Title
Saying what's on your mind : working memory effects on syntactic production

Author
Slevc, Lloyd Robert

Publication Date
2007

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Saying what's on your mind: Working memory effects on syntactic production

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

Doctor of Philosophy

in

Psychology

by

Lloyd Robert Slevc

Committee in charge:

Professor Victor Ferreira, Chair
Professor Tamar Gollan
Professor Marta Kutas
Professor Harold Pashler
Professor John Wixted

2007

The Dissertation of Lloyd Robert Slevc is approved, and it is acceptable in quality and form for publication on microfilm:

____

Chair

University of California, San Diego

2007

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Portions of this work were presented at the Architectures and Mechanisms for Language Processing (AMLaP) conference in August, 2006, and at the 47th Annual Meeting of the Psychonomic Society in November, 2006.

I would like to thank my advisor, Vic Ferreira, for his guidance, interest, and enthusiasm throughout my time at UCSD. His advice and support have been invaluable, and I know of nobody else who would even attempt to read, comment on, and return the first draft of a dissertation within six hours of being asked.

I would also like to thank Kay Bock, Jeremy Boyd, Naomi Friedman, Tamar Gollan, Florian Jaeger, and Liane Wardlow Lane for helpful discussions and advice on the research reported here. I owe still more thanks to Stephanie Ang, Serina Chang, Ruby Cho, Katie Doyle, Paige Heath, Ramsey Khouri, Yanny Siu, and Neil Spears, without whom I would have considerably less data.

Finally, I would like to thank my family for steadfastly supporting my desire to remain a student for nearly 25 continuous years, and Eileen Cardillo and her puppy Milo for reminding me that, while dissertations are important business, sometimes you just need to sit down, relax, and chew on a plastic giraffe.

VITA

Educational History

2007	Ph.D. in Psychology emphasis in Cognitive Psychology (Psycholinguistics) University of California, San Diego
1999	B.A. in Psychology, Summa cum Laude B.A. in Chinese University of Colorado, Boulder

Peer-Reviewed Publications

Slevc, L. R. & Miyake, A. (2006). Individual differences in second language proficiency: does musical ability matter? *Psychological Science, 17(8),* 675–681.

McCleery, J. P., Tully, L. M., Slevc, L. R., & Schreibman, L. (2006). Consonant production patterns of young severely language-delayed children with autism. *Journal of Communication Disorders, 39(3),* 217-231.

Slevc, L. R. & Ferreira, V. S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language, 54*, 515–540.

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

Chapters and Proceedings

Ferreira, V. S. & Slevc, L. R. (in press). Grammatical encoding. To appear in M. Gareth Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford University Press.

Slevc, L. R., Wardlow Lane, L., & Ferreira, V. S. (2007). Pronoun production: Word or world knowledge? In C. T. Schütze & V. S. Ferreira (Eds.), *The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop. MIT Working Papers in Linguistics, 53*, 191-203.

Wardlow Lane, L., Slevc, L. R., & Ferreira, V. S. (2007). Exchanging Elicits: Stem-Exchange Errors and Syntactic Category Shifts. In C. T. Schütze & V. S. Ferreira (Eds.), *The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop. MIT Working Papers in Linguistics, 53*, 221-232.

ABSTRACT OF THE DISSERTATION

Saying what's on your mind: Working memory effects on syntactic production

by

Lloyd Robert Slevc

Doctor of Philosophy in Psychology

University of California, San Diego, 2007

Professor Victor Ferreira, Chair

A fundamental difficulty with producing language is that non-linear conceptual information must be expressed as a linear order of words. Consequently, speakers must often temporarily maintain ideas and lexical representations until they can be grammatically produced, a process that likely relies on working memory (WM). One way speakers might reduce demand on WM is by using syntactic structures that allow early mention of more accessible information (*accessibility effects*), thereby avoiding having to buffer ready-to-produce material. Speakers also use syntactic structures in which previously mentioned (given) information precedes new information. These *given-new ordering* effects are generally assumed to be communicatively driven, but might actually reduce to accessibility effects because given information is usually also highly accessible.

Five experiments investigated the role and the type of WM involved in syntactic production by examining WM influences on accessibility effects and on given-new ordering. Experiment 1 showed that accessibility effects are sensitive to standard WM factors, suggesting these effects do, in fact, reflect WM processes. Experiment 2 showed that given-new ordering, like accessibility effects, is sensitive to WM factors, suggesting that it also arises from demands on speakers' limited capacity WM.

Although these experiments show that WM demands influence syntactic production, they used a sentence-internal WM load, and thus do not distinguish between a form of WM specific to syntactic processing and domain-general WM processes. Experiments 3 and 4 manipulated WM load sentence-externally, and showed that speakers' use of accessibility effects and of given-new ordering are reduced when under an external load, thereby implicating domain general verbal WM in these processes. A potential concern is that the effects in Experiments 1-4 reflect not WM mechanisms but rather increased task difficulty unrelated to WM. Experiment 5 addressed this possibility by relying on the established distinction between verbal and spatial WM, comparing speakers' use of given-new ordering when under verbal and spatial WM load. A concurrent verbal WM load reduced speakers' use of syntactic structures respecting given-new ordering, but a concurrent spatial WM load did not. Together, these experiments show that both accessibility effects and given-new ordering reflect limitations in domain-general verbal WM.

1 Introduction

An integral part of producing a linguistic expression is retrieving the words in a sentence. Such retrieval, like any form of retrieval from memory, is a variable and relatively unpredictable process. Another essential feature of spoken utterances (at least in fixed word-order languages like English) is the order in which words are produced. There is thus a tension between this variable lexical retrieval process and the syntactic encoding process that requires certain words to be used at certain times (Bock, 1982).

One way this tension is dealt with is by the grammar of the language and the way that syntactic encoding processes create syntactic structures (Bock, 1982; Ferreira, 1996). More specifically, speakers can benefit to the extent that the grammar of the language allows information to be produced as it becomes accessible, rather than having to be buffered or stored until it can be grammatically produced (this notion follows from the incremental nature of production, see, e.g., Kempen & Hoenkamp, 1987). This motivates an otherwise puzzling phenomena, namely that languages often offer more than one way to say the same thing (see Levin, 1993, for a variety of examples). For example, to convey information about a committee-agent, a student-patient, and a degree being awarded, a speaker could use a prepositional dative structure, as in (1), or a double-object dative structure, as in (2):

(1) The dissertation committee awarded a degree to the student.

(2) The dissertation committee awarded the student a degree.

These grammatical options might benefit speakers because they allow words to be produced as they become available. And, consistent with this account, speakers are quicker and less prone to errors when producing sentences with verbs that allow variable word orders, like *give*, compared to producing sentences with non-alternating verbs like *donate* (Ferreira, 1996). In fact, it seems that pressures of incremental processing can even lead speakers to produce ungrammatical sentences, at least by stereotypical high-school English-teacher standards. As an (admittedly not very scientifically rigorous) example, a simple web search reveals many instances of *donate* appearing in a double-object dative form, such as “donate me the money.”[1]

So the incremental nature of sentence production means that speakers’ choice of syntactic structures may be influenced by various non-syntactic pressures that operate as a sentence unfolds (Bock, 1982; Bock, Irwin, Davidson, & Levelt, 2003; Ferreira, 1996; Levelt, 1989), and it seems likely that one relevant pressure is to reduce the need to buffer information until it can be produced. (In fact, it seems that speakers attempt to avoid buffering to-be-produced information even in very short, simple utterances, Griffin, 2003.) Little, if any, work has addressed this directly, however two lines of research do so indirectly: investigations of accessibility effects and of given-new ordering.

1.1 Accessibility effects

One class of effects that has been motivated as an instance of memory based pressures on speakers is *accessibility effects*. Accessibility effects refer to speakers’ tendency to produce syntactic structures that allow for earlier production of information that is more accessible (defined in a variety of ways; see below). Essentially, the idea is that production is easier when information can be produced soon after it becomes active, rather than when it must be buffered until the grammar allows for its production.

[1] From: http://news.bbc.co.uk/1/hi/wales/3638757.stm

An important distinction can be made between the *accessibility* and the *availability* of an item in memory (Tulving & Pearlstone, 1966). Availability refers to the existence of an item in memory, whether or not it can be retrieved at any particular moment. Accessibility, in contrast, refers to how easily an item can be retrieved from memory under a particular set of conditions. This distinction is analogous to Bjork and Bjork's (1992) distinction between *storage strength*, corresponding to availability, and *retrieval strength*, corresponding to accessibility. While the availability of a lexical item is obviously a crucial prerequisite for its production, the question of interest in this work is how the moment-by-moment accessibility of conceptual and/or lexical items influences production.[2]

Considerable evidence has shown that both conceptual accessibility (i.e., the accessibility of a specific meaning) and lexical accessibility (i.e., the accessibility of a specific word form) can influence speakers' choice of syntactic structures. Bock and Irwin (1980) used a paradigm where speakers hear a list of questions, followed by a list of answers, then hear the questions again and are required to produce the appropriate answer for each one. Because people typically have a relatively poor memory of the syntactic form of sentences (relative to their memory of the meaning of sentences, e.g., Sachs, 1967), memory-based paradigms such as this constrain speakers to produce sentences with the same basic meaning, but allow for the production of different syntactic structures. Thus, while this paradigm is somewhat odd and unnatural, it does circumvent one of the primary difficulties in language production research, namely how one can

[2] Note that there is a related use of the term *accessibility* that refers to the accessibility of a referent in a discourse, or the (estimated) accessibility of a referent to a listener, rather than to accessibility in memory (see section 1.2.1, below). Here, however, *accessibility* is used in Tulving and Pearlstone's (1966) sense.

exert some control over the wide range of possible things that a speaker could choose to talk about.

Bock and Irwin (1980) manipulated lexical and conceptual accessibility with *wh*-questions. For example, a question like (3a) makes the cowboy (and the rancher) conceptually and lexically accessible (as well as *given*; see section 1.2 below), whereas a question like (3b) makes the horse (and the rancher) conceptually and lexically accessible.

(3a) A rancher received an inquiry from a cowboy about something he needed for his act. What did the rancher do?

(3b) A rancher had a horse who kept running away. What did the rancher do?

Supporting the idea that conceptually and lexically accessible items tend to be mentioned early, subjects were more likely to produce a sentence like *"The rancher sold the cowboy the horse"* in response to (3a), and *"The rancher sold the horse to the cowboy"* in response to (3b), regardless of which answer was originally encoded. Furthermore effects were similar (though smaller) when the accessible item was only conceptually accessible (i.e., when question (3a) was about *"an inquiry from Roy Rogers,"* subjects still tended to answer *"...sold the cowboy the horse,"* and when (3b) was about *"a stallion"* subjects still tended to answer *"...sold the horse to the cowboy"*), suggesting that conceptual accessibility alone is sufficient pressure to warrant early mention. This effect is relatively general, and has been replicated in Japanese (Ferreira & Yoshita, 2003), as well as in college-aged and older adults in a different task (Davidson, Zacks, & Ferreira, 2003).

This experiments discussed so far, however, confounded lexical with conceptual accessibility – it could be the accessibility of the word, rather than of the referent, that leads to early mention. In a second experiment, Bork and Irwin (1980) found similar effects when prompting sentences with single words instead of questions, where referential accessibility likely plays less of a role. That is, speakers tended to produce sentences like *"The falling tree crushed the lumberjack"* when cued by *tree*, and *"The lumberjack was crushed by the falling tree"* when cued by *lumberjack*, suggesting that lexical accessibility in the (relative) absence of conceptual accessibility can influence speakers' use of syntactic structures.

Another confound with this paradigm is that it manipulates both accessibility and the pragmatic factor of givenness (see also Bock, 1977). Thus it may be that these effects do not reflect accessibility, but rather reflect speakers' use of a pragmatic principle to produce given information early. Bock and Warren (1985) addressed this by investigating conceptual accessibility without also manipulating lexical accessibility or pragmatic information. Consistent with Tulving and Pearlstone (1966), they considered conceptual accessibility to be "the ease with which the mental representation of some potential referent can be activated in or retrieved from memory," (p. 50), and operationalized this accessibility by manipulating the imageability of referents (Pavio, Yuille, & Madigan, 1968). Subjects heard a list of sentences that contained nouns differing in imageability, and then recalled sentences when prompted by the verb. Supporting a role of conceptual accessibility in syntactic production, subjects tended to produce imageable nouns earlier in both transitive sentences (e.g., *The doctor administered the shock,* rather than *The shock was administered by the doctor,*) and dative sentences (e.g., *The hermit left the university the property,* rather than *The hermit left the property to the university.*) Subjects did not show this tendency with phrasal conjuncts (e.g., *The hiker fought time and winter*) where the two nouns in the phrase (*time* and *winter*) do not differ in syntactic roles, suggesting that conceptual accessibility affects syntactic choice, but not word order per se. McDonald, Bock and Kelly (1993) operationalized conceptual accessibility as animacy (Cooper & Ross, 1975) and found a similar pattern – subjects often mis-recalled sentences such that animate nouns preceded inanimate ones, but only when the ordering was syntactic.

Bock (1986) manipulated conceptual accessibility in yet another way, by relying on semantic priming. Speakers described pictures with transitive (i.e., active or passive) sentences after reading priming words that were either semantically or phonologically related to the subject or the object of the target sentence. Further supporting the role of conceptual accessibility in syntactic choice, speakers were more likely to produce active sentences when the subject had been semantically primed (e.g, tended to say *Lighting is striking the church* when the prime was *thunder*), and more likely to produce passive sentences when the object had been semantically primed (e.g., tended to say *The church is being struck by lighting* when the prime was *worship*). No differences were found as a function of phonological primes, suggesting that the effect is conceptual in nature.

Fewer studies have directly addressed the effect of lexical, or form-based, accessibility on speakers' choice of syntactic structures, partially because it seems less likely that form-based accessibility is a relevant factor at the point when a speaker must commit to a particular syntactic structure. Levelt and Maassen (1981) found no evidence that speakers preferentially produced hard-to-name words earlier than easy-to-name words (although there was a small, non-significant trend in that direction). This non-effect could not be attributed to an insufficient difference in phonological accessibility because speakers were slower to start speaking when the hard word was produced first. Similarly, McDonald et al. (1993) looked at word length as a different manipulation of lexical accessibility (cf. Zipf, 1949), but found no tendency for speakers to produce short words before long words.

Bock (1987), however, did find evidence for an influence of formal accessibility on speakers' choice of syntactic structures. In a primed picture-description paradigm like that discussed above (Bock, 1986), she had speakers describe pictures that could be described with different word orders (she used both transitive and conjunctive sentences) after primes that were phonologically related to one of the two target noun phrases. For example, speakers described a picture of a horse kicking a cow after reading either a word formally related to horse, like *hearse*, or a word formally related to cow, like *cowl*. In this situation, speakers did show an effect of formal accessibility, such that they tended to produce the phonologically primed word *later* than the unprimed word. While this may seem to contradict the whole motivation of accessibility effects presented above, there is evidence that speakers find it more difficult to produce words after phonological priming (e.g., Colombo, 1986), thus phonological priming can be thought of as making a referent particularly inaccessible. Because the small (non-significant) effect reported in Levelt and Maassen (1981) is similar in size to the effect reported in Bock (1987), it seems likely that lexical accessibility, unlike conceptual or referential accessibility, exerts a small influence on syntactic choice.

1.2 Given-new ordering

Variations in syntactic structure have also been explained with reference to the *information structure* of a sentence (e.g, Halliday, 1967, 1970). Very generally, the information structure of a sentence refers to how the form of that sentence reflects extra-sentential information from the context. For example, it has often been noted that things that are *given* (exact definitions vary, but given information uncontroversially includes things that have previously been mentioned in the discourse) tend to be mentioned earlier than things that are *new*, a phenomenon generally termed *given-new ordering* (Chafe, 1976; H. H. Clark & Clark, 1977; Gundel, 1988; Halliday, 1967, 1970).[3] Evidence for this tendency comes both from corpora (e.g., Thompson, 1990, 1995) and from experimental data (e.g., Bock, 1977; Smyth, Prideaux, & Hogan, 1979). The phenomenon of given-new ordering is not limited to English, but has been demonstrated to exist in a variety of languages (e.g., Ferreira & Yoshita, 2003; Kaiser & Trueswell, 2004) and has even been suggested to be a linguistic universal (E. V. Clark & Clark, 1978).

It should be noted that this distinction between given and new within information structure closely parallels the division between *theme* and *rheme* in thematic structure (the *theme* might be called the starting point of a sentence, and the *rheme* is what is presented within the context set up by the theme). Jackendoff (1972) makes another similar distinction between *focus*, referring to information that the speaker assumes is new for the listener, and *presupposition*, referring to information that the speaker assumes the listener already knows. These terms are often grouped together (given/theme/presupposition and new/rheme/focus) under the broader terms *topic* and

[3] Speakers mark givenness (and newness) in other ways besides with word order. For example, speakers tend to produce given information with less clarity (Bard *et al.*, 2000; Fowler & Housum, 1987), weaker stress (Chafe, 1976; Halliday, 1967), with definite articles (*the* in English; Grieve, 1973), and with less specific terms (e.g., pronouns; Ariel, 1988, 1991; Gundel, Hedberg, & Zacharski, 1993).

comment. While at least some of these concepts are dissociable (e.g., Halliday, 1994), the work presented here makes no attempt to do so, and will use the terms *given* and *new* while recognizing that the conclusions may apply equally well to theme and rheme, foucs and presupposition, and topic and comment more generally.

1.2.1 Given-new ordering as a reflection of listener-based accessibility.

Information can be more or less prominent in the discourse, so can be more or less 'given.' This continuum of givenness has been described as a *givenness hierarchy* (Gundel et al., 1993) or an *accessibility hierarchy* (note that this is a different use of the term *accessible* from that used in section 1.1, see below). Collins (1995), for example, proposes that information can have a status in between *given* (information that has been explicitly mentioned or referred to in the discourse) and *new* (newly introduced or identified in the discourse), and calls this intermediate category *accessible*, meaning information that was not explicitly mentioned but can be inferred, or that was mentioned considerably longer ago. Levelt (1989) further breaks down Collins' definition of given information into two parts: *in discourse model* being somewhat less accessible than *in focus* (Levelt also uses *inaccessible* to refer to what Collins calls *new*). Ariel (1988) makes a similar distinction between levels of accessibility, with recent linguistic material being the most accessible, information from physical surroundings being intermediate, and finally general knowledge being the least accessible.[4]

Note that the use the terms 'givenness' and 'accessibility' in these hierarchies refers to givenness and discourse-accessibility *for the listener*, thus, these approaches suggest that given-new ordering is communicatively motivated. What, exactly, it means

---

[4] Prince's (1981) scale of assumed familiarity makes a similar, though somewhat richer, set of distinctions.

to be given for a listener is not entirely clear, and has been defined in at least three general ways: as being salient for a listener (e.g., Chafe, 1976), as being predictable for a listener (e.g., Halliday, 1970), or as being part of the speaker's and listener's shared knowledge (e.g., H. H. Clark & Haviland, 1977). All three accounts do agree that the reason that speakers produce given (or discourse-accessible) information before new information is to communicate the discourse status of that information to the listener.

This idea has been most fully discussed with respect to referent reduction (i.e., speakers' tendency to refer to given/accessible information with less explicit forms; Ariel, 1988, 1991; Garrod & Sanford, 1982; Gundel et al., 1993; Levelt, 1989), but applies to word order indications of givenness as well. Thompson (1990, 1995) analyzed written narratives (two-thirds of which, incidentally, were murder mysteries) and found that earlier material tended to have a variety of properties suggesting higher listener-based accessibility. Many of these properties could apply equally well to speaker-based accessibility, for example, earlier material tended to be animate, pronominal, specific, and short, however, Thompson also found that referents that were coded as *identifiable by the listener* and/or *accessible to the listener* (what she called *active*) tended to occur early.

Arnold, Wasow, Losongco and Ginstrom (2000) point out two ways that given-new ordering is likely to benefit communication. For one, producing given information early provides a link between what is already established in a conversation and what is about to be communicated, which likely makes for a more cohesive and easy-to-understand discourse. Essentially, this allows a listener to 'ground' new information in relevant previous knowledge. Secondly, the conventional use of given-new ordering provides a way to tell listeners what information they are assumed to already know. That is, speakers can convey the information status of referents simply by placing them in an appropriate order. Assuming listeners find it useful to know what information their speakers think they already know about, this conventionality provides an easy way for this information to be conveyed.

If given-new ordering does help comprehension, and really is produced for the benefit of the listener, then it should be the case that comprehension benefits from the use of given-new ordering. Haviland and Clark (1974) provided relevant evidence by showing that readers more quickly understood sentences where information that was seemingly given (specifically, information that came early in the sentence and was marked with the definite NP *the*) had an antecedent in a previous sentence then when it did not. That is, readers were faster to understand the second sentence in a pair like *"We got some beer out of the trunk. The beer was warm,"* than in a pair like *"We checked the picnic supplies. The beer was warm."* This was not based only on lexical repetition; readers were also faster to understand the second sentence in a pair like *"Last Christmas Eugene became absolutely smashed. This Christmas he got very drunk again,"* than in a pair like *"Last Christmas Eugene went to a lot of parties. This Christmas he got very drunk again."*

Vande Kopple (1982) extended this finding beyond pairs of sentences by having large groups of high school students read paragraphs that did and did not respect given-new ordering. Further suggesting that given-new ordering facilitates comprehension, he found that the paragraphs that respected given-new ordering were reported as easier to read and were better remembered than paragraphs that did not respect given-new organization. Similarly, Smyth et al. (1979) read passages of text to listeners that either did or did not have the order of a target sentence motivated by given information in the prior text, and found that listeners were later more likely to notice changes in the structure of the sentences in the motivated conditions. That is, listeners more often noticed that a sentence like *Arlene brought the leftovers to Wendy* was different from the original sentence *Arlene brought Wendy the leftovers* when the original sentence occurred in a paragraph about Wendy than when it occurred in a paragraph about something else.

Clifton and Frazier (2004) reported somewhat more equivocal evidence for a comprehension benefit from given-new ordering. They manipulated givenness by use of definite and indefinite determiners (*the* and *a*) and constructed dative sentences where the definite NP preceded the indefinite NP (i.e., given before new) or where the indefinite NP preceded the definite NP (i.e., new before given). Readers were faster to judge the grammaticality of definite-indefinite sentences than indefinite-definite sentences, but only for sentences in the double-object dative structure (V-NP-NP), and not for sentences in the prepositional dative structure (V-NP-PP). Based on this asymmetry, they argued that the comprehension preference for given-new ordering is constrained to certain constructions.

Of course, no one seriously proposes that speakers somehow know what is and is not accessible/given to their listeners. Instead, givenness and discourse-accessibility must be based on the speaker's estimation of this accessibility (Levelt, 1989). And, because it seems infeasible for speakers to estimate all relevant bits of listener knowledge each time they want to say something, Clark (1992) suggested that we develop and use mental models of our listeners. Clark's proposal draws largely from work on *common* *ground* (i.e., what I know that you know, and what I know that you know that I know, *ad nauseam*). In particular, he suggests that we use a special type of memory representation called a *reference diary* to keep track of the past co-occurrence of addressees and events, and memory of this co-presence constitutes knowledge of the event being in common ground (H. H. Clark & Marshall, 1981). Common ground thus allows speakers a way to represent what is given and new to their listener. Consistent with a listener-based view of given and new information, Wilkes-Gibbs and Clark (1992) showed that speakers who had previously described ambiguous tangram figures referred to those figures again with short terms and definite reference when talking to the same listener, but used longer terms with indefinite reference when talking to a new listener.

However, a speaker's ability to remember with whom s/he had agreed on particular names for a small set of ambiguous figures is a smaller feat than maintaining a model of what, of all information to be communicated, is given and new to the listener. And when looking at a larger space of referents, this idea that speakers maintain a model of their listener has received only mixed support. Brown and Dell (1987) found that speakers' tendency to more frequently mention atypical information (i.e., things that listeners would likely be unable to infer) did not vary as a function of what they knew their listener to know. That is, in describing pictures with atypical instruments (e.g., stabbing with an icepick), speakers' were no more likely to mention the instrument (icepick) when their listener could not also see the picture. This seems to refute even a basic model of the listener because even visual co-presence did not lead to a listener-based use of common ground. However, Lockridge and Brennan (2002) pointed out that the listeners in Brown and Dell's (1987) study were confederates who might have given subtle signals that they did, in fact, know about the picture. When replicating Brown and Dell's study with naive listeners, Lockridge and Brennan found that, when listeners could not see the picture, speakers did more often mention atypical information, more often marked that information as indefinite (i.e., *an icepick*, rather than *the icepick*) and more often included that information in the main clause.

So it seems that speakers at least sometimes take visual co-presence into account when formulating speech. But only a small fraction of information that is given is also visually co-present, so it remains unclear if speakers actually maintain a representation of the given or new status of referents for their listener. Bard et al. (2000) addressed this question with respect to speakers tendency to produce given information less intelligibly than new information, asking if this reduction is based on givenness for the listener or only for the speaker. In a series of experiments, Bard et al. showed that speakers' clarity was not sensitive to the status of information for their listeners, but only to the speakers' own information status. This was true even when listeners provided direct feedback about their knowledge. Furthermore, a reduction in clarity also followed mention by *other* speakers, thus cannot be simply due to repeated production.

Bard et al. explain their results with reference to the *monitoring and adjustment* hypothesis (Brown & Dell, 1987; Horton & Keysar, 1996), that suggests that speakers are, at least initially, relatively egocentric and speak based only on their own perspective. Speakers can take listeners' needs into account, but this is done by a relatively slow and cognitively demanding process of monitoring their production for information not shared with their listeners (also see Keysar, 1997; Keysar, Barr, & Horton, 1998).

1.2.2 Given-new ordering as a reflection of speaker-based accessibility.

So how can given-new ordering occur if speakers only sometimes take their listeners' perspective into account? One possibility is that given-new ordering only seems to be listener based, but actually emerges because information that is given to the listener is usually also highly accessible for the speaker (in the Tulving and Pearlstone (1966) sense, meaning 'accessible to memory'). That is, given-new ordering might simply be a manifestation of the accessibility effects discussed in section 1.1. Consistent with this claim, speakers seem to have difficulty recognizing when visually present (and thus accessible) objects are *not* part of common ground. That is, speakers tend to refer to things that are visually present, but obscured from their listeners' view, as if they were also part of their listeners' knowledge (Wardlow Lane, Groisman, & Ferreira, 2006).

Arnold et al (2000) investigated whether speakers' tendency to produce given-new ordering (which Arnold et al. refer to as *newness*, i.e., speakers' tendency to produce new information late) might result from *end-weight* or *heaviness*, which refer to speakers' processing-based tendency to produce structurally complex information late (Wasow, 1997). In both an analysis of a language corpus[5] and in a picture-description experiment that manipulated the givenness and complexity of referents, they found that givenness and heaviness made independent contributions to constituent ordering. These results suggest that newness is not reducible to heaviness, and show that information structure plays an important role in the syntactic realization of sentences. However, while these results suggest that information status (givenness/newness) is relevant to syntactic production, they are consistent with both an explanation of given-new ordering that is

[5] Specifically, a corpus of debates from the Canadian parliament, eh?

listener-centered, that is, based on givenness for the listener, *and* with an explanation that is speaker-centered, that is, based on givenness for the speaker.

The only previous study that directly compares the listener-centered and speaker-centered explanations for given-new ordering was reported by Branigan, McLean, and Reeve (2003). They had speakers describe pictures to listeners, who had to decide if the description matched their own pictures, and also manipulated whether the speaker or the speaker and listener saw an initial 'context' picture that introduced one of the referents. Speakers did preferentially produce structures allowing for early mention of the referent from the initial context card (i.e., speakers used given-new ordering), but this was equally true whether or not the listener had also seen the context card (importantly, the speaker knew whether or not the listener saw the card). That is, speakers' were insensitive to the information status of referents from their listeners' perspective, supporting an entirely speaker-centric approach to given-new ordering.

However, Branigan et al. (2003) suggest that the failure to find effects of listener-based givenness may be due to the very simple nature of the listeners' task – that is, perhaps speakers do not bother considering given/new status for the listener when there are only three possible referents – and that a more demanding comprehension task might lead speakers to take their listeners' discourse model into account. Also, the manipulation of givenness used by Branigan et al. manipulated speakers' and listeners' prior exposure to a referent, but it is possible that speakers did not construe that previously-seen referent as being given. This might be the case because, while the speakers were told about what their listeners saw, the 'given' referent was not actually co-present, nor was it part of the prior discourse model. Even if speakers did not construe the manipulated referent as being given, the prior exposure is still likely to have manipulated accessibility, so these results might again demonstrate speakers' tendency to produce accessible information early. Thus the question of whether given-new ordering reflects a form of audience design or simply reflects information accessibility for the speaker remains open.

1.3 The role of memory in language processing

This speaker-centric account of given-new ordering, as well as the usual explanation for accessibility effects, assumes that accessibility is a measure of how quickly a word form gets activated and suggests that speakers find it advantageous to produce information as it becomes active. This pressure, however, runs contra to the demands of the grammar, which generally requires speakers to produce words in a particular order (at least in languages like English, with relatively fixed word order). Because of the variable nature of word retrieval and the fixed ordering requirements of syntax, the production process must rely on some kind of buffer to maintain accessible information that cannot yet be produced. This leads quite naturally to the idea that memory processes might underlie these effects, and that the reason speakers tend to produce accessible information as soon as possible is to reduce demands on memory processes.

1.3.1 Language comprehension and memory

Certainly the idea that language processing involves memory processes is not without precedent (and, as is true for most aspects of psycholinguistics, this precedent has been primarily in the domain of language comprehension). Because language comprehension proceeds incrementally (that is, because speakers process each constituent as soon as possible), a comprehender is often faced with a large amount of information that must be kept active in memory in order to reach a correct interpretation. This can be seen quite easily in center-embedded structures like (4). In sentences such as this one, each of the initial noun phrases must be held in memory until the appropriate verb is found, and this likely exceeds most people's working memory capacity (an explanation first advanced by Miller & Chomsky, 1963).[6]

(4) The dissertation the student the committee questioned wrote was eaten by a dog.

Despite this intuitive relationship between comprehension and memory processes, early attempts to find a relationship between measures of memory and measures of comprehension ability were largely unsuccessful (e.g., Perfetti & Goldman, 1976). Daneman and Carpenter (1980) suggested that the reason a relationship had not been observed was because this early work examined memory only as a system of storage. Drawing on Baddeley's influential model of *working memory* (WM) that proposes both storage and processing functions in short term memory (Baddeley, 1986; Baddeley & Hitch, 1974; Baddeley & Logie, 1999), Daneman and Carpenter (1980) developed the *reading span* task as a measure that taps both the storage and processing functions of WM. This task requires participants to read (or listen to) increasingly larger sets of sentences while attempting to remember the last word of each sentence, and thus requires

---

[6] Note that it is relatively easy to understand the same propositions when split into two sentences: *The student the committee questioned wrote a dissertation. The dissertation the student wrote was eaten by a dog.* But also note that this is may not be a purely syntactic process because semantic constraints can make sentences with multiple center embeddings easier to understand. For example, it is easier to parse the same structure in an example like: *The claim the horse he entered in the race at the last minute was a ringer was absolutely false* (Weckerly & Elman, 1992).

participants to store a number of unrelated words while engaging in language comprehension processes.

Daneman and Carpenter found that performance on this task (termed an individual's *WM-span*) correlated with a variety of measures of comprehension ability. This set off a flurry of studies looking at language comprehension as a function of individual differences in WM span – Daneman and Merikle (1996) reported a meta-analysis of 77 studies investigating the relationship between WM capacity and language comprehension, and concluded that individual differences in WM, as measured by the reading span task, are related to accuracy in such areas as assigning pronominal reference, detecting and resolving ambiguity, making inferences, as well as to performance on general tests of comprehension such as the verbal section of the Scholastic Aptitude Test.

WM has also been shown to play a role in both native-language (e.g., Gathercole, Service, Hitch, Adams, & Martin, 1999; Gathercole, Willis, Emslie, & Baddeley, 1992), and second-language acquisition (Baddeley, Papagno, & Vallar, 1988; Ellis & Sinclair, 1996; MacKay, Meador, & Flege, 2001; Miyake & Friedman, 1998; Service, 1992; Slevc & Miyake, 2006), and simultaneous interpreters, who manage the impressive feat of simultaneously comprehending one language while producing another, have especially high WM capacities (Christoffels, de Groot, & Kroll, 2006). This type of evidence has led to the suggestion that aspects of WM may have evolved as a tool for language acquisition (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998), a conclusion supported by the possibility that the processes underlying language use evolved out of a phonological working memory circuit (Aboitiz, García, Brunetti, & Bosman, 2006).

Furthermore, deficits in WM have been linked to deficits in language processing in patients with various types of neurological disorders (e.g., Bayles, 2003; Caspari, Parkinson, LaPointe, & Katz, 1998; Martin & Romani, 1994; see Smith & Geva, 2000, for a review).

Thus it is uncontroversial that language comprehension involves WM. What is controversial is what *aspects* of language involve WM, and what *type* of WM is involved. There is considerable evidence that WM is not a unitary construct. In particular, considerable work has shown that different pools of resources are involved in the maintenance of verbal and visuo-spatial information (e.g., Baddeley, 1986; Baddeley & Hitch, 1974; Friedman & Miyake, 2000; Logie, 1995; Shah & Miyake, 1996) and, more recently, in different aspects of executive processing (e.g., Friedman et al., 2006; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Caplan and Waters (1999) suggest that the verbal component of WM is further divided into a pool of verbal WM dedicated for syntactic processing and another pool of verbal WM used for non-syntactic, verbally mediated tasks. This domain-specific approach fits well with the idea that syntactic processes are a unique property of human language (e.g., Fitch & Hauser, 2004; Hauser, Chomsky, & Fitch, 2002), and with theories of language (and cognition in general) as made up of specialized limited-purpose modules (e.g., Fodor, 1983).

Support for this domain-specific hypothesis has mostly been based on the memory and syntactic processing performance of patients with various disorders, and on the failure to find statistically significant interactions between syntactic complexity and WM in on-line processing measures. A relatively large body of work, spearheaded by Caplan and Waters, argues for this domain-specific WM (Caplan & Waters, 2002, 2003;

DeDe, Caplan, Kemtes, & Waters, 2004; Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994; Waters & Caplan, 2001, 2002, 2004, 2005; Waters, Caplan, & Yampolsky, 2003; Yampolsky, Waters, Caplan, Matthies, & Chiu, 2002; see also earlier references cited in Caplan & Waters, 1999). A critical claim of this domain-specific hypothesis is that language comprehension relies on both a dedicated WM resource for on-line syntactic parsing, and a domain general form of verbal WM that is involved in what Caplan and Waters (1999) call 'post-interpretive processing.' These post-interpretive processes are reflected in a number of measures of comprehension, especially in off-line measures like plausibility judgments. Caplan and Waters argue that much of the evidence linking WM and syntactic comprehension reflects these post-interpretive processes, thus tests between a domain-specific and domain-general account require on-line measures.

Others have suggested that the storage and processing aspects of sentence parsing is carried out by a single WM system. By this domain-general account, increasing storage demands leads to a corresponding decrease in the cognitive resources available for processing. Within a domain-general account, there are two general types of proposals. One type considers the WM system to play a fundamental role in the architecture of the language system. A second type focuses not on the role of WM in parsing per se, but rather on the role of WM in integrating different sources of information.

An example of the first approach, which considers WM and WM constraints to be a fundamental part of the architecture of the linguistic system, is Gibson's Dependency Locality Theory (also called the Syntactic Prediction Locality Theory; Gibson, 1998;

Grodner & Gibson, 2005). This theory claims that comprehenders can only process and integrate a limited amount of information over time, thus a sentence like (4), above, is difficult to parse because it requires several integrations which presumably overwhelm available WM. A number of studies have found support for this type of theory (e.g., Fedorenko, Gibson, & Rohde, 2007; Gibson, 1998; Grodner, Gibson, & Tunstall, 2002; King & Kutas, 1995).

Another proposal of this general type is that sentence parsing is subject to interference constraints in domain-general WM (Lewis, 1996; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006). This account suggests that WM capacity can be conceived of as the amount of information that can be activated in long-term memory (LTM) without interference (e.g., Anderson & Lebiere, 1998; Cowan, 1999; Oberauer & Kliegl, 2006). Evidence in support of this model comes from Gordon, Hendrick and Levine (2002), who had subjects read sentences of high and low syntactic complexity (object- and subject-clefts, respectively) when under a concurrent three-word WM load. Unlike most previous studies of this type, Gordon et al. manipulated the similarity of the WM load words to the words in the sentence, rather than manipulating the presence or absence of a load. In the matched cases, the words that subjects remembered were semantically similar to words in the sentence, but in the unmatched cases, they were from different semantic categories. For example, for the (syntactically complex) sentence *It was the dancer that the fireman liked before the argument began,* a load trial in the matched condition required maintenance of *poet-cartoonist-voter* whereas in the unmatched condition the load was *Joel-Greg-Andy*. Gordon et al. found that subjects were affected by syntactic complexity more when the load and sentence were semantically similar than when the load and sentence were dissimilar, suggesting that syntactic processing relies on domain-general WM resources.

However, Gordon et al. (2002) only found a significant interaction between syntactic complexity and sentence-load-similarity in accuracy data (based on post-sentence comprehension questions; while their RT data was consistent with a domain-general account, the interaction did not reach significance), which Caplan and Waters (1999) would argue is an off-line task involving post-interpretive processing. Fedorenko, Gibson and Rohde (2006) used the same basic paradigm, but manipulated syntactic complexity with subject- and object-extracted relative clauses (e.g., *The physician who consulted the cardiologist checked the files in the office* and *The physician who the cardiologist consulted checked the files in the office*, respectively) and added the additional manipulation of load difficulty (one vs. three words). Fedorenko et al. found that syntactic complexity interacted with sentence-load similarity with a three-word load (but not with a one-word load). Critically, this pattern occurred in reading times, an on-line measure of processing, supporting a domain-general account of syntactic parsing.

A second approach focuses not on the architectural role of WM in parsing, but rather on the role of WM in integrating different sources of information. By this type of theory, WM capacity predicts the extent to which comprehenders can use different strategies or are influenced by different constraints. For example, in Just and Carpenter's (1992) influential theory, people with lower WM spans are less able than people with higher WM spans to take pragmatic information into account when encountering syntactic ambiguity. Thus syntactic processing is effectively informationally encapsulated for people with low WM spans (or people under high WM load), whereas parsing is more interactive when WM resources are available (other proposals in this vain include Carpenter, Miyake, & Just, 1995; Dick et al., 2001; Felser, Marinis, & Clahsen, 2003; Frazier & Fodor, 1987; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Miyake, Carpenter, & Just, 1994; Pearlmutter & MacDonald, 1995; Rawson, 2007; Traxler, Williams, Blozis, & Morris, 2005).

Within this theoretical approach, Swets, Desmet, Hambrick and Ferreira[7] (2007) conducted what may be the most statistically powerful test of domain-general versus domain-specific WM involvement in syntactic parsing to date (in two experiments, they tested a whopping 538 subjects). They investigated WM involvement in both English and Dutch parsing of relative clause attachment ambiguities in sentences like *The maid of the princess who scratched herself in public was terribly embarrassed,* to see if WM span predicted attachment strategies. In both English and Dutch, lower span speakers were more likely to attach high (i.e., to assume the maid did the scratching) than were higher span speakers. Swets et al. suggested that this is due to different segmentation strategies used by high- and low-WM span readers, reflecting lower-span readers' tendency to break up text into smaller processing 'chunks'. Swets et al. also measured both verbal and spatial WM and, based on their very large sample, were able to use the intercorrelation of these measures to estimate a latent factor corresponding to a very-domain-general form of WM.[8] Interestingly, structural equation modeling showed that both (domain-general) verbal WM and this very-domain-general WM contributed to

---

[7] The other Ferreira.

[8] Swets et al. (2007) use the term "domain-specific WM" to refer to verbal WM (which I have been calling domain-general), and use "domain-general WM" to refer to this latent factor. For the sake of keeping the terminology constant here, I call their domain-general WM "very-domain-general WM".

individual differences in attachment preferences. Swets et al. (2007) follow Kane et al. (2004) in suggesting that this very-domain-general factor corresponds to the ability to control attention.

A final approach to the role of WM in parsing takes issue with the assumption that WM and language processing are separable systems at all. MacDonald and Christiansen (2002) proposed that both memory and language related tasks tap the same underlying processes, which are based on linguistic experience and differences in general processing architecture. They thus point to a single construct, *processing capacity*, that encompasses both language and memory tasks. This approach has not received wide support, however, and at least some evidence speaks against this type of account (Caplan & Waters, 2002; Swets et al., 2007). Nevertheless, especially in light of newer theoretical models of WM that differ from Baddeley's (1986) multiple-component model (e.g., Cowan et al., 2005; Oberauer & Kliegl, 2006; Unsworth & Engle, 2007), it is important to consider how alternative concepts of WM relate to language processing.

In sum, WM clearly plays an important role in many aspects of language comprehension. While the type and role of WM in syntactic parsing are controversial, recent evidence suggests that domain general verbal WM does play a role in syntactic parsing. This role seems to be both an architectural one, such that WM is involved in the processes of creating initial syntactic parses (e.g., Fedorenko et al., 2006) and a constraint-based one, such that parsing strategies and integration of different sources of information rely on WM (e.g., Swets et al., 2007).

1.3.2 Language production and memory

While most work on WM and language has investigated language comprehension, some aspects of language production have also been linked to WM. Bock (1982) points out that people speak slower, less fluently, and less elaborately while driving in traffic, and Kemper, Herman and Lian (2003) found decrements in speed, fluency and elaborateness of production while walking or finger tapping. Kemper and colleagues (Kemper et al., 2003; Kemper & Sumner, 2001) have also shown that the age-related decline in grammatical complexity and propositional content of utterances is attributable to a corresponding age-related decline in WM capacity.

Early experimental evidence linking WM and language production came from Daneman and Green (1986). In this experiment, subjects read short phrases, then were required to replace the last word with an alternate word with a similar meaning. For example, a subject might read *"The belief that we are descended from apes is now quite"* then press a button to see *"WIDESPREAD,"* then generate a context-appropriate replacement word (e.g., *common*) as quickly as possible. They found that speakers with higher WM spans were faster to produce context-appropriate replacements than were speakers with low WM spans, when WM was assessed with a production analog of Daneman and Carpenter's (1980) reading-span task called the *speaking span* task (see section 2.1.3, below, for a description of this task). Daneman and Green interpreted these results as showing that speakers' ability to activate and select appropriate words depends on the same memory capacity that is required to maintain the semantic and syntactic context.

Daneman (1991) extended this finding by showing that speaking span performance was significantly correlated with performance on three measures of verbal fluency. In a speech generation task that required speakers to speak for one minute about a picture, WM span was positively correlated with number of content words produced and with subjective ratings of 'originality and richness of content.' WM span was negatively correlated with how long it took speakers to read aloud a passage of text, and with reading errors on that text. And, in an experimental error-elicitation task (the SLIP procedure, which uses phonological priming to elicit exchange errors, see Baars, 1992), WM span was negatively correlated with error rate. These results again suggest that lexical access relies on WM resources.

WM also seems to be related to the quality of metaphorical language production. Chiappe and Chiappe (2007) had subjects generate metaphors on specific topics, and measured WM with the *listening span* task. In this task, subjects hear sets of sentence fragments (e.g., *"When we are sick we often go to the*") provide words to complete those fragments, then recall the sentence completions. Chiappe and Chiappe compared the metaphors generated by people whose listening span scores fell in the top and bottom quartiles (an 'extreme-groups design'), and found that high span speakers produced better metaphors, even after partialling out variance shared with individual differences in vocabulary (though it should be noted that WM span accounted for only a small amount of unique variance in metaphor production).

Some evidence suggests that WM is related to higher-level aspects of language production as well. While Caspari and Parkinson (2000) found no evidence that individual differences in WM relate to aspects of discourse production, they examined only a very small group of subjects (one brain-damaged patient and four control subjects), thus this may reflect a lack of power. A related aspect of language production that would seemingly rely heavily on memory processes is audience design, or more specifically, common ground (cf. section 1.2.1). Indeed, an early conception of common ground tied it to memory processes, proposing that a specialized memory system (a *reference diary*) was used to keep track of the co-occurrence of specific speakers, listeners and referents (H. H. Clark & Marshall, 1981).

Recently, Horton and Gerrig (2005a, 2005b) have proposed that common ground (and audience design more generally) rely not on a specialized memory system, but on general memory mechanisms. Horton and Gerrig (2005a) showed that this is a feasible claim based on a corpus analysis showing that much of speakers use of common ground could be accounted for by general memory processes. They also provided experimental evidence showing that speakers' use of common ground depends on the situational memory demands in a referential communication task (Horton & Gerrig, 2005b). In this task, speakers described two different sets of cards to two different listeners, and then described both sets of cards to each listener. If speakers successfully use common ground, their second set of descriptions should reflect which cards were previously described to (i.e., are given for) each listener. Horton and Gerrig's (2005b) relevant finding was that speakers more often took common ground into account when the cards were split into easier to remember sets. This supports their claim that common ground relies on general processes of association that cause referents (here, cards) to become associated with specific people. This evidence fits with other work showing that speakers do not take listeners' perspectives into account when under memory load (Roßnagel, 2000), even when listeners give feedback (Roßnagel, 2004).

Note that while the experiments discussed so far demonstrate a role for memory in language production, the role seems limited to either lexical access (Daneman, 1991; Daneman & Green, 1986; see also Kellogg, 2004, for evidence that a WM load impairs lexical access in written production) or conceptual processes (Chiappe & Chiappe, 2007; Horton & Gerrig, 2005a, 2005b; Kemper et al., 2003; Roßnagel, 2000, 2004). In a sense, these conclusions are not very surprising – most people would expect memory mechanisms to be involved in accessing words from a mental lexicon, and to be involved in planning what to say. A more controversial issue (at least based on research on language comprehension) is how WM is involved in the domain of syntax, and whether syntactic production relies on domain-specific or domain-general memory mechanisms.

One type of evidence comes from various language-disordered populations, who often show WM deficits corresponding to syntactic production as well as comprehension. For example, aphasic patients with a lexical-semantic STM deficit (who have trouble maintaining multiple lexical-semantic representations like *hair*, *short*, and *brown*), do poorly when they must produce many words within a phrase (e.g., *short brown hair*) but have little trouble when the concepts can be produced in separate phrases (e.g., *The hair is short and brown*; Martin & Freedman, 2001; Martin, Miller, & Vu, 2004). This suggests that language production at the lexical level is planned phrase-by-phrase, and that WM plays an integral role in the activation and maintenance (and perhaps integration) of items within this phrasal unit of planning.

There is only a small body of work addressing the role of WM in syntactic production in non brain-damaged speakers, and all of this has addressed the process of syntactic *agreement*. Agreement, generally, is when the form of one word (the *target*) changes due to its relationship with another word (the *controller*); one function of this agreement process is to link together non-contiguous linguistic elements that go together conceptually (Bock, 1995). It has often been demonstrated that speakers make agreement errors by, for example, producing verbs that agree with an inappropriate controller (e.g., *The time for fun and games* *are* *over*; Bock & Miller, 1991). Such errors are termed *attraction errors* in which the inappropriate controller is said to have attracted agreement. Bock and Cutting (1992) elicited attraction errors in spoken English, but found that WM span (as assessed with the speaking span task) only correlated with agreement error rates in one of three experiments. However, Fayol, Largy and Lemaire (1994), investigated agreement errors in written French (when subjects were taking dictation), and found that agreement errors were more likely when under a WM load (remembering five words or counting clicks).[9]

Hartsuiker and Barkhuysen (2006) elicited subject-verb agreement errors in spoken Dutch while manipulating WM load and measuring individual differences in WM span with the speaking span task. Specifically, they had speakers produce agreeing sentences, and half of the speakers had to maintain three unrelated words in memory during each trial. Agreement error rates were negatively correlated with WM span, and errors were more likely for the speakers under WM load, but only for the subset of

[9] Although note that Fayol et al (1994) explained their results not in terms of WM involvement in the processing of agreement, but in terms of a controlled process of spelling-verification.

speakers with low WM spans (as assessed by a median split). This suggests that agreement processes do rely on domain-general WM resources, which fits with Badecker and Kuminiak's (2007) proposal that the reason attraction errors occur is because of competition in WM-based agreement processes. More specifically, they suggest that agreement is a process of cue-based retrieval where an agreeing element must retrieve its controller from WM. Attraction errors happen when other memory representations match the retrieval cues for the correct controller. This proposal is conceptually similar to some WM based accounts of syntactic comprehension (e.g., Lewis & Vasishth, 2005), and because agreement, under this proposal, is computed 'on demand', this account is nicely compatible with the incremental nature of production (e.g., Bock, 1982).

As is the case with WM involvement in language comprehension, it is relatively clear and uncontroversial that WM plays a role in conceptual and lexical aspects of language production. The role of WM in syntactic production, however, is less clear, with only one explicit theory (Badecker & Kuminiak, 2007) and both the single theory and all current evidence confined to a single aspect of syntactic production (namely, agreement processes).

1.4 Working memory, accessibility effects, and given-new ordering.

Although accessibility effects (see section 1.1), are often assumed to result from pressures on WM, there is, as of yet, no evidence that actually shows this to be the case. Such evidence would need to show that accessibility effects do, in fact, interact with manipulations of WM load and/or measures of WM capacity. One possibility, of course, is that accessibility effects do not interact with WM factors at all, which would suggest that accessibility effects are not a memory-based phenomenon. However, if WM factors do influence accessibility effects, the way in which these factors interact should reveal the nature of WM involvement in these effects. A WM load might affect speakers' use of accessibility effects in two general ways. One possibility is that a WM load, by virtue of increasing overall production difficulty, could lead to greater use of accessibility effects. This follows from the motivation of accessibility effects as a 'release valve' that speakers use to ease the production process. Alternatively, a WM load could interfere with the maintenance of otherwise-accessible items, thereby leading to reduced use of accessibility effects when under load. This outcome would suggest that, though accessibility effects reflect WM-based accessibility, they might not be used to ease production per se. Alternatively, this second possibility might result if WM load interferes with otherwise-accessible information so much that this information is no longer differentially accessible. These two possibilities are not mutually exclusive, and might emerge differently in speakers with different WM capacities. For example, even if a WM load pressures speakers to rely more on accessibility effects, this might only be possible for speakers with higher WM capacities. For lower WM-capacity speakers, the WM load might interfere with the maintenance of otherwise-accessible information, leading to reduced accessibility effects when under load.

If accessibility effects do interact with WM factors, this raises the important question of what *type* of WM is relevant to these effects. If syntactic comprehension relies on a dedicated, syntax-specific form of WM (a position most strongly associated with Caplan & Waters, 1999), syntactic production might also rely on this domain specific type of WM. By this account, accessibility effects should only interact with a type of WM load that is syntactic in nature. This finding would support the idea that syntactic processing is a specialized cognitive module (in the sense of Fodor, 1983) and is consistent with the claim that syntactic processing is the aspect of language unique to humans (e.g., Hauser et al., 2002). Alternatively, syntactic production might rely on general verbal WM mechanisms (e.g., Just & Carpenter, 1992), predicting that accessibility effects should interact with any type of verbal WM load. This result would support a non-modular account of syntactic processing, but still fit well with multiple-component models of WM (e.g., Baddeley, 1986) by suggesting that syntactic processing relies on verbal WM processes. A third possibility is that syntactic production relies on WM as a whole, in which case any type of WM load (even a non-verbal load) would interact with accessibility effects. This, too, would support a non-modular account of syntactic processing, but would also suggest that the relevant aspect of WM is not a component for verbal storage and processing, but is rather an even more general cognitive mechanism such as focused attention (cf. Swets et al., 2007).

Syntactic production has also been shown to be sensitive to pragmatic factors, such as speakers' tendency to produce given-new ordering (see section 1.2). This tendency has been primarily motivated as a manifestation of audience design (e.g., Collins, 1995), but might instead be an emergent property of accessibility (Branigan et al., 2003) especially based on other evidence that audience design may be primarily motivated by speaker-centric factors (e.g., Horton & Keysar, 1996). An audience design account of given-new ordering would be supported by evidence that WM plays a different role in accessibility effects and given-new ordering, suggesting that speakers really do maintain a model of their listeners knowledge. In contrast, a speaker-centric account would be supported to the extent that given-new ordering patterns like accessibility effects with respect to WM influences, which would fit with other findings that seemingly communicative effects can emerge simply from cognitive pressures on speakers (e.g., Ferreira & Dell, 2000).

The experiments reported here address these three general issues. Experiment 1 investigates if accessibility effects do, in fact, interact with manipulations of WM load and measures of WM capacity. Experiment 2 extends the paradigm developed in Experiment 1 to examine if, and how, given-new ordering interacts with WM constraints. Experiments 3 and 4 contrast the predictions of domain-specific and domain-general accounts of WM in syntactic production are contrasted with respect to accessibility effects and given-new ordering. Finally, Experiment 5 addresses a persistent criticism of many previous experiments on the role of WM in syntactic processing – that effects may be due not to WM processes, but instead to manipulations of difficulty or attentional factors – by contrasting the roles of verbal and spatial WM.

These experiments address the role of WM in syntactic processing, and so are relevant to debates over the modularity or interactivity of language processing more generally. Furthermore, these experiments go beyond past work both by investigating a different type of syntactic process and by working to delineate the type of WM relevant to syntactic processing. The results of these experiments should thus prove informative regarding the role of WM in language processing, and also have implications for the role of language in WM.

1.5 A Note on Data Analysis.

In an influential paper, Herb Clark (1973) pointed out that the most straightforward way of analyzing psycholinguistic data – by simply submitting average responses for each condition for each subject to an ANOVA – is often problematic for psycholinguistic experiments because it ignores the fact that the experimental items are generally sampled from a larger population of items, just as the experimental subjects are sampled from the larger population of language-users (i.e., it implicitly treats items as a fixed-effect). Clark called this the "language-as-fixed-effect fallacy," and showed that it leads to an increased probability of Type I errors (i.e., incorrectly concluding that an effect exists). Clark suggested that the solution is to treat the items-factor as a random effect, just as the subject-factor is treated, allowing effects to be tested with a quasi *F* ratio. Because missing data often makes it impossible to calculate a quasi *F* ratio, Clark further advocates use of the lower bound for the quasi *F* ratio – *minF'* – which can be easily calculated from the results of two separate analyses: one on means obtained by averaging over items (F_1, or the by-subjects analysis) and one on means obtained by averaging over items (F_2, or the by-items analysis).

However, *minF'* came to be viewed as an overly conservative test (see Raaijmakers, Schrijnemakers, & Gremmen, 1999, for a discussion), leading to the current standard practice in psycholinguistics of reporting both the F_1 and F_2 analyses, but not reporting *minF'*. Recently, a number of problems with this practice have been pointed out. For one, Raaijmakers and colleagues (Raaijmakers, 2003; Raaijmakers et al., 1999) point out that doing separate analyses by-subjects and by items does not adequately address the language-as-fixed-effect-fallacy, and further note that the use of *minF'* (or the use of both F_1 and F_2 analyses) is unnecessary in many designs and leads to considerably reduced power (in particular, Raaijmakers argues that F_1, alone, is appropriate in experiments that counterbalance items, or that match sets of items, across conditions).

Another problem that affects all of these analyses (*minF'*, F_1 and F_2, or F_1 alone), is that the averaging across items and/or subjects leads to systematic data loss (Baayen, Tweedie, & Schreuder, 2002).

1.5.1 Mixed-effects models.

A recent approach to these problems is to use *mixed-effects modeling* (also called *multilevel modeling*, *multilevel regression*, and *hierarchical linear modeling*), which simultaneously incorporates both fixed and random effects (Baayen, in press; Faraway, 2006; Pinheiro & Bates, 2000). Such models allow for the inclusion of both subjects and items as random effects, along with the independent variables of interest as fixed-effects, in the same statistical model. By using Maximum Likelihood estimation, fixed effect factors are modeled by means of contrasts (as in traditional ANOVA) and random effects are modeled as random variables with means of zero and unknown variance. The variability of subjects and items are accounted for by these estimates of random-effects variance (typically reported in terms of standard deviations).

These models are gaining in popularity because they offer a variety of advantages over repeated-measures ANOVA. In particular, mixed-effects modeling does not require aggregation across trials and does not require fully balanced data (i.e., it does not require individuals to have the same number of observations). Such mixed-effects models thus avoid the "language-as-fixed-effect" fallacy, while also avoiding the lack of power and applicability of *minF'* or the quasi-*F'* ratio (Baayen, 2004).

1.5.2 Logistic mixed-effects modeling.

Because of these advantages, mixed-effects models have started to appear in the psycholinguistic literature, especially in work measuring reading times (e.g., Baayen et al., 2002; Traxler et al., 2005) and in work using corpus data (e.g., Bresnan, Cueni, Nikitina, & Baayen, 2007). Mixed-models can also apply to the analysis of binary data (like that in the experiments presented here, measuring speakers' choices between syntactic structures). One common, but problematic, approach to analyzing binary outcome variables is to apply standard least squares ANOVA or regression models to proportions or percentages created by averaging over items and subjects respectively. This suffers from a number of problems, however, including the problem of data loss due to averaging discussed above. Furthermore, the mathematical form of a linear ANOVA model is inappropriate for binary responses (e.g., predicted probabilities can be outside the range of 0 to 1), and this approach violates the assumption of homoscedasticity because binary outcome variables are inherently heteroscedastic. That is, the variance of a binary variable is not consistent across values, the upshot of which is that effects close to 50% are overestimated, while effects close to 0% or 100% are underestimated.[10]

A solution to these problems is to use a logistic regression model, which requires no prior averaging and makes no assumptions about distributions of predictor variables (Tabachnick & Fidell, 2001). Instead, the dependent variable is coded as 1 or 0, and the logarithm of the odds of the outcome is modeled as a linear function of the dependent variables. In the experiments presented here, prepositional datives were (arbitrarily) coded as 1, and double object datives were coded as 0, thus the probability that a speaker produces a prepositional dative, p_i, is the probability that $Y_i = 1$, and is modeled in terms

[10] More specifically, the variance of a binomial variable is lower at more extreme values, so as probabilities approach 1 or 0, then the variance approaches zero. For example, when 50% of people have values of 1, the variance equals .25, but when only 10% of people have values of 1, the variance equals .09.

of the *log-odds* of that probability (also called the *logit*), based on the dependent variables (*β*) and their interactions. The basic form of a logistic regression is:

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i}$$

where $i = 1,\ldots,n$. As with mixed-models, parameters are estimated using maximum likelihood, and the statistical significance of individual regression coefficients are determined with the Wald *z* statistic (in which each coefficient is divided by its standard error, yielding a normally-distributed z statistic; Tabachnick & Fidell, 2001).[11]

Combining these two methods, the experiments presented here were analyzed with logistic mixed-effects models, as implemented in the lme4 package (Bates, 2005) in the statistical software R (version 2.4.1; R Development Core Team, 2006). Parameter estimates were calculated with maximum likelihood modeling using Laplace approximations, and the statistical significance of individual fixed-effect estimates were determined with the Wald *z* statistic. Because this is still a relatively new approach within psycholinguistic research, "traditional" by-subjects (F_1) and by-items (F_2) analyses were also conducted on arcsine-transformed proportions (see footnote 11), although it should be noted that by-items (F_2) analyses may not be appropriate for designs such as this where items are counterbalanced across conditions (Raaijmakers, 2003; Raaijmakers et al., 1999). For readability and for purposes of graphical presentation, values are described as proportions rather than as log-odds ratios.

---

[11] Another solution is to arcsine square-root transform proportional measures and analyze these transformed data with a standard linear model. While this approach corrects for the heteroscedasticity of proportional data (Winer, Brown, & Michels, 1991), it still requires averaging across items and/or subjects.

2 Experiment 1: Does WM underlie accessibility effects?

The goal of Experiment 1 was to test the idea that lexical accessibility effects result from WM processes. To this end, speakers described dative-eliciting pictures where either the name of the theme or goal constituent was made accessible, and did so under low or high WM-load conditions. This, plus a measure of individual differences in WM-Span, allowed speakers' choices of syntactic structure to be analyzed as a function of WM capacity limits.

2.1 Method

2.1.1 Participants.

51 students at UCSD participated in exchange for class credit. All but one of the speakers reported English as their native language; data from this non-native English speaker were excluded from the analyses. Data were also excluded from two additional students: one who misunderstood the instructions and performed the task incorrectly, and one because of an equipment malfunction.

2.1.2 Materials and design.

Experimental materials for the dative-eliciting task consisted of 24 pictures of dative actions (see Figure 1 for an example) with dative verbs printed underneath (e.g., *give*). Twelve dative verbs were used – each with two pictures. An additional 12 pictures of transitive actions with transitive verbs printed underneath were used as fillers. These 36 pictures were printed on an approximately 3.5 by 3.5 foot poster in six rows of six pictures each. The pictures in each row had the same agent, but different themes and goals (and patients, in the case of the two fillers in each row). Half of the pictures in each row were oriented with the subject on the right, and half with the subject on the left. Two additional filler pictures were printed on a separate sheet of paper and used for examples and practice during the instructions.

Questions about the six agents in the pictures were created with both short and long subject NPs (e.g., *What's going on with the pirate?* and *What's going on with the pirate with the peg-leg?*) to manipulate WM-load. Each picture was paired with a short and long question, and with both a theme and goal cue to create four versions of each item, counterbalanced across four lists such that each item appeared an equal number of times in each condition across the experiment, and such that each speaker was presented with each item only once.

Figure 1: Example of the dative-eliciting pictures used in all experiments.

Materials for the speaking-span task were 70 unrelated seven-letter words, arranged in five sets, each of two, three, four, and five words.

2.1.3 Procedure.

The experiment was administered with PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). Speakers sat facing the computer screen, and the experimenter sat beside the speakers such that the computer screen was obscured from the experimenter's view by a cardboard divider. Both the speakers and the experimenter could easily see the poster of line-drawn pictures.

Each experimental session lasted about 40 minutes, and consisted of two parts. The first part began with instructions and two practice items (which were not used in the experimental trials). In the experimental trials, speakers described dative-eliciting pictures from a poster on the wall (e.g., of a pirate giving a monk a book) in response to questions asked by the experimenter about the subject of a picture, which was described with a short (e.g., *What's going on with the pirate?*) or long (e.g., *What's going on with the pirate with the peg-leg?*) noun phrase. Speakers then read one of the post-verbal NPs (*BOOK* or *MONK*) on the computer screen (which was not visible to the experimenter) thereby making that NP more accessible, and then described the relevant picture to the questioner. Speakers were instructed to describe the picture using the same noun phrase that the experimenter used in asking the question, and using the verb printed under the picture.

In the second part of the session, speakers were administered a version of the Daneman and Green (1986) speaking-span task. In this task, speakers saw words presented individually on the computer screen for 1 second each, and then heard a beep. At the end of each set (i.e., after the beep), speakers were instructed to generate a separate sentence containing each word for as many words as they could remember. Sets increased in size from two to five words, with five trials of each set size. Speakers were instructed to produce grammatical sentences, however no restrictions were given as to the length or complexity of the sentences or to the order of recall.

2.1.4 Analysis.

Speakers' sentences from the picture-description task were transcribed and coded as prepositional datives (PD), double-object datives (DO), or other. Trials in which speakers did not produce a dative sentence, did not produce both post-verbal noun phrases, or did not produce the correct verb were excluded from the analysis, causing 13.2% of critical trials to be excluded (11.6% of trials in the low-WM load condition and 14.8% of trials in the high-WM load condition). Trials in which speakers did not produce the appropriate subject NP length (i.e., produced a short subject NP in the high-WM-load condition) were also excluded, leading to the exclusion of an additional 5.6% of critical trials in the high WM-load condition.

The speaking span task was scored online by the experimenter, who marked each word that was successfully recalled on a score-sheet. A word was considered to be successfully recalled when the speaker produced a sentence that included either the target word or a morphological variant of the target word (e.g., *contained* for *contain*). Sentences that contained two target words from a given set were only scored as one correct recall. Following Daneman and Green (1986), a speaker's WM-span was operationalized as the highest set size at which that speaker successfully produced sentences for all words in a majority of sets (cf. Daneman & Carpenter, 1980; Miyake et al., 1994). For example, a speaker who did not successfully recall even the two-word sets more than once was given a score of 1 (note that every speaker in these five experiments were able to produce sentences for at least one word on a majority of trials), and a speaker who recalled at least three sets of five words was given a score of 5. Additionally, speakers who recalled all words for two out of the five sets of a given set size were given 'partial credit,' so a speaker who recalled all words for two of the five four-word sets would be given a score of 3.5. Thus WM-span ranged from 1 to 5, in intervals of 0.5.

2.2 Results and discussion.

Figure 2 shows the proportion of prepositional datives produced as a function of cue-type (theme or goal) and WM-load (short vs. long subject NP). Surprisingly, speakers did not show standard accessibility effects at all – that is, speakers produced no more prepositional datives when cued with the theme argument than when cued with the goal argument, as reflected by no significant main effect of cue-type. Speakers did produce 5.5% more prepositional datives in sentences with long subject NPs (i.e., when under WM-load), resulting in a significant main effect of WM-load ($z = 6.39$, $p < .001$), however the (non-existent) accessibility effect did not interact with WM-load.[12]

The failure to find basic accessibility effects is surprising given that accessibility effects have been demonstrated several times previously (e.g., Bock, 1977; Bock & Irwin, 1980; Ferreira & Yoshita, 2003; McDonald et al., 1993; see section 1.1 above). One explanation for the failure to find accessibility effects in this experiment comes from the nature of the experimental task. Most previous experiments investigating accessibility effects have used sentence recall tasks, which are quite different than the question-answering picture-description task used here. It seems likely that sentence recall tasks impose stronger memory demands than does this task, in which the to-be-described picture provides all information necessary to produce the sentence. If the lack

[12] Traditional by-subject and by-item analyses reveal the same pattern of results: a main effect of WM-load significant by-subjects ($F_1(1,47) = 13.84$, CI = ±0.175), but not by-items ($F_2(1,23) < 1$, CI = ±0.528), and no other significant effects.

of accessibility effects reported above are, in fact, due to this tasks' relatively low memory demands (and perhaps this task's relatively weak manipulation of accessibility; see the discussion of Experiment 2, below, for more on this possibility), then accessibility effects are likely to interact with speakers' memory spans, with lower-span speakers showing effects more comparable to those found in previous experiments. Similarly, any influence WM-load has on accessibility effects might differ as a function of WM-span.

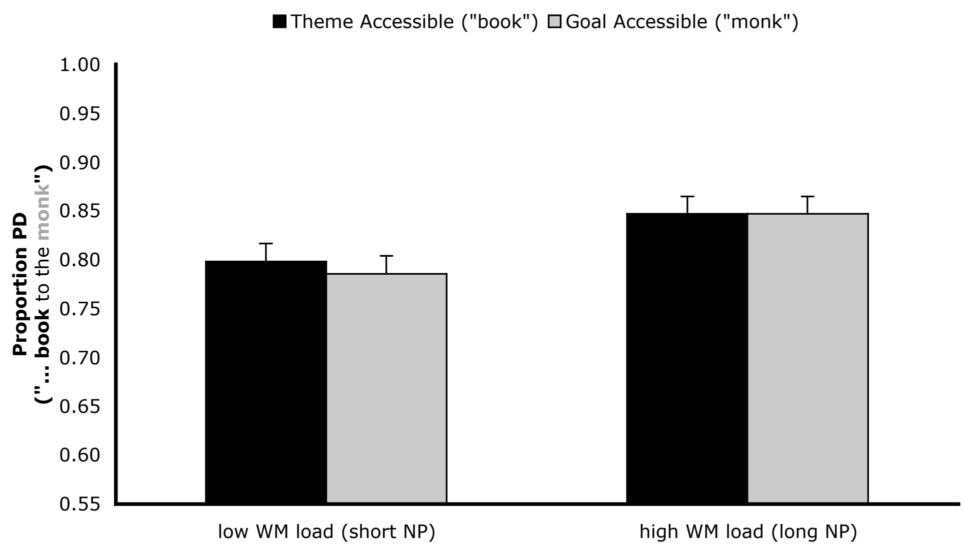

Figure 2: Proportion of prepositional dative (PD) sentences produced in Experiment 1 as a function of cue-type (i.e., which post-verbal argument was made accessible: theme or goal) and WM load (short or long subject NP). Error bars represent standard errors.

Supporting this account, an expanded statistical model that included WM-span along with cue-type and WM-load as fixed-effect factors (and both subjects and items as random-effect factors) revealed a significant 3-way interaction between the three fixed-effect factors ($z = -2.37$, $p < .05$)[13]. Figure 3 shows speakers' average accessibility effects as a function of WM load for speakers with below-average and above-average performance on the speaking-span task. Inspection of Figure 3 reveals that speakers with below-average WM-spans showed accessibility effects in sentences with short subject NPs, but not in sentences with long subject NPs, where the WM-load presumably overwhelmed their WM capacity. In contrast speakers with above-average WM spans only showed accessibility effects in sentences with long subject NPs, presumably because the high WM load adds pressure to rely on accessibility effects to ease production.

Interestingly, there is some indication of "backwards" accessibility effects for the lower-span, high-load, and higher-span, low-load conditions. That is, in these conditions, speakers tended to choose structures allowing the accessible NP to be produced *later* rather than earlier. One possible explanation for these "backwards" effects is that less accessible words are especially sensitive to output interference, and so it is advantageous to produce those words early when both constituents can be maintained in WM (this is similar to Brainerd's (1995) explanation of *cognitive triage*, but see Wixted, Ghadisha and Vera (1997) for evidence against this account). It is not clear, however, why accessibility would sometimes favor early mention and other times

---

[13] An analysis on arcsine-transformed proportions revealed a 3-way interaction significant by subjects ($F_1(1,47) = 4.35$, CI = ±0.36) but not by items ($F_2(1,23) = 1.25$, *n.s.*) although maximum likelihood estimates did not converge, suggesting an overfitted model. Another approach is to classify WM-Span as above- or below-average (as in Figure 3) and perform a 2×2×2 ANOVA with WM-Span nested under Subject. While this did not yield a significant interaction, a similar analysis that included only at the top & bottom quartiles of WM-Span revealed a significant 3-way interaction by subjects ($F_1(1,23) = 6.03$, CI = ±0.25), though it was not possible to carry out the complementary items analysis because of missing values.

favor late mention (though one possibility is that the *degree* of accessibility plays a role). Of course, this backwards effect could also simply be due to "noise" in the data, so its cause will be revisited if a similar effect appears in the next four experiments.

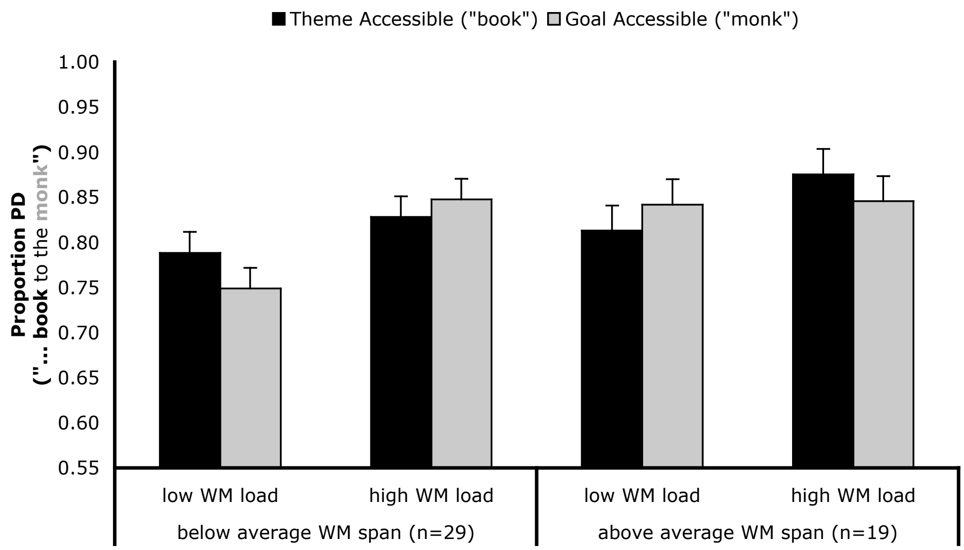

Figure 3: Proportion of prepositional dative (PD) sentences produced in Experiment 1 as a function of cue-type (theme or goal), WM load (short or long subject NP) and WM Span (below or above average). WM span was treated as a continuous variable in the analyses, but speakers were divided into two groups for illustrative purposes. Error bars represent standard errors.

It is worth noting that speakers overwhelmingly preferred to use prepositional dative structures in these data (81.9% of all dative sentences produced in Experiment 1 were prepositional datives). This is somewhat surprising because data from corpora of naturally produced language suggest that prepositional datives and double-object datives are about equally common. For example, Zhong, Stent and Swift (2006), in an analysis of dative verbs from four different automatically-parsed corpora, found that about 48% appeared in sentences of the form V-NP-NP (i.e., were probably double-object datives), and 52% were of the form V-NP-PP (i.e., were probably prepositional datives). While this preponderance of prepositional datives was unexpected, there is no reason to assume it is related to the interaction between accessibility effects and WM based factors, thus it should not detract from the main pattern of results. It does, however, again point to the advantages of using logistic models that are not constrained by the underestimation of proportions near extreme values.

One likely explanation for the high rate of prepositional datives in these data is simply that the theme argument is always between the subject and goal arguments. So if a speaker first finds the subject, then scans the picture starting from the subject, the order in which the speaker sees the arguments will correspond to subject-theme-goal, i.e., a prepositional dative. This explanation suggests that the prepositional dative structure is the preferred form, at least for describing this type of picture. In fact, there is evidence that the prepositional dative structure is the preferred ('unmarked') form more generally. For example, adults learning English as a second language tend to acquire the prepositional dative before the double-object construction (Mazurkewich, 1984). And, while some data suggest that children more commonly use the double-object construction (Campbell & Tomasello, 2001; Snyder & Stromswold, 1997), this is complicated by a dissociation between token and type frequency. That is, the most commonly used dative verbs are more often used in the double-object structure, which leads to a high token frequency of double-object structures. However, children use a greater *number* of verbs in prepositional dative structures, reflected in a higher type frequency for prepositional datives (Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989), and children preferentially use prepositional datives when generalizing dative structures to new verbs (Conwell & Demuth, 2007), suggesting that children treat the prepositional dative as the more broadly applicable (i.e., default) structure.

The default status of the prepositional dative (and/or the markedness of the double-object dative) may also account for the fact that speakers more often produced prepositional dative structures when under WM-load (reflected in the main effect of WM-load mentioned above). Speakers might be more inclined to rely on the default structure (or the structure that better maps onto the visual structure of the picture) when production is more difficult, again showing a role of WM resources in syntactic production. The default status of the prepositional dative may also explain the surprising asymmetry found in Clifton and Frazier's (2004) study (see section 1.2.1, above), where readers only benefited from given-new ordering in the double-object dative construction. If prepositional datives are the default structure, they probably are less likely than double-object datives to consistently reflect pragmatic information like givenness, and thus comprehenders may learn that order information is only informative (with regard to information structure) in the context of double-object datives. Alternatively, this effect might instead simply reflect a preference to use prepositional dative structures in sentences with long subject NPs.

3 Experiment 2 – Does WM underlie given-new ordering?

As mentioned in the introduction, speakers not only tend to mention more accessible information early, but also tend to order given information before new information. This preference for given-new ordering has been noted often in the linguistics literature (e.g., Chafe, 1976; Halliday, 1970), and has even been suggested to be a property of all languages (E. V. Clark & Clark, 1978). Given-new ordering is, by some accounts, a manifestation of audience design; that is, speakers' use given-new ordering in order to facilitate comprehension by their listener (H. H. Clark & Haviland, 1977; Vande Kopple, 1982). Others argue that given-new ordering is not based on a cognitive model of the listeners' knowledge, but occurs simply because information that is given (by, for example, previous mention or perceptual salience) is often also relatively more accessible than new information. So, by this second account, given-new ordering is essentially the same as accessibility effects and only coincidentally facilitates comprehension (Branigan et al., 2003). Of course, interactions of WM with given-new ordering would still be consistent with a model of givenness as a property of *WM-dependent* audience design, however showing no effect of WM on given-new ordering would suggest that givenness cannot be reduced to accessibility.

3.1 Method.

3.1.1 Participants.

48 students at UCSD participated in exchange for class credit. All speakers reported English as their native language.

3.1.2 Materials, design, and procedure.

The experimental design and materials were identical to those used in Experiment 1. The procedure was very similar to that of Experiment 1, with one major exception: instead of manipulating the accessibility of a post-verbal NP (by showing that word on a computer screen privileged to the speaker), one post-verbal argument was made *given* by the interlocutor/experimenter. Specifically, speakers described pictures (e.g., of a pirate giving a book to a monk), in response to questions asked by the experimenter, who manipulated the given argument by mentioning one of the post-verbal NPs (underlined in the examples below; contrast sentences (5a) and (6a) with sentences (5b) and (6b)) and manipulated WM-load with the length of the subject NP as in Experiment 1 (contrast sentences (5a) and (5b) with (6a) and (6b), below).

(5a) What's going on with the pirate and <u>the book</u>?
(5b) What's going on with the pirate and <u>the monk</u>?
(6a) What's going on with the pirate with the peg-leg and <u>the book</u>?
(6b) What's going on with the pirate with the peg-leg and <u>the monk?</u>

The experimenter sat facing away from the poster (unlike in Experiment 1) so that the pictures themselves would not be construed as being given information, and the given argument was always referred to with the definite determiner *the*, which has been shown to be consistently related to givenness (Grieve, 1973). Finally, participants were asked to read instructions from a sheet of paper instead of from the computer screen. The speaking-span task was administered just as in Experiment 1.

3.2 Results and Discussion.

Following the same criteria used in Experiment 1, trials in which speakers did not produce a dative sentence, did not produce both post-verbal noun phrases, or did not produce the appropriate verb were excluded (accounting for 3.6% of all critical trials; 3.1% of trials in the low WM-load condition and 4.2% of trials in the high WM-load condition), as were trials in which speakers did not produce the appropriate subject NP length (i.e., produced a short subject NP in the high-WM-load condition; an additional 4.0% of trials in the high load condition).

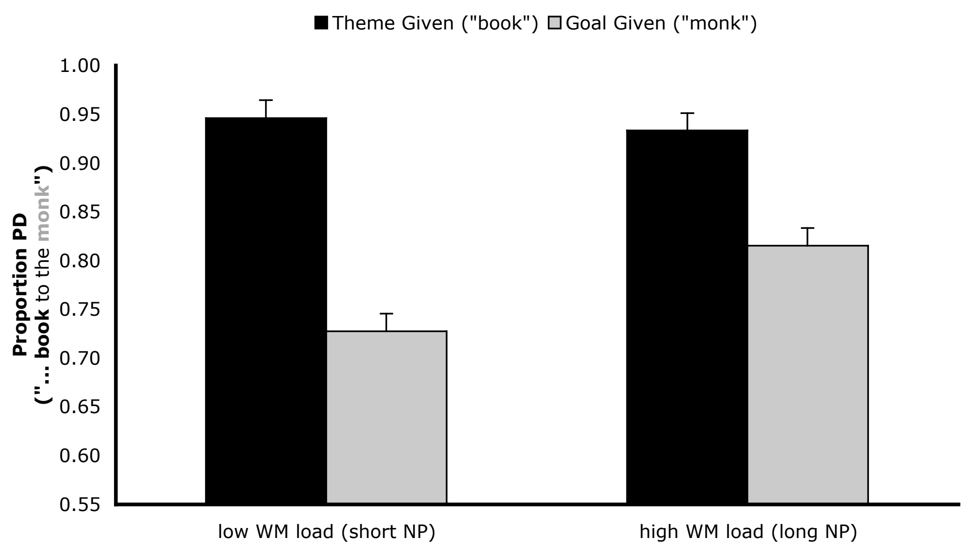

Figure 4: Proportion of prepositional dative (PD) sentences produced in Experiment 2 as a function of cue-givenness (i.e., which post-verbal argument was made given: theme or goal) and WM load (short or long subject NP). Error bars represent standard errors.

Figure 4 shows the proportion of prepositional datives produced as a function of cue-givenness (theme or goal) and WM-load (short vs. long subject NP). As in Experiment 1, speakers produced significantly more prepositional dative structures when under a higher WM load ($z = -3.15$, $p < .01$), perhaps reflecting a default status for prepositional dative structures (see above). There was also an effect of givenness, such that speakers tended to produce given information before new information ($z = 5.98$, $p < .001$) but, importantly, a significant interaction between cue-givenness and WM-load ($z = 2.41$, $p < .05$) shows that this tendency was notably reduced when under WM-load (a

22% givenness effect in sentences with short subjects, but only a 12% effect in sentences with long subjects).[14]

In Experiment 2, WM-span did not interact with cue-givenness and WM-load ($z < 1$, *n.s.*), suggesting that WM-load affects given-new ordering in the same way, regardless of WM-span. This differs from the pattern of results in Experiment 1, where a three-way interaction showed that the effect of WM load on accessibility effects did depend on WM-span. This difference could indicate that accessibility effects and given-new ordering rely on WM in different ways, as predicted from an account where given-new ordering arises not from information accessibility per se, but from a non-accessibility-based (but WM-mediated) form of audience design (e.g., H. H. Clark & Marshall, 1981). This conclusion is suspect, however, as can be seen from Figure 5, which shows Experiment 2 broken down by below- and above-average WM span. Note that the lower-span speakers show the same pattern as the lower span speakers in Experiment 1 (see Figure 3): greater use of given-new ordering when not under WM load. The three-way interaction with WM-Span found in Experiment 1 seems to be driven by the higher-span speakers in that experiment, who did not produce accessible information early when *not* under WM-load, which was clearly not the case for the higher-span speakers in Experiment 2. It is not clear why the higher-span speakers did not produce accessible information early in the no-load condition of Experiment 1, but

---

[14] Traditional by-subjects and by-items analyses give a similar pattern of results: a main effect of WM-load significant by subjects ($F_1(1,47) = 5.81$, CI = ±0.16), though not by items ($F_2(1,23) < 1$, CI = ±0.21, *n.s.*), a significant effect of cue-givenness ($F_1(1,47) = 46.53$, CI = ±0.23; $F_2(1,23) = 41.77$, CI = ±0.25) and a significant interaction between these factors ($F_1(1,47) = 10.13$, CI = ±0.17; $F_2(1,23) = 8.72$, CI = ±0.16).

it is equally unclear how a different pattern only for higher-span speakers could represent a different role of WM in given-new ordering and accessibility effects.

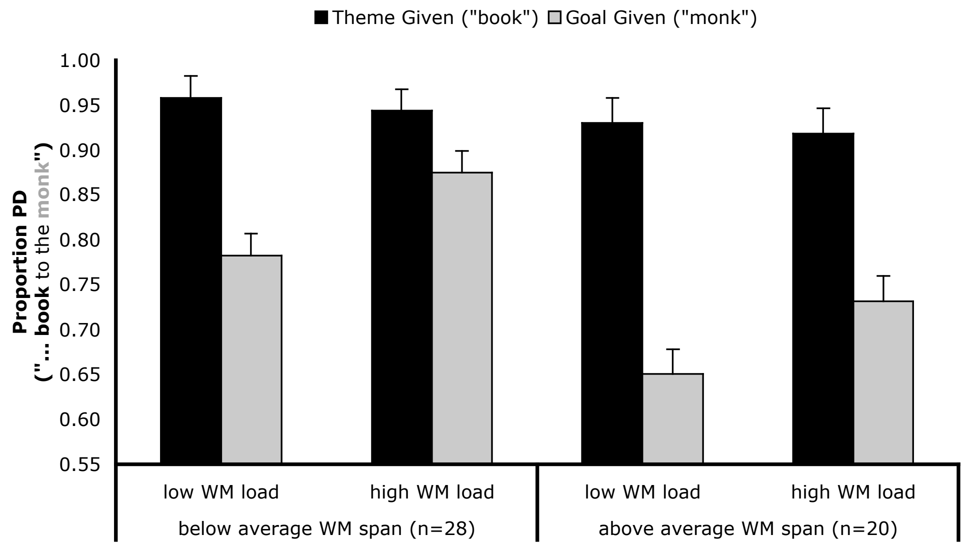

Figure 5: Proportion of prepositional dative (PD) sentences produced in Experiment 2 as a function of cue-givenness (theme or goal), WM load (short or long subject NP) and WM Span (below or above average). WM span was treated as a continuous variable in the analyses, but speakers were divided into two groups for illustrative purposes. Error bars represent standard errors.

One account for the difference between these experiments has to do with the strength of the manipulation. Note that the given-new ordering effects in Experiment 2 were considerably larger than the accessibility effects in Experiment 1. This, plus the assumption that given-new ordering reflects information accessibility, suggests that the manipulation used in Experiment 2 was a considerably stronger manipulation of accessibility than was the manipulation used in Experiment 1. Perhaps the higher span speakers in Experiment 1 simply did not find the "accessible" referent accessible enough to warrant early mention, except when under WM load. Of course, another reasonable account for the differences between these experiments is that the two experiments used a different sample of speakers. Although the distribution of WM-span in these experiments was similar, speakers in Experiment 1 displayed a greater range of WM-spans than did speakers in Experiment 2 (the mean WM-span in Experminent 1 was 2.25 (SD = 0.66) and ranged from 1 to 4.5, whereas the mean WM-span in Experiment 2 was 2.19 (SD = 0.46) and ranged from 1 to 3).

This issue of the nature of givenness and its relation to accessibility will be revisited below, in the discussion of Experiments 3 and 4. These next two experiments, however, were designed to address a different question: what type of WM underlies accessibility effects and given-new ordering?

4 Experiments 3 and 4: Domain-specific or domain-general WM?

As noted in the introduction, it has been suggested that the processing involved in linguistic abilities is distinct from other types of cognitive processing (e.g., Caplan & Waters, 1999; Pinker, 1994), and conversely that domain general WM is involved in language processing (e.g., Just & Carpenter, 1992; King & Just, 1991; Miyake et al., 1994). Experiments 1 and 2 suggest that WM constraints can affect linguistic processing; at least as far as accessibility effects and given-new ordering are concerned. Importantly, however, the imposed load in Experiments 1 and 2 has been a load *internal* to the production system because WM-load was manipulated by varying the length of the to-be-produced subject NP. Thus, the WM-load in the first two experiments did not just require maintenance of some verbal load, but also had a syntactic component (i.e., it led to the production of more syntactically complex sentences). Thus both types of syntactic processing accounts – based on domain-general or based on domain-specific WM resources – predict a sentence-internal load to interact with accessibility effects, as found in Experiments 1 and 2.

However, these two accounts make different predictions with respect to a WM load that is *not* internal to the language production process. An account where syntactic production relies on domain-specific WM resources predicts that a sentence-external load without a syntactic component should not affect speakers syntactic processing, and thus should not affect speakers' use of accessibility effects and/or given-new ordering. In contrast, a domain-general account predicts the same pattern of results as found in the first two experiments because a sentence external load should tax the same WM resources that are involved in syntactic production. Of course, a sentence-external load might still affect syntactic production differently than a sentence-internal load even if syntactic production does rely on domain-general WM. This might be the case because the sentence-internal load used in Experiments 1 and 2 – an elaboration of the subject NP – is more similar to the rest of the sentence than is an unrelated external WM load. Under the reasonable assumption that maintenance in WM is susceptible to similarity, then this sentence-internal load might be predicted to exert greater interference with WM processes. However, finding that sentence-internal and sentence-external WM loads exert similar effects on syntactic production would clearly support a domain general account. Experiments 3 and 4 will address this question; namely, does a WM load need to be sentence internal (i.e., syntax-specific) in order to affect syntactic choice, or can a non-sentence internal load cause similar effects?

4.1 Experiment 3: What type of WM underlies accessibility effects?

4.1.1 Method.

4.1.1.1 Participants.

49 students at UCSD participated in exchange for class credit. Data were excluded from one student because of an equipment malfunction. All speakers reported English as their native language.

4.1.1.2 Materials, design, and procedure.

The experimental design and materials were identical to those used in Experiments 1 and 2. Unlike in Experiments 1 and 2, however, the WM load was not manipulated with the length of the subject NP. Instead, at the beginning of each load trial, speakers were presented with two unrelated words in boldfaced, blue, 14-point font, which they were instructed to remember, and recall aloud when prompted at the end of the trial. With the exception of this WM load manipulation, the procedure was identical to that of Experiment 1: on each trial, speakers were asked a question by the experimenter (all questions had short subject NPs, e.g., *What's going on with the pirate?*), read a word corresponding to a post-verbal NP of the to-be-described picture (e.g., *BOOK* or *MONK*), thereby making that NP more accessible, then described the appropriate picture from a poster on the wall (e.g., of a pirate giving a monk a book; see Figure 1). The speaking-span task was administered just as in Experiments 1 and 2.

4.1.2 Results and discussion.

Trials in which speakers did not produce a dative sentence, did not produce both post-verbal noun phrases, or did not produce the appropriate verb were excluded from the analysis (9.7% of all critical trials; 10.2% of trials in the WM load condition and 7.8% of trials in the no load condition), as were trials in which speakers did not correctly recall at least one word of the WM-load (another 2.4% of trials in the WM load condition).

Figure 6 shows the proportion of prepositional datives produced as a function of cue-type (theme or goal) and WM-load (none vs. two words). As in Experiments 1 and 2, speakers produced significantly more prepositional dative structures when under a WM load ($z = 3.47$, $p < .001$). While there was no main effect of cue-type ($z = 1.15$, *n.s.*), the significant interaction of WM-load and cue-type ($z = -1.99$, $p < .05$) reflects speakers' use of accessibility effects *only* when not under a concurrent WM-load (an

8% accessibility effect in the no-load condition, but a -1% effect in the concurrent 2-word load condition).[15]

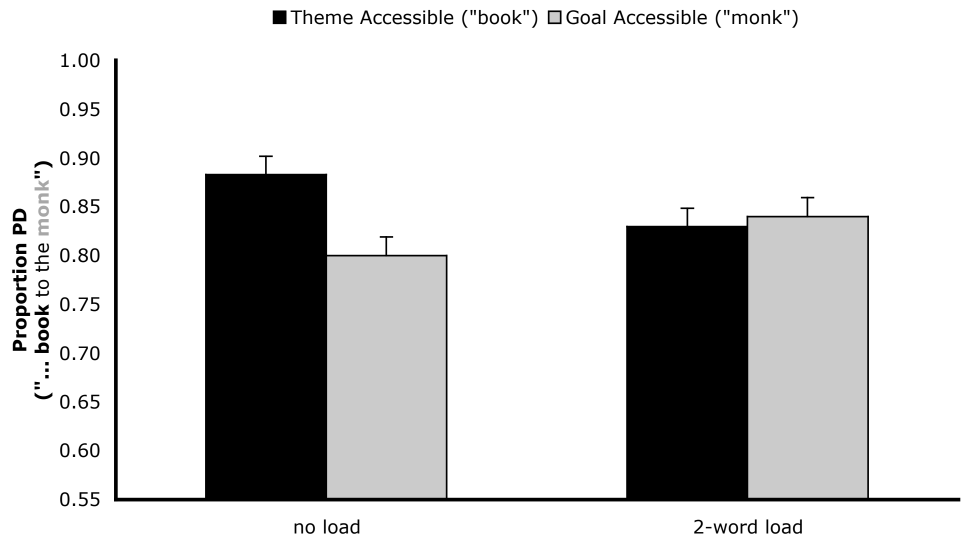

Figure 6: Proportion of prepositional dative (PD) sentences produced in Experiment 3 as a function of cue-type (theme or goal) and WM load (no load or sentence-external verbal load). Error bars represent standard errors.

Speakers did not show any evidence of accessibility effects when under an external WM load, giving clear evidence that accessibility effects rely on a type of WM that need not be sentence internal. These results do not exactly mirror the results of Experiment 1, however, in that WM-span did not interact with cue-accessibility and WM-load. This may be because the external load manipulation is more demanding than the sentence-internal load imposed in Experiment 1 (some evidence for this is that speakers forgot at least one of the load words on an average of 13.9% of WM-load trials in Experiments 3 and 4, but only forgot the long subject NPs on an average of 4.8% of WM-load trials in Experiments 1 and 2). This more difficult load may have led even

---

[15] Traditional by-subjects and by-items analyses also show a significant interaction between WM-load and cue-type by subjects ($F_1(1,47) = 5.82$, CI = ±0.18), but not by items ($F_2(1,23) = 1.51$, *n.s.*). Neither main effect reached significance.

the higher-span speakers to rely more generally on mechanisms that ease sentence processing. That is, in situations where memory is often heavily taxed, speakers might tend to rely more on syntactic flexibility to ease the production process even in sentences that are not particularly taxing.

There are at least two ways that this could come about. For one, it may be that the more difficult load in Experiment 3 caused speakers' WM to be somewhat taxed overall in Experiment 3 (i.e., WM was somewhat 'fatigued' even in the no-load conditions), thus speakers relied more on cognitive mechanisms to ease production when they were able to do so. Another possibility is that cognitive strategies like producing more accessible information earlier are broadly applied. If so, speakers may be more likely to use such strategies in harder tasks, when WM is sometimes overloaded, than in tasks that are relatively easy. Of course, as mentioned above, this could instead simply reflect a different group of speakers in these experiments.

4.2 Experiment 4: What type of WM underlies given-new ordering?

If given-new ordering also occurs because of pressures on WM, then an external WM load should reduce speakers' use of given-new ordering, just as the sentence internal WM load did in Experiment 2. To address this, Experiment 4 replicated Experiment 2, but imposed a sentence-external WM load.

4.2.1 Method.

4.2.1.1 Participants.

48 students at UCSD participated in exchange for class credit. All speakers reported English as their native language.

4.2.1.2 Materials, design, and procedure.

The experimental design and materials were identical to those used in Experiments 1, 2 and 3. The procedure was identical to that of Experiment 2 (in which the experimenter manipulates the givenness of post-verbal NPs by mentioning one in the question, e.g., *What's going on with the pirate and the apple?*), except that the WM load was an external two-word load as in Experiment 3. On load trials, the experimenter first said aloud two unrelated words for speakers to maintain throughout the question and picture description, and all questions (and thus all answers) had short subject NPs. The speaking-span task was administered just as in the three previous experiments.

4.2.2 Results and discussion.

As in Experiment 3, trials in which speakers did not produce a dative sentence, did not produce both post-verbal noun phrases, or did not produce the appropriate verb were excluded from the analysis (5.9% of trials in each condition), as were trials in which speakers did not recall at least one word of the WM-load (an additional 8.3% of trials in the WM load condition).

Figure 7 shows the proportion of prepositional datives produced as a function of cue-givenness (theme or goal) and WM-load (none vs. two words). Unlike in the previous three experiments, speakers were no more likely to produce prepositional dative structures when under a WM load ($z = 1.31$, *n.s.*). There was a main effect of cue-givenness ($z = 6.92$, $p < .001$), reflecting speakers' strong tendency to produce given information before new information. However, this tendency was reduced when speakers were under a concurrent WM-load, reflected in a significant cue-givenness by WM-load interaction (z = -2.07, $p < .05$).[16]

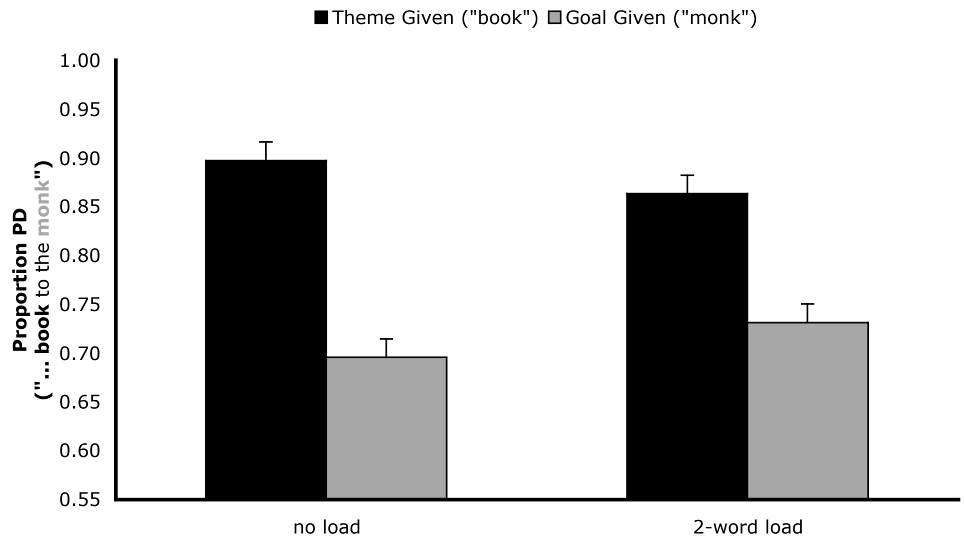

Figure 7: Proportion of prepositional dative (PD) sentences produced in Experiment 4 as a function of cue-givenness (theme or goal) and WM load (no load or sentence-external verbal load). Error bars represent standard errors.

These results support an account of speakers' tendency to use given-new ordering as an emergent property of accessibility, based on the fact that given information tends to also be accessible information. This account fits well with other evidence that general processing mechanisms underlie seemingly communicative phenomena such as speakers' use of optional words and prosody to avoid ambiguity (Ferreira & Dell, 2000; Kraljic & Brennan, 2005) and speakers' use of common ground (Horton & Gerrig, 2005a, 2005b).

---

[16] Traditional by-subjects and by-items analyses also show a significant main effect of cue-givenness ($F_1(1,47) = 25.69$, CI = ±0.33; $F_2(1,23) = 26.42$, CI = ±0.23) and a significant interaction between WM-load and cue-givenness by subjects ($F_1(1,47) =$ 3.35, CI = ±0.17) but not by items ($F_2(1,23) = 2.60$, *n.s.*).

4.3 Is given-new ordering an emergent property of accessibility?

At first glance, these results might seem to contradict an account of given-new ordering as resulting from communicatively-motivated audience design (i.e., from speakers forming their utterances based on a model of their listeners mental states; Ariel, 1988; H. H. Clark & Haviland, 1977; Vande Kopple, 1982). However, audience design might also be affected by WM manipulations – after all, maintaining and updating a model of the listener, including what referents are given and new to that listener, seems a likely candidate to rely on WM processes (but see Sailor, Baird, & Gallerani, 2006, who found no relationship between memory ability and children's marking of given and new information). This idea that givenness is represented in a WM-dependent model of the listener is similar to Clark and Marshall's (1981) proposal that personal common ground relies on explicit memory traces of the co-occurrence of topics with addressees. Similarly, Ariel (1988, 1991) claims that the choice of referring expressions in discourse arises from the accessibility of the referent *for the listener*, presumably based on a similar kind of model of the listener.

This seemingly makes it difficult to determine which account is correct. However, while an account of given-new ordering based on a WM-dependent model of the listener does predict interactions with WM, it also suggests that givenness should interact with WM *independently* of accessibility. Thus, this account would be supported by evidence that givenness and accessibility interact with WM processes differently, or to different extents. In contrast, if given-new ordering effects result from information accessibility and incremental processing, then given-new effects should be sensitive to memory pressures in the same way as are accessibility effects.

To address this issue, an analysis was carried out on combined data from Experiments 1 through 4 treating cue-type (theme / goal) and WM-load as within-subjects factors, and cue-manipulation (accessibility / givenness) as a between subjects factor (i.e., comparing Experiments 1 and 3 to Experiments 2 and 4). If cue-manipulation interacts with WM-load (either in a simple interaction or within a three-way interaction), this would show that WM affects accessibility effects and given-new ordering in different ways. Thus this finding would suggest that while given-new ordering relies on WM, it does so differently than accessibility effects, thereby supporting an account of given-new ordering as based on WM-dependent audience design. In contrast, if cue-manipulation does not interact with WM-load, then this suggests that given-new ordering and accessibility effects rely on WM in the same way. More explicitly, because the accessibility manipulation in Experiments 1 and 3 was manipulated *without* also manipulating givenness (i.e., items were made accessible only for the speaker), speakers' tendency to produce accessible information earlier in these experiments cannot be communicatively motivated. Thus, if the effect of WM on accessibility does not differ from the effect of WM on givenness, this would suggest that speakers use of given-new ordering is also not communicatively motivated, but rather reflects speaker-based accessibility.

Figure 8 shows the proportion of prepositional datives produced as a function of cue-type, WM-load, and whether the cued word was made accessible or given. A logistic mixed-effects analysis revealed a main effect of WM-load ($z = -3.48$, $p < .001$), a marginally significant main effect of cued-word ($z = -1.65$, $p < .10$), and a marginally significant interaction between WM-load and cued-word ($z = 1.72$, $p < .10$).

Surprisingly, given the rather dramatic difference in the size of given-new ordering effects and accessibility effects in these experiments (in the no WM-load conditions, the average effect of accessibility on syntactic choice was 4.7% while the average effect of givenness was 21.0%), the mixed-effects model finds no significant interaction of cued-word with cue-manipulation (accessible/given) nor, in fact, any significant role of cue-manipulation whatsoever. Inspection of the logistic mixed model reveals that the estimate for the cue-type by cue-manipulation interaction term is quite large (in fact, it is larger than any other estimated effect in the model), making the lack of significance even more surprising. Furthermore, the 'traditional' by subjects and by items analyses do show such an effect (see below).

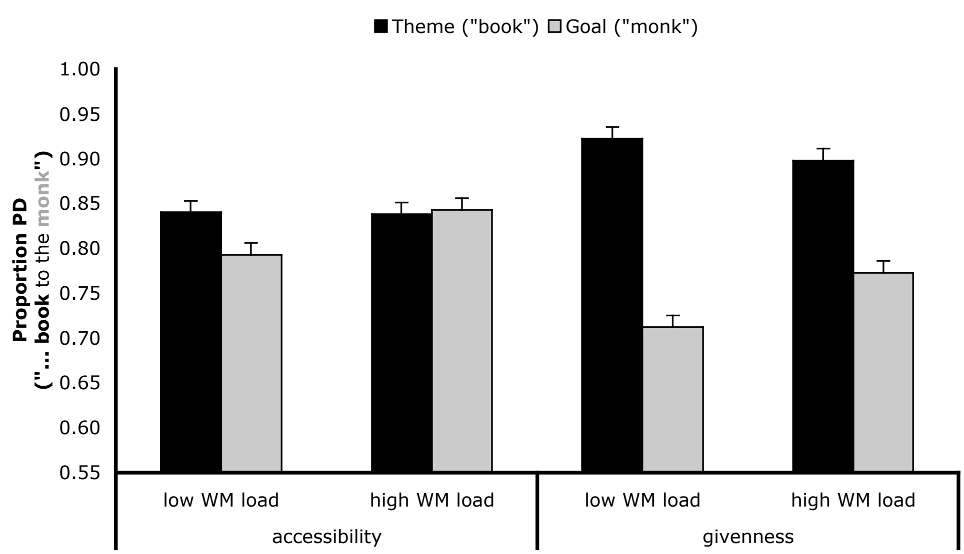

Figure 8: Proportion of prepositional dative (PD) sentences produced in Experiments 1 through 4 as a function of cue-type, WM load, and cue-manipulation (whether cues were made accessible or given). Error bars represent standard errors.

It is likely, especially in light of the contradictory results from the traditional analyses, that this reflects an instance of the *Hauck-Donner effect* (Hauck & Donner, 1977; see also Vaeth, 1985) – a problem with the Wald statistic used to test for significance in the mixed-model. Essentially, the standard error of the estimate can be overestimated as the parameter estimate moves away from zero (i.e., for parameters with particularly large effects) leading to an overly small *z* value (recall that Wald's *z* simply equals the parameter estimate divided by its standard error). One solution to this problem is to use log likelihood ratios, rather than Wald's *z*, to determine significance. As the overall pattern of data suggests, log likelihood ratio test shows that the inclusion of cue-manipulation leads to a significantly better model ($\chi^2(5) = 60.54$, $p < .001$), further suggesting that the lack of any cue-manipulation effect in the model above reflects the Hauck-Donner effect.

Note, however, that this log likelihood test only shows that the cue-manipulation factor matters – it does not show with which, if any, other factors it interacts. Thus, in this instance, it seems that the traditional by-subjects and by-items tests (on arcsine square-root transformed proportions) are preferable. As expected based on the results of the individual experiments, these analyses reveal that speakers tended to produce accessible/given items early, reflected in a main effect of cue-type ($F_1(1,190) = 33.66$, CI = ±0.045; $F_2(1,23) = 32.33$, CI = ±0.30), and tended to produce more prepositional dative structures when under WM-load, shown by a main effect of WM-load significant by subjects ($F_1(1,190) = 5.62$, CI = ±0.081) but not by items ($F_2(1,23) < 1$, *n.s.*). Furthermore, these two factors interacted, showing that speakers were less likely to produce accessible/given information early when under WM-load ($F_1(1,190) = 13.84$, CI = ±0.026; $F_2(1,23) = 10.03$, CI = ±0.20).

Of particular interest here, however, is any effect or interactions with cue-manipulation. Recall that an account where given-new ordering emerges from information accessibility predicts that cue-manipulation should not interact directly with WM load, nor with any interactions including WM load. Such interactions would show that WM load affects given-new ordering *differently* than it affects accessibility, and would therefore support an account where given-new ordering is based on a WM-dependent model of the listener (or model of the discourse), but not based simply on information accessibility.

Inspection of Figure 8 suggests that making a cue given had a greater effect on constituent ordering than did making a cue accessible (a 16.8% effect of cue-givenness vs. a 2.1% effect of cue-accessibility), reflected in a significant interaction of cue-manipulation with cue-type ($F_1(1,190) = 20.26$, MSE = 0.05088, CI = ±0.045; $F_2(1,23) = 16.34$, CI = ±0.26). Crucially, cue-manipulation was involved in no other effects or interactions, suggesting that the relationship between WM and givenness is no different than the relationship between WM and accessibility. Instead, these results are consistent with an account where givenness effectively acts as a very strong manipulation of accessibility.

5 Experiment 5: Contrasting the effects of verbal and spatial WM.

Although the previous experiments consistently found WM load to reduce speakers use of accessibility effects and given-new ordering, it is possible that these results arise not from a load on WM per se, but simply because the task is somehow more complicated or demanding (in a way not specific to WM) on the WM load trials. For example, perhaps speakers relied less on accessibility effects and given-new ordering in the WM load conditions simply because the load trials effectively imposed a sort of a dual task (e.g., Pashler, 1994; see Caplan & Waters, 1999, for a similar criticism applied to other studies).

There is, however, a straightforward way to address this concern. The claim that domain-general WM affects syntactic processing does not imply that all types of WM are involved. WM theorists have pointed to an important distinction between verbal and spatial WM (e.g., Shah & Miyake, 1996) and most work suggesting a role of WM in syntactic processes have, at least implicitly, claimed that language processing involves verbal but not spatial WM (see references in section 1.3 above, although note that there are suggestions that at least some aspects of language processing are inherently spatial, e.g., Chatterjee, Southwood, & Basilico, 1999; Richardson, Spivey, Barsalou, & McRae, 2003).[17] This distinction between verbal and spatial WM processes suggests that external loads on verbal WM should affect accessibility use, but loads on spatial WM should not because verbal and spatial WM are separate subsystems. Frameworks that define WM in terms of processes that act on LTM

[17] This is also likely to be quite different for signed languages (cf. Wilson & Emmorey, 1997, 2003).

representations and that point to WMs susceptibility to interference (e.g., Cowan, 1999, 2005; Lewis, 1996) make the same prediction, although for different reasons. Specifically, these models claim that the same system is used for both phonological and spatial maintenance, however interference occurs between representations that are similar so that a load that is quite different from the materials in a task should not interfere with that task.

5.1 Method.

5.1.1 Participants.

55 students at UCSD participated in exchange for class credit. Data were excluded from seven students, six of whom did not produce analyzable sentences on a majority of trials, and one because of an experimenter error.

5.1.2 Materials and design.

The main part of Experiment 5 used the same experimental design and materials as used in the previous four experiments. In addition, a spatial load and a spatial (letter-rotation) task were created, and thirteen additional verbal load words were selected and digitally recorded at 44.1 kHz by the author.

The spatial WM load was based on Ichikawa's (1981, 1983) dot-in-matrix patterns (see also Miyake et al., 2001). These patterns are five-by-five grids with dots randomly assigned to spaces in the grid. Dot configurations that were duplicates or that formed systematic patterns were eliminated (Ichikawa, 1981; Miyake et al., 2001) and these patterns were presented for 750 ms on a computer screen. This duration was chosen following Miyake et al. (2001), who found 750 ms to be a brief enough presentation to discourage verbal or other idiosyncratic coding strategies (e.g., "two over, three down"), while still allowing enough exposure to encode the pattern into spatial WM. An example is shown in Figure 9.

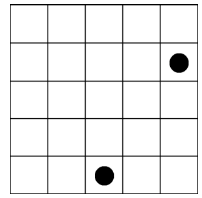

Figure 9: Example dot-in-matrix pattern used to impose a spatial WM load in Experiment 5.

In order to determine the appropriate difficulty of this dot memory task, a norming experiment was conducted with a separate group of ten participants. On each trial of this norming task, participants either saw a verbal (two-word) or a spatial (dot-in-matrix pattern) memory load, then saw a sentence on the computer screen, which they were instructed to say aloud. The sentence disappeared as soon as the participants started to speak (triggered by a head-warn microphone connected to a PsyScope button box), thus participants had to process the sentences sufficiently to produce them from memory. Finally, participants either said aloud the verbal load, or marked the location of the dots on an answer sheet containing several five-by-five matrix grids. The verbal load trials used the same 18 two-word verbal load stimuli used in Experiments 3 and 4, and the spatial load trials were 48 dot-in-matrix patterns containing two, three, four, or five dots (12 patterns of each difficulty). The 66 sentences were mostly intransitive or predicate-adjective structures. Two lists were created that randomly paired sentences with load stimuli, with the constraint that the verbal load stimuli were not paired with sentences to which they shared any obvious semantic or phonological relationship. Trial order was randomly ordered within each list.

On average, participants accurately recalled both words on 47.2% of the verbal-load trials and accurately recalled the dot pattern on 52% of the two-dot trials, 25% of the three-dot trials, 8% of the four-dot trials, and 7% of the five-dot trials. Matched-pairs *t*-tests revealed that performance on the verbal load and on the two-dot spatial load did not differ significantly ($t(9) = 0.53$, n.s.), whereas performance on the verbal load was significantly better than performance on the three-dot ($t(9) = -3.71$, $p < .005$), four-dot ($t(9) = -6.87$, $p < .001$), and five-dot ($t(9) = -5.96$, $p < .005$) spatial load trials. Based on these norming data, which suggest that a two-dot pattern is equivalently difficult to the two-word verbal load, 34 unique two-dot patterns were created (as above) to serve as the spatial load in Experiment 5.

The spatial processing task was a letter rotation task adapted from one used in Shah and Miyake (1996).[18] On each trial, participants saw a series of the same capital letter (F, J, L, P, or R) that was either normal or mirror-imaged, and that was presented in a non-canonical orientation (either rotated 45˚, 90˚, 135˚, 180˚, 225˚, 270˚, or 315˚ from upright). Participants' task was to press the right/green button on the button-box to indicate if the letter was "normal", or the left/red button if the letter was "mirror imaged" as quickly and accurately as possible. There was a practice block of 30 upright letters (three normal and three mirrored versions of each of the six letters, presented in random order), then a second practice block with three trials of two rotated letters per trial. The experimental trials were in four blocks of six trials each, which progressed from two to five letters per trial. Thus there were 24 experimental trials, six of each size.

[18] Thanks to Naomi Friedman for helpful advice regarding this task.

5.1.3 Procedure.

Experiment 5 consisted of three parts. The procedure of the first part was identical to that of Experiment 4 (with the manipulations of givenness and sentence-external WM load), except that the no-load trials were replaced with trials requiring maintenance of a spatial WM load. On each spatial-load trial, participants first pressed a button and saw a five-by-five grid with dots in two of the squares (see Figure 9) for 750 ms on the computer screen. The experimenter then asked a question (e.g., *What's going on with the pirate and the apple?*), to which the speaker responded by describing the relevant picture from the poster on the wall, and then indicated the two dot locations on a five-by-five grid printed on a paper answer sheet. The procedure for the verbal load trials was the same, except that the experimenter read aloud the two unrelated load words, which participants recalled aloud at the end of the trial (just as on the load trials in Experiment 4).

In the second part of Experiment 5, participants judged whether letters at different rotations were normal or mirror imaged while maintaining either a verbal or a spatial WM load (just as in the sentence production task above). A practice block familiarized participants with the normal/mirrored distinction, and a second practice block familiarized participants with the rotation task while under verbal and spatial loads (with one verbal-load, and two spatial-load practice trials). On each trial of the second practice block and of the experimental blocks, participants either saw a two-dot pattern for 750 ms or heard two unrelated words through a speaker next to the computer, then made normal/mirror judgments for two, three, four, or five rotated letters, then recalled aloud the verbal load or marked the spatial load dot locations on an answer grid. See Figure 10 for a schematic of this spatial task.

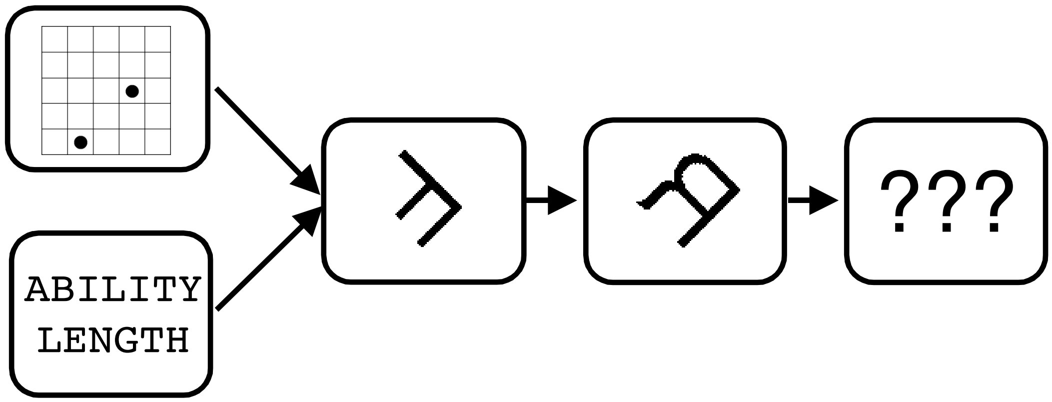

Figure 10: Schematic of the letter rotation task administered as the second part of Experiment 5.

In the third part of Experiment 5, the speaking-span task was administered just as in the four previous experiments.

5.2 Analysis, results and discussion

5.2.1 Sentence production task.

As in the previous four experiments, speakers' sentences from the picture-description task were transcribed and coded as prepositional datives (PD), double-object datives (DO) or other. Trials in which speakers did not produce a dative sentence or did not produce both post-verbal noun phrases were excluded from the analysis, causing 7.9% of all critical trials to be excluded (8.7% of trials in the verbal-load condition and 7.1% of trials in the spatial-load condition). Because the dot-in-matrix load trials were scored as correct when both dots were in the correct grid and incorrect otherwise, the verbal load was scored similarly – as correct only when both words were correctly recalled. Note that this differs from the more lenient criterion used in Experiments 3 and 4, where correctly recalling one of the two verbal load words was considered sufficient recall for inclusion. Trials on which speakers did not correctly remember the WM-load were excluded from the analysis, leading to the exclusion of an additional 34.9% of all critical trials (43.8% of spatial-load trials and 26.0% of verbal load trials). These exclusions again point to the utility of mixed-effect modeling, which is robust to unequal numbers of observations (Baayen, in press).

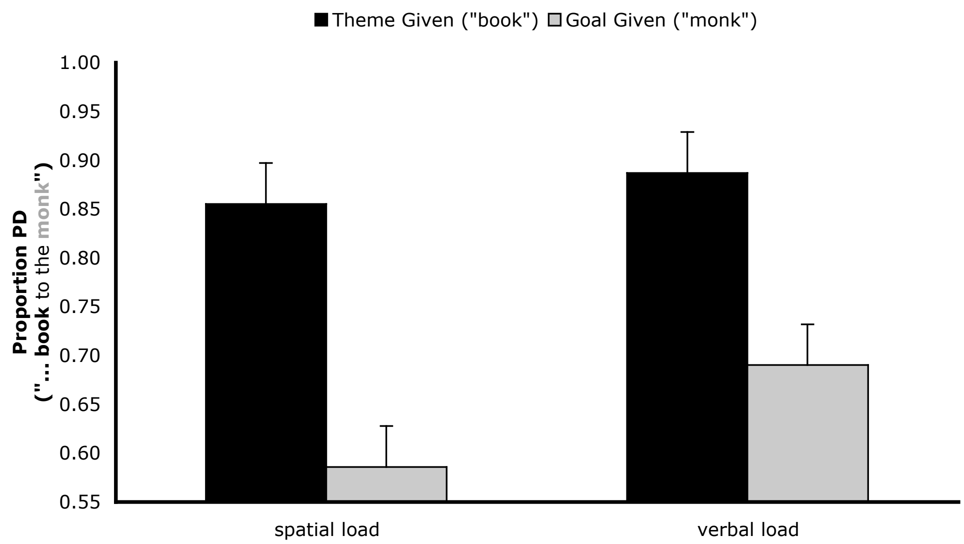

Figure 11: Proportion of prepositional dative (PD) sentences produced in Experiment 5 as a function of cue-type (i.e., which post-verbal argument was made given: theme or goal) and WM load (spatial vs. verbal). Error bars represent standard errors.

Figure 11 shows the proportion of prepositional datives produced as a function of cue-givenness (theme or goal) and WM-load (verbal vs. spatial). Speakers were more likely to produce a PD sentence when under verbal load than when under spatial load, reflected in a significant main effect of WM-load ($z = 2.69$, $p < .01$). There was also a significant effect of cue-givenness, reflecting speakers' tendency to produce given information before new information ($z = 6.60$, $p < .001$). Importantly, these two factors significantly interacted ($z = -2.09$, $p < .05$), reflecting the reduced use of given-new ordering when under verbal load (an 20% effect) compared to when under spatial load (a 27% effect).[19]

5.2.2 Letter rotation task.

Trials in which subjects did not correctly remember the WM load were excluded from the analysis according to the same criteria as in section 5.2.1, above, leading to the exclusion of 33.0% of all trials. Speakers seem to have found the spatial load more difficult than the verbal load as this criteria led to the exclusion of 64.6% of trials in the spatial-load condition and only 5.2% of trials in the verbal load condition. (Note that fewer trials can be excluded by using a more lenient criterion for load recall, which leads to the exclusion of 17.8% of trials – 37.2% of spatial-load trials and 2.8% of verbal load trials – and shows the same numeric and statistical pattern of results). Additionally, reaction times (RT) deviating more than two standard deviations from each subject's mean RT were excluded from the analysis, leading to the exclusion of 4.3% of all trials (2.1% of trials in the spatial load condition and 2.6% of trials in the verbal load condition). Reaction times were analyzed with a linear mixed-effects model treating subject as a random effect and load-type as a fixed effect. Because it is not clear how to determine degrees of freedom for this kind of model (an upper bound is to take the number of observations minus the number of fixed-effect terms, but this is

---

[19] Because traditional by-subjects and by-items analyses are not robust to empty cells (which arose when subjects did not provide any analyzable responses in one or more conditions, usually by not successfully maintaining the WM-load), 5 subjects had to be excluded from the subjects analysis, and one item from the items analysis. After these exclusions, there was a significant main effect of WM-load by subjects ($F_1(1,42) = 5.12$, CI = ±0.13) but not by items ($F_2(1,22) < 1$, *n.s.*), a significant main effect of cue-givenness ($F_1(1,42) = 18.10$, CI = ±0.25; $F_2(1,22) = 26.67$, CI = ±0.38), and an interaction between these factors that was not significant by subjects ($F_1(1,42) < 1$, *n.s.*) but significant by items ($F_2(1,22) = 4.62$, CI = ±0.27).

probably insufficiently conservative), *p* values were determined with Markov chain Monte Carlo (MCMC) sampling (Baayen, in press).

Figure 12 shows normal/mirror judgment response latencies as a function of type of WM load. Although the accuracy of letter rotation judgments did not differ as a function of load, participants were an average of 71 ms slower to make letter rotation judgments when under a spatial WM load than when under a verbal WM load, as shown by a significant effect of load-type ($t = -2.81$, $p < .01$).[20]

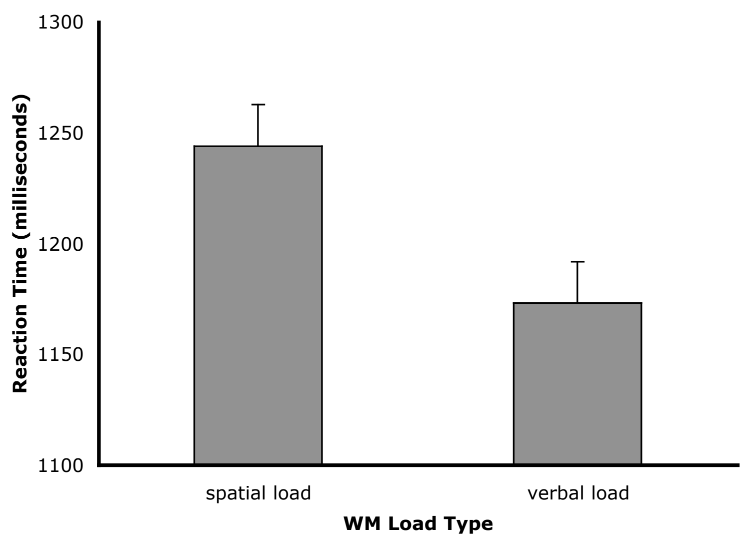

Figure 12: Average response latencies to categorize rotated letters as normal or mirror imaged in Experiment 5 as a function of WM load type (spatial or verbal). Error bars represent standard errors.

Experiment 5 provides further evidence that given-new ordering depends on verbal WM because a concurrent load on verbal WM led to decreased use of given-new ordering whereas a load on spatial WM did not. This shows that the effect of a verbal WM load on syntactic production found here and in the previous four experiments is not

---

[20] A traditional analysis on RT data averaged over subjects also shows a significant effect of load-type ($F(1,45) = 7.02$, CI = ±53.5 ms); note that two subjects were excluded due to missing values.

attributable to the dual-task nature of the load trials, nor to the overall increased difficulty in the WM load trials. If anything, the spatial WM load seems to have been more difficult overall, as evidenced by how often speakers were unable to remember the spatial load relative to the verbal load. Furthermore, the exact opposite pattern was found in a letter-rotation task that requires spatial processing, namely that the spatial WM load led to worse performance than the verbal WM load. This both shows that the spatial load was, in principle, difficult enough to lead to influence processing, and also adds further support to the distinction between processes of verbal and spatial WM processes.

6 General discussion.

These experiments make two general points. First, both accessibility effects and given-new ordering rely on WM mechanisms. Experiments 1 and 3 show that WM load reduces speakers' tendency to produce accessible information early, and this effect further interacted with individual differences in WM span in Experiment 1. This finding confirms what has been suggested before – that accessibility effects reflect a pressure to avoid buffering otherwise ready-to-produce material (cf. Bock, 1982; Ferreira, 1996). Experiments 2, 4, and 5 show that maintenance of a WM load also reduces speakers' tendency to use given-new ordering. This is somewhat surprising because the concept of givenness has primarily been construed as reflecting givenness *for the listener*, and therefore given-new ordering has been considered to be a communicatively based effect (e.g., Ariel, 1988, 1991; Chafe, 1976; H. H. Clark & Haviland, 1977; Collins, 1995; Halliday, 1970; Levelt, 1989; Thompson, 1995). The experiments presented here suggest that given-new ordering instead reflects cognitive pressures on speech production, especially as both givenness and accessibility are affected by WM load in the same way (as shown in the combined analysis of Experiments 1 through 4). This likely reflects the fact that speaking is a difficult process, and so many of speakers' tendencies can be ascribed to cognitive pressures on the speaker rather than to communicative pressures (cf. Bock, 1982). It is important to recognize, however, that this conclusion does not imply that given-new ordering does not serve an important communicative function. So long as speakers reliably produce given information early, which they are likely to do because given information is generally also relatively more accessible than new information, then listeners can rely on constituent order to reflect givenness and newness.

The second general finding of these experiments is that speakers' choices of syntactic structures are influenced by demands on domain-general verbal WM. Speakers relied less on accessibility effects and on given-new ordering when under both sentence-internal and sentence-external WM loads, and the equivalent effects of these types of WM load was supported by a combined analysis of Experiments 1 through 4 that showed no difference in the effect of internal and external WM load on accessibility effects and given-new ordering.[21] The most straightforward (and possibly also the most controversial) implication of this finding is that syntactic production does not rely only on a dedicated pool of memory resources (as has been claimed to be the case for syntactic comprehension, e.g., Caplan & Waters, 1999), but rather relies on a more general system of WM. This further suggests that syntactic production is not an informationally encapsulated process (in the sense of Fodor, 1983) because syntactic choice is influenced by non-syntactic pressure from buffering of accessible information in domain-general verbal WM.

Many demonstrations of domain-general WM effects on syntactic comprehension have not controlled for the possibility that WM load effects on parsing

---

[21] Specifically, in an analysis treating load-type as a between subjects factor, load-type did not interact with cue-type nor with the cue-type by WM-load interaction. In fact, the only significant role of load-type was an interaction with WM-load ($F_1(1,190) = 9.70$, CI = ±.08, $F_2(1,23) = 2.87$, *n.s.*). This somewhat unexpected result shows that speakers' tendency to produce more prepositional dative (PD) structures while under a WM load was greater in the experiments manipulating WM load sentence-internally. Thus it seems that the main effect of WM-load discussed in Experiment 1 may not result from a reversion to the grammatical default when under cognitive load, but rather reflects a preference to use PDs for sentences with longer subject NPs.

reflect other sources of difficulty (i.e., perhaps the load condition imposes a sort of dual-task that accounts for the effects purported to be WM-related; see Caplan & Waters, 1999, for criticisms of this sort). Experiment 5 included such a control by imposing either a verbal WM or a spatial WM load, and found that while the verbal WM load led to reduced use of given-new ordering, the spatial WM load did not. In fact, the given-new ordering effects in the spatial WM load condition were numerically the largest effects across all five experiments. This observation, combined with the difficulty speakers had with maintaining the spatial WM load, suggest that increasing general cognitive difficulty may lead to *increased* reliance on given-new ordering. This pattern follows nicely from the idea that accessibility effects arise as a way to reduce demand on cognitive systems (Bock, 1982) combined with the idea that WM is particularly sensitive to similarity (e.g., Cowan, 1999; Gordon et al., 2002; Lewis & Vasishth, 2005). Specifically, increasing cognitive pressure in a general sense (here, with the spatial WM load) leads to *more* reliance on accessibility effects to help mitigate that pressure. However, because the verbal WM load is relatively similar to the to-be-produced material, the accessible information is subject to interference from the verbal WM load.

This invokes a somewhat different concept of WM than the traditional multiple-components model. Instead of conceptualizing WM as a workspace into which representations from long-term memory (LTM) are transferred and held (e.g., Baddeley, 1986; Jackendoff, 2002), this account fits better with frameworks that consider WM as a set of activated LTM representations, and suggest that WM capacity is determined not by some fixed amount of 'workspace,' but is rather based on similarity-based interference (e.g., Cowan, 1999, 2005; Cowan et al., 2005; Oberauer & Kliegl, 2006; Unsworth & Engle, 2007). Recently, this kind of approach has been fruitfully applied to models of WM and language comprehension (Fedorenko et al., 2006; Gordon et al., 2002; Lewis & Vasishth, 2005), suggesting that this kind of approach can profitably be applied toward understanding the role of WM in syntactic production as well.

The discussion thus far has implicitly assumed that the WM load interfered with the *maintenance* of the accessible item, that is, the WM load presumably makes it difficult to keep information sufficiently active to warrant early mention. However, an equally plausible alternative is that the WM load interferes with the *encoding* of the accessible item. By this account, the WM load causes the accessible (or given) information to be less thoroughly processed, causing that information to sometimes be insufficiently active to warrant early mention. The data presented here cannot distinguish between these accounts because, in these experiments, the WM load always preceded the presentation of the accessible / given information, thus I leave this question for future research. It is, however, important to note that under either account, the strength of a domain-general memory representation influences speakers' syntactic production processes. Therefore this distinction does not detract from the main conclusions that accessibility effects and given-new ordering result from domain-general WM processes.

6.1 Conclusions and future directions.

One of the fundamental difficulties with producing spoken language is that non-linear conceptual information must be expressed as a linear order of words. Because of this, speakers often must buffer information that is otherwise ready to be produced. The experiments reported here show that this buffering process employs domain-general WM mechanisms. Speakers help reduce these demands on WM by relying on the flexibility afforded by the grammar and choosing syntactic structures that allow more accessible information to be produced early. This tendency underlies not only accessibility effects per se, but also seem to account for speakers' (usually pragmatically-motivated) tendency to produce given information before new.

As with most research, these conclusions lead to a number of other questions. One concerns the pragmatic nature of givenness – do all forms of givenness reflect speaker-centric processes? Some of the ways in which speakers mark givenness could plausibly reduce to accessibility, for example the tendency to produce given information with less clarity (Bard et al., 2000; Fowler & Housum, 1987), weaker stress (Chafe, 1976; Halliday, 1967), and with less specific terms (Ariel, 1988, 1991; Gundel et al., 1993). However, it is less clear how, for example, the tendency to refer to given information with definite articles (i.e., *the* in English; Grieve, 1973) would ease processing in any substantive way. So while it seems very unlikely that there is no pragmatic dimension of givenness from speakers' perspectives, the results presented here suggest that future research might profitably investigate what other kinds of pragmatically and communicatively motivated effects might instead be ascribed to speakers' strategies to reduce cognitive load. Of course, there is a balance between speaker-centric and listener based processes (if language were all speaker-centric, then nobody would talk at all), but many things that are seemingly communicatively based might instead be driven by the processing demands of production. A related example is the finding that speakers use optional words and prosody not to avoid ambiguity for their listeners, but rather as a way to ease their own processing (e.g., Ferreira & Dell, 2000; Kraljic & Brennan, 2005).

The other major implication of the data presented here is that syntactic production relies on domain general verbal WM processes. Recent evidence on the role of WM in syntactic comprehension suggests that the crucial factor is not the amount of material in the WM load, but rather the similarity of the WM load to the material in the to-be-produced sentence (Fedorenko et al., 2006; Gordon et al., 2002). Based on these findings, an obvious extension of the work presented here is to look at the influence of similarity in syntactic production. Simply by manipulating the similarity of the WM load to the target sentence in the paradigm used in these experiments, it should be possible to determine if the verbal WM load imposed in these experiments really *interferes* with the representation of the accessible/given information. Not only could this line of research be informative regarding the nature of the WM processes involved in sentence production, but it also would likely inform the nature of WM more generally.

Indeed, investigations like those presented here have implications not just for accounts of language processing, but also for models of WM. Investigations of WM processes often rely on language production with the implicit assumption that production is not influenced by the topic of investigation. Bock (1996) calls this the *mind in the mouth assumption* (i.e., the assumption that the mental representation of investigation is transparently reflected in the verbal response). Fortunately, this assumption is becoming increasingly uncommon, and increasing interest and energy is focusing on the role that linguistic representations play in memory tasks (e.g., Howard

& Nickels, 2005; Jacquemot & Scott, 2006; Martin & Freedman, 2001). Based on the data presented here, it seems likely that syntactic processes may also play a role in WM processing (for some evidence that this might be the case, see Schweppe & Raummer, 2007).

7 References

Aboitiz, F., García, R., Brunetti, E., & Bosman, C. (2006). The origin of Broca's area and its connections from an ancestral working memory network. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's Region* (pp. 3-16). New York: Oxford University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics, 24*, 65-87.

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics, 16*, 443-463.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heavyness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*, 28-55.

Baars, B. J. (Ed.). (1992). *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum.

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers, 1*, 1-45.

Baayen, R. H. (in press). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge University Press.

Baayen, R. H., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language. Special Issue: Mental lexicon II, 81*, 55-65.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY: Cambridge University Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586-595.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory & Language, 56*, 65-85.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory & Language, 42*, 1-22.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News, 5*, 27-30.

Bayles, K. A. (2003). Effects of working memory deficits on the communicative functioning of Alzheimer's dementia patients. *Journal of Communication Disorders, 36*, 209-219.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory* (Vol. 1, pp. 35-67). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bock, J. K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning & Verbal Behavior, 16*, 723-734.

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

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

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

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

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395-421.

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

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

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

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

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

Brainerd, C. J. (1995). Interference processes in memory development: The case of cognitive triage. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 105-139). San Diego, CA: Academic Press.

Branigan, H. P., McLean, J. F., & Reeve, H. M. (2003). *Something old, something new: Addressee knowledge and the given-new contract.* Paper presented at the Proceedings of the 25th annual meeting of the Cognitive Science Society.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation.* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*, 441-472.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied Psycholinguistics, 22*, 253-267.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77-126.

Caplan, D., & Waters, G. S. (2002). Working memory and connectionist models of parsing: A reply to MacDonald and Christiansen (2002). *Psychological Review, 109*, 66-74.

Caplan, D., & Waters, G. S. (2003). On-line syntactic processing in aphasia: Studies with auditory moving window presentation. *Brain and Language, 84*, 222-249.

Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology, 46*, 91-121.

Caspari, I., & Parkinson, S. R. (2000). Effects of memory impairment on discourse. *Journal of Neurolinguistics, 13*, 15-36.

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition, 37*, 205-223.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.), *Subject and Topic* (pp. 25-55). New York: Academic Press.

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia, 37*, 395-402.

Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory & Language, 56*, 172-188.

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language, 54*, 324-345.

Clark, E. V., & Clark, H. H. (1978). Universals, relativity, and language processing. In J. Greenberg (Ed.), *Universals of Human Language, Volume 1: Method and Theory* (pp. 225-277). Stanford: Stanford University Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. Freedle (Ed.), *Discourse production and comprehension*. New York, NY: Ablex.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clifton, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition, 32*, 886-895.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers, 25*, 257-271.

Collins, P. (1995). The indirect object construction in English: An informational approach. *Linguistics, 33*, 35-49.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 226-234.

Conwell, E., & Demuth, K. (2007). Early syntactic productivity: Evidence from dative shift. *Cognition, 103*, 163-179.

Cooper, W. E., & Ross, J. R. (1975). World order. In R. E. Grossman, L. J. San & T. J. Vance (Eds.), *Papers from the parasession on functionalism* (pp. 63-111). Chicago: Chicago Linguistic Society.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York, NY: Cambridge University Press.

Cowan, N. (2005). *Working Memory Capacity*. New York: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42-100.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*, 445-464.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450-466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory & Language, 25*, 1-18.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422-433.

Davidson, D. J., Zacks, R. T., & Ferreira, F. (2003). Age preservation of the syntactic processor in production. *Journal of Psycholinguistic Research, 32*, 541-566.

DeDe, G., Caplan, D., Kemtes, K., & Waters, G. S. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging, 19*, 601-616.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review, 108*, 759-788.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology. Special Issue: Working Memory, 49A*, 234-250.

Faraway, J. J. (2006). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Boca Raton: Chapman & Hall/CRC.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject-verb agreement errors: A study in French written language. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 47A*, 437-464.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory & Language, 54*, 541-553.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory & Language, 56*, 246-269.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition: A Journal of Developmental Linguistics, 11*, 127-163.

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

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

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

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377-380.

Fodor, J. A. (1983). *Modularity of mind: an essay on faculty psychology*. Cambridge, Mass.: MIT Press.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language, 26*, 489-504.

Frazier, L., & Fodor, J. D. (1987). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General, 129*, 61-83.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*, 172-179.

Garrod, S. C., & Sanford, A. J. (1982). The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics, 1*, 21-41.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65-77.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887-898.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425-430.

Grieve, R. (1973). Definiteness in discourse. *Language and Speech, 16*, 365-372.

Griffin, Z. M. (2003). A revised word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10*, 603-609.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science, 29*, 261-290.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory & Language, 46*, 267-295.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language, 65*, 203-257.

Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. A. Moravcsik & J. R. Wirth (Eds.), *Studies in syntactic typology* (pp. 209-239). Philadelphia, PA: J. Benjamins Pub. Co.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274-307.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: II. *Journal of Linguistics, 3*, 199-244.

Halliday, M. A. K. (1970). Language structure and language function. In J. Lyons (Ed.), *New Horizons in Linguistics* (pp. 140-165). New York: Penguin Books.

Halliday, M. A. K. (1994). The construction of knowledge and value in the grammar of scientific discourse: With reference to Charles Darwin's 'The Origin of Species'. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 136-156). London: Routledge.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: the case of subject-verb agreement. *Language and Cognitive Processes, 21,* 181-204.

Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in Logit analysis. *Journal of the American Statistical Association, 72*, 851-853.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior, 13*, 512-521.

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1-35.

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition, 96*, 127-142.

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

Howard, D., & Nickels, L. (2005). Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22*, 42-77.

Ichikawa, S.-i. (1981). Rated ease of memorization of dot-in-matrix patterns: Multiple regression analysis by physical variables which describe configuration of dots. *Japanese Psychological Research, 23*, 69-78.

Ichikawa, S.-i. (1983). Verbal memory span, visual memory span, and their correlations with cognitive tasks. *Japanese Psychological Research, 25*, 173-180.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, Mass.: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10*, 480-486.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*, 113-147.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.

Kellogg, R. T. (2004). Working memory components in written sentence generation. *American Journal of Psychology, 117*, 341-361.

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

Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology and Aging, 18*, 181-192.

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16*, 312-322.

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24*, 253-270.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*, 46-50.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory & Language, 30*, 580-602.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

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

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

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.

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

Lewis, R. L. (1996). Interference in short-term memory: the magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*, 93-115.

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science. Special Issue: 2004 Rumelhart Prize Special Issue Honoring John R. Anderson: Theoretical Advances and Applications of Unified Computational Models, 29*, 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 447-454.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*, 550-557.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale: Lawrence Erlbaum.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35-54.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56-98.

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103-125.

Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory, 9*, 261-280.

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology, 21*, 625-644.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology, 8*, 506-523.

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory & Language, 33*, 83-111.

Mazurkewich, I. (1984). The acquisition of the dative alternation by second language learners and linguistic theory. *Language Learning, 34*, 91-109.

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

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2, pp. 269-321). New York: Wiley.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11*, 671-717.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*, 621-640.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory & Language, 55*, 601-626.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin, 116*, 220-244.

Pavio, A., Yuille, J. C., & Madigan, S. (1968). Concreteness, imagry, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*, 1-25.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory & Language, 34*, 521-542.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning & Verbal Behavior, 15*, 33-42.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow & Co, Inc.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-255). New York: Academic Press.

R Development Core Team. (2006). R: A language and environment for statistical computing (Version 2.4.1): R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Raaijmakers, J. G. W. (2003). A further look at the "language-as-fixed-effect fallacy". *Canadian Journal of Experimental Psychology, 57*, 141-151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416-426.

Rawson, K. A. (2007). Testing the shared resource assumption in theories of text processing. *Cognitive Psychology, 54*, 155-183.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767-780.

Roßnagel, C. (2000). Cognitive load and perspective-taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology, 30*, 429-445.

Roßnagel, C. (2004). Lost in Thought: Cognitive Load and the Processing of Addressees' Feedback in Verbal Communication. *Experimental Psychology, 51*, 191-200.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437-442.

Sailor, M. M., Baird, J. A., & Gallerani, C. (2006). Telling others what's new: Preschoolers' adherence to the given-new contract. *Journal of Cognition and Development, 7*, 341-379.

Schweppe, J., & Raummer, R. (2007). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory & Language, 56*, 336-356.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21-50.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*, 675-681.

Smith, E. E., & Geva, A. (2000). Verbal working memory and its connections to language processing. In Y. Grodzinsky, L. Shapiro & D. Swinney (Eds.), *Language and the Brain: Representation and Processing* (pp. 123-141). San Diego, CA: Academic Press.

Smyth, R. H., Prideaux, G. D., & Hogan, J. T. (1979). The effect of context on dative position. *Lingua, 47*, 27-42.

Snyder, W., & Stromswold, K. (1997). The structure and acquisition of the English dative constructions. *Linguistic Inquiry, 28*, 281-317.

Swets, B., Desmit, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*, 64-81.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Thompson, S. A. (1990). Information flow and dative shift in English discourse. In J. A. Edmondson, C. Feagin & P. Mühlhäusler (Eds.), *Development and Diversity: Language Variation across Time and Space* (pp. 239-253). Dallas: Summer Institute of Linguistics.

Thompson, S. A. (1995). The iconicity of "dative shift" in English: Considerations from information flow in discourse. In M. E. Landsberg (Ed.), *Syntactic Iconicity and Linguistic Freezes: The Human Dimension* (pp. 155-175). New York: Mouton de Gruyter.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204-224.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381-391.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132.

Vaeth, M. (1985). On the use of Wald's test in exponential families. *International Statistical Review, 53*, 199-214.

Vande Kopple, W. J. (1982). The given-new strategy of comprehension and some natural expository paragraphs. *Journal of Psycholinguistic Research, 11*, 501-520.

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! *Psychological Science, 17*, 273-277.

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research, 26*, 347-361.

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16*, 128-144.

Waters, G. S., & Caplan, D. (2002). Working memory and online syntactic processing in Alzheimer's disease: Studies with auditory moving window presentation. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 57B*, 298-P311.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57A*, 129-163.

Waters, G. S., & Caplan, D. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory, 13*, 403-413.

Waters, G. S., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review, 10*, 88-95.

Weckerly, J., & Elman, J. L. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Earlbaum.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183-194.

Wilson, M., & Emmorey, K. (1997). A visuospatial "phonological loop" in working memory: Evidence from American Sign Language. *Memory & Cognition, 25*, 313-320.

Wilson, M., & Emmorey, K. (2003). The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education, 8*, 97-103.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York, NY: McGraw-Hill.

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 523-538.

Yampolsky, S., Waters, G. S., Caplan, D., Matthies, M., & Chiu, P. (2002). Effects of acoustic degradtion on syntactic processing: Implications for the nature of the resource system used in language processing. *Brain and Cognition, 48*, 617-625.

Zhong, H., Stent, A., & Swift, M. (2006). Modeling the dative alternation with automatically extracted features. In *Proceedings of the Workshop on Statistical and Empirical Approaches for Spoken Dialog Systems*: American Association for Artificial Intelligence (AAAI).

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.