are at least three levels at which one can maintain the parser-generator identity claim. First is the claim that the parser and the generator share the same grammar. This is not especially controversial, as what a speaker can understand and say normally obey the same grammatical constraints. We call this *representation-level identity*: the output representation of parsing and generation is the same. In the vast majority of cases, sentences that are judged to be unacceptable in comprehension are also avoided in production (but see Section 5.2 for some discussion). Experimentally, structural priming occurs from comprehension to production (Bock et al. 2004) and from production to comprehension (Kim et al. 2014), suggesting that whatever is primed is shared between comprehension and production. Additionally, neuroimaging evidence suggests that overlapping neural resources are used during comprehension and production (Menenti et al. 2011; Segaert et al. 2011). Thus, at least the representations constructed during understanding and speaking are overwhelmingly the same.

Second, not only the representations but also the mechanisms by which those representations are built during speaking and understanding could be identical. We call this *mechanism-level identity*.

Third, the procedures for building sentence structures could be identical between speaking and understanding. We call this *algorithm-level identity*. Parsing and generation algorithms might specify which parts of a sentence representation are created first, second, etc. Under the algorithm-level identity view, this order matches between parsing and generation. In principle, algorithm-level identity could hold without mechanism-level identity being true. But we consider this unlikely, so here we assume that algorithm-level identity entails mechanism-level identity.

These three levels of identity claims are progressively harder to maintain, in the order discussed, as the relationship between the parser and the generator gets tighter. As we have noted, the tighter the relationship between the parser and the generator, the more plausible the transparency between grammar and real-time structure building. Below we evaluate the mechanism-level and algorithm-level identity claims, taking the representation-level identity as given.

4.2 Evaluating the mechanism-level identity view

We evaluate the mechanism-level identity view by breaking it into several testable questions that we can address using existing research. The list is by no means exhaustive, but not random either. Most properties that we focus on are research topics in their own right in psycholinguistics. Our aim is to show that the mechanism-level identity view is testable, plausible, and relevant to existing psycholinguistic research.

4.2.1 Long-term memory access

Starting from a relatively simple and well-studied process, lexical access is unequivocally involved in parsing and generation. It is often suggested that lexical access is based on content-addressable memory mechanisms (e.g., connectionist models of lexical access like McClelland & Elman 1986 in comprehension; Roelofs 1992 in production; but see Forster 1992 for the involvement of a serial mechanism). The mechanism-level identity view claims that this mechanism is identical in parsing and generation.

How can a memory mechanism for comprehension and production be the same when the input to each is different in kind? A key property of content-addressable memory mechanisms is that items in memory are accessed in parallel by matching the featural contents of memory items and retrieval cues. This removes the need to assume distinct mechanisms for lexical access in comprehension and production. To see this, consider how a single content-addressable memory mechanism can interact with different modalities in other domains of cognition, using the case of recalling memories of real world events. This task involves accessing episodic memory, and it can be accomplished using various cues, including visual, auditory, and olfactory cues, or even more abstract cues like a person, place, or thing that are part of the target episodic memory. Despite the many different kinds of input, we do not conclude that there are distinct memory mechanisms for each type of cue. This is the sense in which a single memory mechanism can use different kinds of inputs.

Furthermore, it is not accurate that the inputs to lexical access are entirely different in speaking and understanding. There is evidence that multiple types of cues can be used in parsing and generation. For example, as noted by Pickering & Garrod (2007) lexical access in generation is helped by the existence of phonological cues associated with the target item. This is also shown in picture-word interference studies, where seeing phonologically similar words helps speakers produce the target word faster, both in isolated production tasks (e.g., Lupker 1979; Schriefers et al. 1990 among others) and in sentence-level production tasks (e.g., Meyer 1996). Meanwhile, lexical access in parsing can benefit from semantic cues, just like in generation. This can be seen in lexical priming effects (e.g., Neely 1977) and in lexical prediction effects using event related potentials (DeLong et al. 2005; Van Berkum et al. 2005) or visual-world eye-tracking (Altmann & Kamide 1999 among many others).

There is also suggestive evidence that syntactic cues are used for lexical access in similar ways. In comprehension Wright & Garrett (1984) showed that lexical recognition is facilitated when the syntactic category of the target word is predictable from the prior syntactic context, independent of meaning. In production it is well known that that substitution and exchange errors rarely occur between words of different syntactic categories (Nooteboom 1973). In an experimental context, Momma et al. (2016) showed that the semantic interference effect in picture naming that is normally observed between semantic associates (e.g., *walking* and *running*) disappears when the two associates mismatch in syntactic category (e.g., *walking* as a noun and *running* as a verb). They showed this using a novel “sentence-picture interference” task. This suggests that lexical competition is restricted to a specific syntactic category: speakers are able to restrict their retrieval to nouns when accessing a noun.

In sum, we have evidence that phonological, semantic and syntactic cues are used for lexical access both in comprehension and production when available. This follows from the view that a single content-addressable memory mechanism is involved in lexical access in parsing and generation.

The argument that we applied to lexical access may also hold for units larger than a single word. It is reasonable to assume that memory encodings of structural configurations larger than a word are stored and accessed in parsing and generation, as in many computational models of parsing (e.g., R. Lewis & Vasishth 2005; MacDonald et al. 1994; St John & Mcclelland 1990; Vosse & Kempen 2000). Just like lexical memory, structural memory may be associated with semantic and phonological cues as well as syntactic cues. For instance, the double object dative construction (e.g., *John gave him an apple*) might be associated with semantic features such as transfer of possession and completion of transfer events (e.g., Gleitman 1990). Under this view, structural memories may have semantic content independent of the lexical items that fill them. Via content-addressable memory access, structural memories could be retrieved using semantic content as a cue, both in comprehension and production. The same goes for phonological cues. For example, certain items in structural memory may be associated with prosodic cues, due to systematic correspondence between syntactic and prosodic structures. An empirical advantage of this view is that the rapid use of message-level information and prosodic information in structural processing is expected in both parsing and generation. The rapid use of message-level information in parsing is well documented (e.g., Tanenhaus et al. 1995 among many others), and the rapid use of semantic cues in generating syntactic structure is what speakers do all the time. Also, predictive processes in comprehension are likely based on event memory representations, which presumably share a representational format with message representations (e.g., Chow et al. 2016; Kim et al. 2016). Likewise, the rapid use of prosodic cues to select an appropriate syntactic structure is also documented in parsing (e.g., Snedeker & Trueswell 2003 among many others), although it is unclear how prosodic cues can be used for structure generation in speaking. Syntactic cues are clearly used during both parsing and generation. For example, projection of phrase structure can be understood as the process of accessing phrasal nodes in memory using syntactic category as a retrieval cue. The verb's structural information (subcategorization) has also been shown to rapidly affect parsing (e.g., Kim & Trueswell 1998) and generation (e.g., Melinger & Dobel 2005; Pickering & Branigan 1998), and this also can be understood as retrieving relevant structural memories using the verb's subcategorization features as retrieval cues. In sum, memory for words and for structures may be accessed via phonological, syntactic, and semantic cues when they are available. Thus, a single content-addressable memory mechanism may underlie lexical and structural memory access in parsing and generation.

4.2.2. Eagerness of structural integration

Once a structural building block is retrieved from memory, it needs to be integrated into the developing representation of a sentence. In parsing, models that immediately integrate newly accessed items are described as *strongly incremental* (e.g., Crocker 1996; Lombardo & Sturt 2002; Schneider 1999; Sturt & Lombardo 2005). Those incremental models maintain a connected structural representation throughout the comprehension process. This is in contrast to some earlier parsing models like Marcus (1980) that employed a look-ahead buffer. If the mechanism-level identity view is correct, we should expect parsing and generation to be equally 'eager' in how structural integration occurs.

In evaluating this view it is important to note that the term *incrementality* is commonly used in production research, but often in a different sense than in the parsing literature. In production, the term *incrementality* often is used to refer to the degree of synchronization between planning and articulation processes (e.g., F. Ferreira & Swets 2002; Iwasaki 2011; Konopka 2012; Norcliffe et al. 2015; Wagner et al. 2010), and is closely associated with cascading of information across different levels of processing (De Smedt 1996; Kempen & Hoenkamp 1987;

Levelt 1989). This is not the same as the notion of incrementality used in parsing. Incrementality in parsing refers to the speed and grain-size of integration and updating within a level of representation.

In production, it is intuitively the case that speakers do not always utter words immediately upon planning them. Speakers can certainly plan an entire sentence before starting to say it if they so choose. Experimentally, the scope of advance (lexical) planning in speaking has been shown to vary across studies (Allum & Wheeldon 2007; 2009; Griffin 2001; Meyer 1996; Smith & Wheeldon 1999; Wheeldon & Smith 2003; among others) and even within an individual study, suggesting that how long we wait to utter a planned word is under strategic control and/or modulated by resource availability (F. Ferreira & Swets 2002; Konopka 2012; Wagner et al. 2010). However, this does not mean that the generator is less incremental than the parser. The relevant question is whether the generator rapidly integrates retrieved words/structures to maintain a connected structure, in the same way that the incremental parser rapidly integrates incoming words into the structural parse. There is little existing evidence on this issue, but we take it as a reasonable initial assumption that parsing and generation closely parallel one another in this regard.

4.2.3. Grain size of structure building unit

After building some structure, comprehenders and speakers use it to understand and speak. In both comprehension and production, structures must expand by a certain increment, and the size of this increment determines how much structure speakers/comprehenders build before they use it for understanding or speaking. Under the mechanism-level identity view, the grain size of the units by which structures expand should match between parsing and generation.[2] The grain size can be studied by examining how frequently structure building and post-structural processes (interpretation and articulation) can interleave in comprehension and production.

Modern parsing models normally assume that little needs to be known before interpretation can start. Interpretation of a partial input can and often does occur immediately as each word is received (Kamide et al. 2003; Kutas & Hillyard 1980; Marslen-Wilson 1973 among others). Intuitively, the parser does not need to know an entire clausal structure before it can start to interpret the sentence. Thus, the grain size of structure building units in parsing is usually considered to be rather small, definitely smaller than a clause.

In contrast to most modern comprehension models, some influential production models assume that speakers plan a lot of structure before they can start articulation (e.g., Boomer 1965; Bock & Cutting 1992; F. Ferreira 2000; Fodor et al. 1974; Ford & Holmes 1978; Griffin & Weinstein-Tull 2003). For instance, Garrett (1980; 1988) assumes that the abstract representations of words (lemmas) along with their grammatical roles (subject, object, verb, etc) are all specified before phonological encoding initiates. More recently, F. Ferreira (2000) proposed a production

[2] In production research, this question is often asked in the context of whether sentence production is incremental (e.g., Levelt 1989; De Smedt 1996; Wagner et al. 2010). But we reserve the term incrementality to mean the immediate integration of retrieved words into grammatical structures, to draw a closer parallel between parsing and generation.

model based on lexicalized tree-adjoining grammar (see, e.g., Joshi & Schabes 1997), in which basic clausal structure along with a verb head is specified before the initiation of phonological encoding. This follows a long tradition in production research in which the ‘performance unit’ is assumed to be a clause rather than a word (see Bock & Cutting 1992 for a historical overview).

Based on previous research, we can be confident that comprehenders do not need to know (or at least do not need to be certain about) the upcoming sentence structure to initiate interpretive processes (Aoshima et al. 2009; Kamide et al. 2003; Kutas & Hillyard 1980; Marslen-Wilson 1973). In production, however, we still know relatively little about how much structure speakers need to plan before articulation is initiated.[3] Certainly, there are models that assume a small grain size for structure building units (e.g., de Smedt 1996; Levelt 1989). However, much previous work on sentence planning has focused on whether a particular word that occurs downstream in a sentence is accessed before the first word of the sentence is articulated, and relatively little attention has been paid to the question of whether structural aspects of the sentence need to be planned ahead (but see Wheeldon et al. 2013 for some relevant work). This issue is important in answering the grain size question. If a large chunk of structure is planned without the words that fill it, then whether a particular downstream word is planned in advance is not informative about the size of structural unit. Consequently, existing evidence about lexical look-ahead does not necessarily tell us whether the structural representation needs to be planned ahead. Indeed, there is evidence that suggests that structural aspects of a sentence are planned prior to lexical access (Dell et al. 2008; V. Ferreira & Humphreys 2001; Momma et al. 2016; Nooteboom 1973).

We recently tested whether speakers indeed plan basic clausal structure before sentence onset (Momma et al. 2017), as predicted by models of production that assume a large grain size (Bock & Cutting 1992; Boomer 1965; F. Ferreira 2000; Fodor, Bever & Garrett 1974; Ford & Holmes 1978). We tested this by combining structural priming with measurements of changes in speech rate throughout a sentence. Normally, structural priming is measured as an increase in the proportion of choosing one structure over another given a structural choice (Bock 1986; see Pickering & V. Ferreira 2008 for a review). For instance, if a speaker hears/says a prepositional dative (PD) sentence like *John gave an apple to Mary* before describing a picture involving a ditransitive event, he is more likely to use a PD construction in his subsequent description (e.g., *Bill sent a letter to his parents*). In contrast, we measured the effect of structural priming on speech rate during the production of ditransitive sentences (see Wheeldon & Smith 2003 and Segaert et al. 2016 for studies that show effects of structural priming on speech onset latency). We expected that repeating the PD structure would speed up production, but the question was not whether but when in the sentence this speed up effect would occur. If speakers plan the

[3] Gleitman et al. (2007) showed that, in an event picture description task, perceptual attention modulates the choice of sentence structure, e.g., speakers are more likely to say a passive sentence when the patient entity in a transitive picture is perceptually cued. Some researchers suggest that this offers strong evidence for word-by-word production with little or no structural look-ahead (e.g., Bock & V. Ferreira 2014; Konopka 2012). We disagree with this interpretation. The evidence suggests an early influence of conceptual processes on linguistic formulation (as Gleitman et al. 2007 already argued), but it does not argue against advance planning of structure beyond the initial part of the sentence. These are logically independent issues.

structure of the verb phrase prior to utterance onset, then they should be faster to begin saying a sentence with the same verb phrase structure. If speakers instead construct the verb phrase structure while they are uttering the preceding part of the sentence, then they should speed up in the middle of the sentence, at the point when they have to make a decision about the structure of the verb phrase. The results show that speakers are not faster to begin uttering the sentence, but they are faster to say the verb itself when the ditransitive structures are primed. Assuming that the timing of priming effects corresponds to the timing of planning, this suggests that speakers can and do start saying the verb itself without planning the internal structure of verb phrase. This points to a rather small grain size of structure building in generation, consistent with a standard property of modern parsing models.

4.2.4. Structural decision making

One of the primary tasks for the parser is to resolve structural ambiguity in the input. This is not the case for the generator. Speakers generally know what they want to say. How is it possible that the parser and the generator are the same given this difference? Again, a task difference need not imply a mechanism difference. Although generation does not face a message ambiguity problem, both the parser and the generator have to make structural decisions. The parser needs to choose a structure from among structural alternatives that are consistent with the sound input, while the generator needs to choose a structure from among structural alternatives that are consistent with the same message (e.g., when choosing among active vs. passive structures or double object vs. prepositional object structures). In other words, the parser and the generator share the same computational goal of selecting a structural representation that is consistent with the input. The question is whether the same mechanism underlies this computation.

In parsing research, there has been an major line of research on whether the candidate structural analyses are generated serially or in parallel (R. Lewis 2000; see Clifton & Staub 2008 for discussion). The same question can be asked, and has been asked under different terms, in production research. One theoretical choice point in production models is whether structural alternatives are selected competitively (e.g., Dell & O'Seaghdha 1992). Competitive models assume parallelism in selecting structural alternatives, while non-competitive models usually assume serial processes (as in V. Ferreira 1996). Parallel parsing models and competitive generation models predict that processing is more costly when there are multiple structural alternatives, even before the input becomes incompatible with one of the preferred structural alternatives. Contrary to this prediction, one notable empirical parallel between parsing and generation research is that the (temporary) availability of structural choice (e.g., ambiguity in parsing and structural flexibility in generation) *facilitates* rather than inhibits processing, until the structure becomes incompatible with the input (V. Ferreira 1996; Traxler et al. 1998). This is suggestive evidence that the same non-competitive mechanism for structural decision making might be at play in both parsing and generation.

4.3. Long-distance dependency formation

Both parsing and generation involve establishing long-distance dependencies, including filler-gap dependencies (e.g., wh-questions, relative clauses), anaphoric dependencies (e.g., reflexives, pronouns), morphosyntactic dependencies (e.g., subject-verb agreement), among others. Under the mechanism-level identity view, the parser and the generator should engage the same mechanisms for processing each of these dependencies, although they may exhibit different surface phenomena, due to differences in the content and timing of the input.

4.3.1. Filler-gap dependencies

In parsing, two prominent processes presumed to be involved in filler-gap dependencies are active gap search (e.g., Fodor 1978) and the reactivation and integration of fillers (e.g., McElree & Bever 1989). In active gap search the parser initiates a search for a potential gap site upon encountering a filler (e.g., wh-words, evidence for a relative clause, etc.). As a result, the parser slows down when a potential host for the gap is filled by overt elements (*filled gap effect*: Crain & Fodor 1985; Stowe 1986), or even when the first-encountered verb (in a non-island context) is an intransitive verb that cannot host a gap (hyper-active gap-filling; Omaki et al. 2015). This suggests that the parser actively creates a gap representation as soon as it encounters a filler.

Is the process of building filler-gap dependencies similarly ‘active’ in production? We suspect that it is, but this does not mean that we should expect to find filled-gap effects in production in the same constructions that elicit them in comprehension. Here, as in other cases, parsing and generation differ in terms of their starting point. Comparing surface effects can be misleading. In comprehension in English, the filler signals that a filler-gap dependency is present. The parser immediately knows that there is a gap in the sentence, but does not know where it is. Filled-gap effects arise when the parser commits to gap sites that turn out to be in the wrong position. In production, in contrast, we can presume that speakers have a better sense of the meaning of the sentence, and hence a better idea of where the gap should be. For example, in an English direct wh-question the speaker must commit early to the existence of a filler, because it is uttered first, and at that point should already know what is being questioned, i.e., the gap position. However, at the point of committing to a filler-gap dependency, the speaker may know the message, but might not yet know how that translates into linguistic form, and hence whether the filler-gap relation will turn out to be grammatically licit. Therefore, the production equivalent of filled gap effects is cases where the producer creates a filler-gap dependency and then discovers that it is illicit, because the intervening material creates a syntactic island, requiring the gap to be replaced with a resumptive pronoun. For example, F. Ferreira & Swets (2005) used a picture description task to elicit sentences with relative clauses. In the critical conditions, the filler-gap dependency turned out to include a wh-island, and hence required a resumptive pronoun instead of a gap, e.g., *This is the donkey that I don’t know where it came from*. Ferreira and Swets used measures of speech rate to show that speakers detected this problem in advance of articulating the wh-island. This suggests that the generator actively posits gaps before planning the syntactic structure of the clause containing the gap.

Another process in filler-gap dependency is the integration of fillers at the gap site. This has been shown in experiments that probe whether the filler representation is 'reactivated' at the gap site (McElree & Bever 1989), using such paradigms as cross-modal lexical priming (Nicol & Swinney 1989), probe recognition tasks (Bever & Sanz 1997; MacDonald 1989), and online semantic anomaly detection tasks (Wagers & Phillips 2014). It is also suggested by EEG effects found at the completion of entirely well-formed filler-gap dependencies (Kaan et al. 2000; Phillips et al. 2005). We know of no model of production that explicitly specifies what happens at the gap site. In both comprehension and production the gap site marks the completion of the dependency, but since producers know what they mean to say, this involves less reduction in uncertainty than the corresponding step in comprehension. Under the mechanism-level identity view, however, the retrieval of fillers might be expected to occur when the generator needs to link the filler with the gap position. Some preliminary evidence suggests this is the case (Momma et al., in prep) Thus, the parsing model provides a reasonable default hypothesis about what happens at the gap site in production.

4.3.2. Subject-verb agreement

In many languages, including English, the parser and the generator must be able to establish morpho-syntactic agreement between the subject and the verb in terms of grammatical features, e.g., number features. In this process, both the parser and the generator are known to be susceptible to interference from nearby material (Bock & Miller 1991; Wagers et al. 2009). For example, speakers often mis-produce a plural verb when the subject noun is singular but a structurally irrelevant plural noun is present: *the key to the cabinets are...* Likewise, comprehenders often fail to notice the ungrammaticality of the plural verb when the same sentence is presented to them. Like in filler-gap dependency processing, logically speaking, the parser and the generator could share or not share the same mechanism for computing subject-verb agreement. Whatever the right explanation of this phenomenon (see, e.g., Eberhard Cutting & Bock 2005; Wagers et al. 2009 for two different accounts), when agreement processing is successful and when it is unsuccessful seem to generally overlap (cf. Bock & Cutting 1992). This is surprising if the parser and the generator compute subject-verb agreement using different mechanisms.

However, despite the similarities in the surface effects of agreement attraction, Badecker & R. Lewis (2007) caution that different sets of cues are available for establishing subject-verb agreement in comprehension and production. In comprehension, the verb form is provided to the comprehender, so its morphological features (e.g., number, gender) can be used as cues for retrieving the subject. In production, the speaker's task is to choose an appropriate verb form, so the verb cannot possibly provide morphological retrieval cues for accessing the subject. This does not mean that the mechanism for establishing subject-verb agreement, namely content-addressable memory, is different in speaking and understanding, but it raises the possibility of finer grained differences between agreement attraction in production and comprehension. Also, it poses a challenge for algorithm-level identity with respect to subject-verb agreement.

4.3.3. Anaphoric dependencies

Both the parser and the generator must be able to establish referential dependencies between anaphoric elements (e.g., *herself, himself, them*) and their antecedents, matching in grammatical features such as number and gender. This process is superficially similar to subject-verb agreement: both involve establishing a morpho-syntactic dependency with respect to some grammatical features. However, unlike subject-verb agreement, the parser and the generator seem to behave differently. Bock et al. (1999) reported that reflexive number agreement is susceptible to interference from non-subject interveners in generation. In contrast, this attraction effect is either weak or nonexistent in parsing (Dillon et al. 2013). We now know of situations that more reliably elicit attraction effects for reflexives (Parker & Phillips 2016), but these require more specific configurations. If this discrepancy is real, it suggests that either the task, the mechanism or the procedures for establishing anaphoric dependencies are different between generation and parsing. We currently do not know which is to be blamed or if the observed empirical pattern is consistent across studies.

One possible cause of this discrepancy is the differences in starting point. In parsing, comprehenders use agreement morphology to encode coreference relationship. In production, speakers use coreference information to encode agreement morphology. Thus, the parser needs to rely heavily on morphological information, while the generator needs to rely heavily on semantic (coreference) information. Such a difference in the representational type of retrieval cues does not exist in subject-verb agreement. It is possible that retrieval processes using morphological cues vs. semantic cues with respect to antecedent-reflexive dependencies are differentially sensitive to grammatical constraints, i.e., that a reflexive must be locally c-commanded by its antecedent. The challenge to this approach is to explain exactly why retrieval based on morphological cues on reflexives should be more sensitive to grammatical constraints than retrieval based on semantic cues. This issue is in need of further investigation.

4.3.4. Is mechanism-level identity plausible?

In this section we derived some predictions and questions from the mechanism-level identity view, and reviewed relevant empirical findings, ranging from lexical access to long-distance dependency formation. We argued that mechanism-level identity is largely consistent with what we know about parsing and generation. Our knowledge of production processes is clearly more limited, but addressing them in parallel with comprehension is both feasible and instructive.

4.4. Evaluating the algorithm-level identity view

In the previous section we asked whether the same structure creation mechanisms are engaged in parsing and generation, but with no commitment to structures being assembled in the same order in both tasks. If input is provided to the parser and the generator in different orders, then the same mechanisms could build structures in different orders. The order of structure building in parsing is relatively fixed, because the comprehender is at the mercy of the order of the input. In generation the speaker has more latitude, at least in principle. The sentence has to be uttered using the correct word order, but that is no guarantee that the structure is planned in the order that the words are uttered. In this section we examine the stronger claim that the order of structure-building operations is the same in parsing and generation. We call this the *algorithm-level identity* view.

4.4.1. Manner of structure building

The parsing literature distinguishes whether constituents are built in a bottom-up or top-down fashion. In bottom-up structure building words are accessed first and then structures are built by combining retrieved words. In top-down structure building phrasal structures are built first and then retrieved words are inserted into available structural positions. These approaches differ in the relative timing of structural vs. lexical processes, and we can test their status in parsing and generation.

In parsing research, it is widely acknowledged that a mix of bottom-up and top-down parsing methods are necessary to parse a sentence. Miller & Chomsky (1963) noted that consistently left-branching structures (e.g., *John's brother's cat despises rats*) and consistently right-branching structures (e.g., *the dog that chased the cat that bit the rat likes the cheese*) do not cause substantial processing difficulty, relative to structures with mixed branching direction (e.g., center embedded sentences). This is unexpected if the human parser relies on purely bottom-up or purely top-down structure building. Assuming standard phrase structure rules, a purely bottom-up parser cannot process a left-branching structure without substantial memory cost. Meanwhile, a purely top-down parser cannot process right-branching structures without making many incorrect structural predictions. A solution is to adopt a mixed method, in which bottom-up input filters top-down structure building, as in a left-corner parser. The left-corner model predicts that center embedded sentences are hard and left-branching and right-branching sentences are easy. Thus, it captures when the human parser experiences difficulty, and it has been adopted by models of comprehension (e.g., Abney & Johnson 1991; Johnson-Laird 1983; R. Lewis & Vasishth 2005; Resnik 1992). We thus take this as a good first approximation to how the human parser processes a sentence.

Turning to generation, some models of production assume a radically top-down approach in which the basic structure of a sentence is planned before lexical items are integrated into the structure (e.g., Garrett 1980; 1988). Other models of production adopt a more bottom-up approach, in which structures are built as a result of retrieval of lexical items, in particular the head of the phrase (e.g., Bock & Levelt 1994; Levelt 1989). Thus, there is a potential mismatch in how existing parsing and generation models build structures.[4]

We argue that both top-down and bottom-up structure building are used in generating individual sentence structures. One type of evidence for top-down structure building comes from the experiment described in Section 4.2.3 (Momma et al. 2017). We found that the speed-up effect of repeating a verb-phrase structure in a ditransitive sentence preceded the speed-up effect of repeating a noun internal to that verb-phrase. Assuming that speed-up due to repetition priming reflects the time when the repeated material is planned, this suggests that structural

[4] The model described by Levelt is not purely bottom-up (Levelt 1989, Chapter 7). For example, the model allows a verb's category node to be created based on the need to have a predicate, rather than based on having identified a specific verb. This is a limited case of top-down structure building.

representations are constructed before lexical planning, at least when producing a ditransitive verb phrase. This can be understood as evidence for top-down structure building in the verb phrase. As for bottom-up structure building, there is ample evidence that the availability of a word determines which sentence structure is used (Bock 1987 and much subsequent work). One interpretation of this *availability effect* is that lexical access precedes structure building, i.e., evidence for bottom-up structure building. When combined, these two pieces of evidence suggest that generation involves both bottom-up and top-down structure building processes.

Certainly, this evidence falls short of showing that the generator builds structures in exactly the same order as a left-corner parser. We consider that rather unlikely. For example, when building a noun phrase in a language with grammatical gender, it is unlikely that the determiner is planned before the noun head. However, the mixture of top-down and bottom-up structure building may be shared between parsing and generation. The difference in the order of derivation might simply reflect differences between the two tasks, which affect which pieces of information become available when to the language processor. In a normal comprehension task, the parser receives input in a fixed linear order, so the parsing algorithm ends up looking like a left-corner parser algorithm. In contrast, in picture description tasks, for example, information that corresponds to the head noun is what hits speaker's' eyes. Thus, the generation algorithm may use the head noun to filter top-down structure building, ending up looking like a head-driven algorithm in noun phrase production. This could be sufficient to lead to differences in which parts of the sentence are built in a top-down vs. a bottom-up fashion.

4.4.2. Order of structure building

Even if both parsing and generation involve a mix of top-down and bottom-up structure building, there is freedom in the order of structure building that could give rise to differences between parsing and generation algorithms. Indeed, in the previous section we acknowledged the implausibility of an exact match in the order of structure building between parsing and generation. Here we explore the possibility that this flexibility in the order of structure building could be shared between parsing and generation.

First, it is unlikely that the planning of lexical items in production proceeds in an exactly left-to-right fashion. For example, the determiner in a noun phrase is likely planned after the head noun, especially in gendered languages like German. There is also evidence that a sentence medial adjunct is planned after the sentence final verb in some English intransitive sentences (Momma et al. 2017). However, it is increasingly unclear whether parsing proceeds in an exactly left-to-right order either, due to the involvement of predictive processes (Altmann & Kamide 1999; DeLong et al. 2005; Kamide et al. 2003; Lau et al. 2006; Staub & Clifton 2006; van Berkum et al. 2005). For example, when the parser predicts a noun phrase that occurs later in a sentence, it is possible that the head is planned before the determiner or adjective that modifies it. Thus, when the parser builds structures ahead of the phonological input, it might behave like the generator in terms of the order of structure building. Conversely, when the generator is constrained by phonological input, as in a sentence completion task, where the preamble of the sentence is provided to the speaker, the generator might behave like the parser in terms of the order of structure building. For example, speakers can be given the determiner phonologically as a preamble, and they should have no difficulty completing the rest of the noun phrase. Here, again, differences between parsing and generation might simply reflect differences in what is available in the different task environments.

4.4.3. Is algorithm-level identity plausible?

In this section, we derived some empirical questions from the algorithm-level identity view, and discussed some findings that bear on the order in which the parser and the generator build sentence structures. We argued that the algorithm-level identity view is largely consistent with what we have learned so far, but that it is likely necessary to assume some flexibility in how structures are built in parsing and generation.

5. Challenges

Despite the theoretical appeal of a unified account, there are reasons to think that parsing and generation may be supported by distinct mechanisms. These include: (i) neuropsychological and developmental double-dissociations; (ii) differences between acceptability and producibility and (iii) differences in maximal processing speed. We briefly review each challenge and its seriousness.

5.1. Arguments from double dissociation

Classic neuropsychological investigations have suggested that patients with left frontal vs. temporal lobe damage can show selective impairments in expression or perception of language. Historically, Broca's and Wernicke's aphasia were thought to be deficits of production and perception of speech, respectively, and this double dissociation led to comprehension and production being localized to distinct regions in the brain in the Wernicke-Geschwind model (Geschwind 1967; see Ben Shalom & Poeppel 2008). Following the logic of double dissociation, these selective impacts imply that the parser and the generator are distinct mechanisms.

Also, it has been noted repeatedly in the developmental literature that language production is normally delayed compared to comprehension (e.g., Keenan & MacWhinney 1987). The reverse is also true in some domains of language. For example, 4-6-year old children often misinterpret sentences like *Ernie washed him* to mean *Ernie washed himself* (Chien & Wexler 1990, but see Conroy et al. 2009), but corresponding errors are not found in production (Bloom et al. 1994). Thus, we find cases of double dissociation in developmental time course. This suggests that some components of parsing and generation might be distinct in children.

The interpretation of double dissociations, however, is not so straightforward. Dissociations of comprehension and production tasks do not entail that the mechanisms are dissociable. In neuropsychology, Caramazza & Zurif (1976) have shown that Broca's aphasics' comprehension is far from normal, especially for sentences with non-canonical structures like passives and object relative clauses. Thus, Broca's aphasia may reflect, for example, a deficit in mapping morphosyntactic information and semantics, which are necessary for both systematic comprehension and fluent production. Wernicke's patients' speech production is also far from normal. In addition at least some Wernicke's patients also have preserved reading comprehension (Ellis et al. 1983). From this, it has been claimed that Wernicke's aphasics may have trouble mapping phonological and semantic sources of information onto each other. If these explanations are correct, it could be argued that the apparent double dissociations simply reflect the differential demands of individual sub-tasks in comprehension and production. The same argument could be made for developmental double dissociations. The challenge for an advocate of the single mechanism view is to show explicitly how the same underlying deficit yields different surface effects. An advocate of a two mechanism approach does not face this challenge, but must instead address the related challenge of why superficially different aspects of speaking and understanding are frequently impacted together.

5.2. Differences in acceptability and producibility

Speakers often produce sentences that they themselves find unacceptable. A well-known case in English is resumptive pronouns inside syntactic islands (e.g., McDaniel & Cowart 1999; Zukowski & Larsen 2004). In an elicited production study by F. Ferreira & Swets (2005), speakers produced sentences like, **This is the donkey that I don't know where it lives.* The same sentences were, however, judged to be unacceptable. This mismatch between acceptability and producibility poses an apparent challenge to the notion of shared mechanisms for comprehension and production. However, it is again unclear whether this difference is due to a task difference or a difference in mechanisms. In production tasks, speakers need to complete sentences, and this pressure can overcome the dispreference for certain structures. Such pressure does not exist in acceptability judgement tasks, where comprehenders simply need to provide a rating for a sentence.

5.3. Difference in maximal speed

At least after some amount of practice, comprehenders are surprisingly good at understanding compressed speech (Foulke & Sticht 1969; see also MacKay 1987 for a discussion), which has a much faster speed than fluent speakers can possibly speak. This rate difference could suggest distinct mechanisms. But it again could reflect task differences rather than different mechanisms. Speaking involves much open-ended decision making, whereas comprehenders face fewer and more constrained choices. Also, speakers have to move their articulators, whereas comprehenders merely have to build mental representations.

6. Conclusion

In this article, we advanced the view that a single mechanism for building sentence structure may be sufficient for structure building in comprehension and production, synthesizing previous proposals and challenges on the issue. There are many superficial differences, but these may reflect input differences rather than distinct mechanisms. As argued above, this view is a prerequisite for maintaining a transparent relationship between grammar and real-time structure building, and thus it has far-reaching implications for the architecture of the human language system. This view also provides a useful framework for integrating largely independent research programs on comprehension and production by both constraining the models and uncovering new questions that can drive further research.

Acknowledgments

This research was supported in part by NSF DDRI grant BSC-1530332 to SM and CP, NSF NRT grant DGE-1449815 to CP, and NIH grant R01 HD051030 to Victor Ferreira.

Literature Cited

Abney SP, Johnson M. 1991. Memory requirements and local ambiguities of parsing strategies. *J. Psycholinguist. Res.* 20:233-50

Allum PH, Wheeldon LR. 2007. Planning scope in spoken sentence production: The role of grammatical units. *J. Exp. Psychol.-Learn. Mem. Cogn.* 33(4): 791

Allum PH, Wheeldon LR. 2009. Scope of lexical access in spoken sentence production: Implications for the conceptual-syntactic interface. *J. Exp. Psychol.-Learn. Mem. Cogn.* 35(5):1240

Altmann GTM, Kamide Y. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73:247–64

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

Appelt DE. 1987. Bidirectional grammars and the design of natural language generation systems. In *Proceedings of the 1987 Workshop on Theoretical Issues in Natural Language Processing*, pp. 206-12. Association for Computational Linguistics

Badecker, W, Lewis, R. 2007. *Agreement and sentence formulation: The role of working memory (retrievals) in language production. Unpublished manuscript*

Ben Shalom D, Poeppel D. 2008. Functional anatomic models of language: assembling the pieces. *Neuroscientist* 14(1):119-27

Bever TG, McElree B. 1988. Empty categories access their antecedents during comprehension. *Linguist. Inq.* 19(1):35-43

Bever TG, Sanz M. 1997. Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguist. Inq.* 28:69-91

Bianchi V, Chesi C. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguist. Inq.* 45(4):525-69

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

Bock JK, Cutting JC. 1992. Regulating mental energy: Performance units in language production. *J. Mem. Lang.* 31:99–127

Bock JK, Levelt WJM. 1994. Language production: grammatical encoding. In *Handbook of Psycholinguistics*, ed. M Gernsbacher, pp. 945–84. San Diego: Academic Press

Bock K, Miller CA. 1991. Broken agreement. *Cogn. Psychol.* 23(1):45-93

Bock JK, Ferreira VS. 2014. Syntactically speaking. In *The Oxford Handbook of Language Production*, eds. M Goldrick, VS Ferreira, M Miozzo, pp. 21-46. New York: Oxford University Press

Bock JK. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *J. Exp. Psychol.-Learn. Mem. Cogn.* 12(4):575

Bock K. 1987. An effect of the accessibility of word forms on sentence structures. *J. Mem. Lang.* 26(2):119-37

Bock K. 1991. A sketchbook of production problems. *J. Psycholinguist. Res.* 20(3):141-60

Bock K, Eberhard KM, Cutting JC. 2004. Producing number agreement: How pronouns equal verbs. *J. Mem. Lang.* 51(2):251-78

Bock K, Nicol J, Cutting JC. 1999. The ties that bind: Creating number agreement in speech. *J. Mem. Lang.* 40(3):330-46

Boomer DS. 1965. Hesitation and grammatical encoding. *Lang. and Speech* 8(3):148-58

Caramazza A, Zurif EB. 1976. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain Lang.* 3(4):572-82

Chien Y-C, Wexler K. 1990. Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Lang. Acquis.* 1:225-95

Chow WY, Momma S, Smith C, Lau E, Phillips C. 2016. Prediction as memory retrieval: timing and mechanisms. *Lang. Cogn. Neurosci.* 31(5):617-27

Clifton C, Staub A. 2008. Parallelism and competition in syntactic ambiguity resolution. *Lang. Linguist.* Compass 2(2):234-50

Conroy A, Takahashi E, Lidz J, Phillips C. 2009. Equal treatment for all antecedents: How children succeed with Principle B. *Linguist. Inq.* 40:446-486

Crain S, Fodor JD. 1985. How can grammars help parsers. In *Natural language parsing: Psychological, computational, and theoretical perspectives,* eds. DR Dowty, L Karttunen, AM Zwicky, pp. 94-128. Cambridge University Press

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

Dell GS, Chang F. 2014. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Phil. Trans. R. Soc. B*, *369*(1634):20120394

Dell GS, O'Seaghdha PG. 1992. Stages of lexical access in language production. *Cognition* 42(1):287-314

Dell GS, Oppenheim GM, Kittredge AK. 2008. Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Lang. Cogn. Process.* 23(4):583-608

DeLong KA, Urbach TP, Kutas M. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8(8):1117–21

Dillon B, Mishler A, Sloggett S, Phillips C. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *J. Mem. Lang.* 69(2):85-103

Eberhard KM, Cutting JC, Bock JK. 2005. Making syntax of sense: Number agreement in sentence production. *Psychol. Rev.* 112:531–59

Ellis AW, Miller D, Sin G. 1983. Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition* 15(1):111-44

Federmeier KD. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44:491–505

Ferreira F, Swets B. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* 46(1):57-84

Ferreira F, Swets B. 2005. The production and comprehension of resumptive pronouns in relative clause“ island” contexts. In *Twenty-first Century psycholinguistics: Four cornerstones,* ed. A Cutler, pp. 263-78. Mahway, NJ: Lawrence Erlbaum Associates

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

Ferreira VS. 1996. Is it better to give than to donate? Syntactic flexibility in language production. *J. Mem. Lang.* 35(5):724-55

Ferreira VS, Humphreys KR. 2001. Syntactic influences on lexical and morphological processing in language production. *J. Mem. Lang.* 44(1):52-80

Fodor JD. 1978. Parsing strategies and constraints on transformations. *Linguist. Inq.* 9(3):427-73

Fodor JA, Bever TG, Garrett MF. 1974. *The Psychology of Language: an Introduction to Psycholinguistics and Generative Grammar.* New York: McGraw-Hill

Ford M, Holmes VM. 1978. Planning units and syntax in sentence production. *Cognition* 6(1):35-53

Foulke, E. and Sticht, T.G., 1969. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, *72*(1), p.50.

Forster K. 1992. Memory-addressing mechanisms and lexical access. *Adv. Psychol.* 94:413-34

Frazier L. 1987. Sentence processing: a tutorial review. In *Attention and Performance XII*, ed. M Coltheart, pp. 559-85. Hillsdale, NJ: Erlbaum

Garrett MF. 1980. Levels of processing in sentence production. In *Language Production, Volume I: Speech and Talk,* ed. BL Butterworth, pp. 177-220. London: Academic Press

Garrett MF. 1988. Processes in language production. In *Linguistics: The Cambridge Survey, HI: Language: Psychological and Biological Aspects,* ed. FJ Newmeyer, pp. 69-96. Cambridge, UK: Cambridge University Press

Garrett MF. 2000. Remarks on the architecture of language processing systems. In *Language and the Brain,* eds. Y Grodzinsky, L Shapiro, D Swinney, p. 31. San Diego, CA: Academic Press

Gennari SP, MacDonald MC. 2009. Linking production and comprehension processes: The case of relative clauses. *Cognition* 111(1):1-23

Geschwind N. 1967. The varieties of naming errors. *Cortex* 3:97-112

Gleitman L. 1990. The structural sources of verb meanings. *Lang. Acquis.* 1(1):3-55

Gleitman LR, January D, Nappa R, Trueswell JC. 2007. On the give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57(4):544-69

Griffin ZM, Weinstein-Tull J. 2003. Conceptual structure modulates structural priming in the production of complex sentences. *J. Mem. Lang.* 49(4):537-55

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

Hartsuiker RJ, Kolk HH. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cogn. Psychol.* 42(2):113-157

Huettig F, Hartsuiker RJ. 2010. Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Lang. Cogn. Process.* 25(3):347-74

Iwasaki N. 2010. Incremental sentence production: Observations from elicited speech errors in Japanese. In *Processing and producing head-final structures*, pp. 131-151. Netherlands: Springer

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

Joshi AK, Schabes Y. 1997. Tree-adjoining grammars. In *Handbook of Formal Languages,* eds. G Rozenber, A Saloma, pp. 69-123. Berlin: Springer Verlag

Kaan E, Harris A, Gibson E, Holcomb P. 2000. The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Proc.* 15:159-201

Kamide Y, Altmann GT, Haywood SL. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *J. Mem. Lang.* 49(1):133-56

Keenan JM, MacWhinney B. 1987. Understanding the relationship between comprehension and production. In *Psycholinguistic Models of Language Production,* eds. HW Dechert, M Raupach. Norwood, NJ: Ablex

Kempen G, Hoenkamp E. 1987. An incremental procedural grammar for sentence formulation. *Cogn. Sci.* 11(2):201-58

Kempen G. 2000. Could grammatical encoding and grammatical decoding be subserved by the same processing module?. *Behav. Brain Sci.* 23(01):38-9

Kempen G. 2014. Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics* 12(1):111-42

Kempen G, Olsthoorn N, Sprenger S. 2012. Grammatical workspace sharing during language production and language comprehension: evidence from grammatical multitasking. *Lang. Cogn. Process.* 27(3):345-80

Kempson, R., Meyer-Viol, W. and Gabbay, D.M., 2000. *Dynamic syntax: The flow of language understanding*. Wiley-Blackwell

Kim CS, Carbary KM, Tanenhaus MK. 2014. Syntactic priming without lexical overlap in reading comprehension. *Lang. and Speech* 57(2):181-95

Kim AE, Oines LD, Sikos L. 2016. Prediction during sentence comprehension is more than a sum of lexical associations: the role of event knowledge. *Lang. Cogn. Neurosci.* 31(5):597-601

Konopka AE. 2012. Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *J. Mem. Lang.* 66(1):143-62

Kutas M, Hillyard SA. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207(4427):203-205

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

Levelt WJ. 1983. Monitoring and self-repair in speech. *Cognition* 14(1):41-104

Levelt WJ. 1989. *Speaking: From intention to articulation.* Cambridge, MA: MIT Press

Lewis RL, Vasishth S. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29(3):375-419

Lewis RL. 2000. Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *J. Psycholinguist. Res.* 29(2):241-8

Lewis S, Phillips C. 2015. Aligning grammatical theories and language processing models. *J. Psycholinguist. Res.* 44(1):27-46

Lombardo V, Sturt P. 2002. Incrementality and lexicalism: A treebank study. In *The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues,* eds. S Stevenson, P Merlo, pp.137-155. Philadelphia, PA: John Benjamins

Lupker SJ. 1979. The semantic nature of response competition in the picture-word interference task. *Mem. Cognit.* 7(6):485-95

MacKay, DG. 1987. Asymmetries in the relationship between speech perception and production. *Perspectives on Perception and Action*, pp.301-333.

MacDonald MC, Pearlmutter NJ, Seidenberg MS. 1994. The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101(4):676

MacDonald MC. 1989. Priming effects from gaps to antecedents. *Lang. Cogn. Process.* 4(1):35-56

MacDonald MC. 2013. How language production shapes language form and comprehension. *Front. Psychol.* 4:226

Marr D. 1982. *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman

Marcus MP. 1980. *Theory of syntactic recognition for natural languages.* Cambridge, MA: MIT press

Marslen-Wilson W. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244:522-3

McClelland JL, Elman JL. 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18(1):1-86

McDaniel D, Cowart W. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70(2):B15-B24

McElree B, Bever TG. 1989. The psychological reality of linguistically defined gaps. *J. Psycholinguist. Res.* 18(1):21-35

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

Menenti L, Gierhan SM, Segaert K, Hagoort P. 2011. Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol. Sci.* 22(9):1173-82

Meyer AS. 1996. Lexical access in phrase and sentence production: Results from picture-word interference experiments. *J. Mem. Lang.* 35:477-96

Miller G, Chomsky N. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology*, eds. RD Luce, RR Bush, E Galanter, 2:419–91. New York, NY: Wiley

Momma S, Bowen Y, Ferreira V. 2017. Non-linear lexical planning in sentence production. Talk given at the *30th annual CUNY Conference on Human Sentence Processing*, Cambridge, MA. March 30-April 1

Momma S, Slevc LR, Buffinton J, Phillips C. 2016. Similar words compete, but only when they're from the same category. Poster presented at the *29th annual CUNY Conference on Human Sentence Processing*, Gainesville, FL. March 3-5

Momma S, Muller H, Phillips C. In prep. *Long-distance dependency in sentence production.*

Momma S, Kraut R, Slevc LR, Phillips C. 2017. Timing of syntactic and lexical priming reveals structure-building mechanisms in production. Poster presented at the *30th annual CUNY Conference on Human Sentence Processing*, Cambridge, MA. March 30-April 1

Neely JH. 1977. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol.* 106(3):226

Nicol J, Swinney D. 1989. The role of structure in coreference assignment during sentence comprehension. *J. Psychol. Res.* 18(1):5-19

Nooteboom SG. 1969. The tongue slips into patterns. In *Leyden Studies in Linguistics and Phonetics*, eds. G Sciarone, A van Essen, A Van Raad, pp. 114–132. The Hague: Mouton

Norcliffe E, Konopka AE, Brown P, Levinson SC. 2015. Word order affects the time course of sentence formulation in Tzeltal. *Lang. Cogn. Neurosci.* 30(9):1187-208

Omaki A, Lau EF, Davidson White I, Dakan ML, Apple A, Phillips C. 2015. Hyper-active gap filling. *Front. Psychol.* 6:384

Parker D, Phillips C. 2017. Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94:272-290

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

Phillips C. 2003. Linear order and constituency. *Linguist. Inq.* 34(1):37-90

Phillips C, Kazanina N, Abada S. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cog. Br. Res.* 22:407-428.

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

Phillips C, Wagers MW, Lau EF. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces* 37:147-80

Pickering MJ, Branigan HP. 1998. The representation of verbs: Evidence from syntactic priming in language production. *J. Mem. Lang.* 39(4):633-51

Pickering MJ, Garrod S. 2007. Do people use language production to make predictions during comprehension?. *Trends in Cogn. Sci.* 11(3):105-10

Pickering MJ, Garrod S. 2013. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36(04):329-47

Pickering MJ, Ferreira VS. 2008. Structural priming: A critical review. *Psychol. Bull.* 134(3):427

Pinker S. 1984. *Language learnability and language development.* Cambridge, MA: Harvard University Press

Postma A. 2000. Detection of errors during speech production: A review of speech monitoring models. *Cognition* 77(2):97-132

Resnik P. 1992. Left-corner parsing and psychological plausibility. In *Proceedings of the 14th Conference on Computational Linguistics*, 1:191-7. Association for Computational Linguistics

Roelofs A. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42:107-42

Schneider DA. 1999. *Parsing and incrementality.* Doctoral dissertation, University of Delaware

Schriefers H, Meyer AS, Levelt WJ. 1990. Exploring the time course of lexical access in language production: Picture-word interference studies. *J. Mem. Lang.* 29(1):86-102

Segaert K, Wheeldon L, Hagoort P. 2016. Unifying structural priming effects on syntactic choices and timing of sentence generation. *J. Mem. Lang.* 91:59-80

Segaert K, Menenti L, Weber K, Petersson KM, Hagoort P. 2011. Shared syntax in language production and language comprehension—an fMRI study. *Cereb. Cortex* p. Bhr249

Shan CC, Barker C. 2006. Explaining crossover and superiority as left-to-right evaluation. *Linguist. Philos.* 29(1):91-134

Shieber SM. 1988. A uniform architecture for parsing and generation. In *Proceedings of the 12th conference on Computational linguistics*, 2:614-619. Association for Computational Linguistics

Smith M, Wheeldon L. 1999. High level processing scope in spoken sentence production. *Cognition* 73(3):205-46

Snedeker J, Trueswell J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *J. Mem. Lang.* 48(1):103-30

St. John MF, McClelland JL. 1990. Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence* 46(1-2):217-57

Staub A, Clifton C Jr. 2006. Syntactic prediction in language comprehension: evidence from either... or. *J. Exp. Psychol.-Learn. Mem. Cogn.* 32(2):425

Steedman M. 2000. *The syntactic process*, vol. 24. Cambridge: MIT Press

Stowe LA. 1986. Parsing WH-constructions: Evidence for on-line gap location. *Lang. Cogn. Process.* 1(3):227-45

Sturt P, Lombardo V. 2005. Processing coordinated structures: Incrementality and connectedness. *Cogn. Sci.* 29(2):291-305

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217):1632-4

Traxler MJ, Pickering MJ, Clifton C Jr. 1998. Adjunct attachment is not a form of lexical ambiguity resolution. *J. Mem. Lang.* 39:558-92

Trueswell JC, Kim AE. 1998. How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *J. Mem. Lang.* 39(1):102-23

Van Berkum JJ, Brown CM, Zwitserlood P, Kooijman V, Hagoort P. 2005. Anticipating upcoming words in discourse: evidence from erps and reading times. *J. Exp. Psychol.-Learn. Mem. Cogn.* 31:443-67

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):105-43

Wagers MW, Lau EF, Phillips C. 2009. Agreement attraction in comprehension: Representations and processes. *J. Mem. Lang.* 61(2):206-37

Wagers, M.W. and Phillips, C., 2014. Going the distance: memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), pp.1274-1304.

Wagner V, Jescheniak JD, Schriefers H. 2010. On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *J. Exp. Psychol.-Learn. Mem. Cogn.* 36(2):423

Wheeldon L, Smith M. 2003. Phrase structure priming: A short-lived effect. *Lang. Cogn. Process.* 18(4):431-42

Wheeldon L, Ohlson N, Ashby A, Gator S. 2013. Lexical availability and grammatical encoding scope during spoken sentence production. *Q. J. Exp. Psychol.* 66(8):1653-73

Wright B, Garrett M. 1984. Lexical decision in sentences: Effects of syntactic structure. *Mem. Cognit.* 12(1):31-45

Zukowski A, Larsen J. 2004. The production of sentences that we fill their gaps. Poster presented at the *CUNY Sentence Processing Conference*, University of Maryland