Choice of Referring Expression Subject to Visual Context, Linguistic Context and Competition

Jessica L. Montag (montag@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Maryellen C. MacDonald (mcmacdonald@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

Language production studies tend to investigate central tendencies, though there are extensive individual differences in any production task. Studying this variability in behavior can be revealing of the motivations and consequences of production choices. In a picture-description task, participants used either active object or passive relative clauses to describe animate or inanimate entities in pictures. Variability in productions was dependent on the animacy of the target noun, such that animate entities were described with almost exclusively passive relative clauses while inanimate entities were described with a combination of actives and passives. Further, visual properties of the scenes affected production choices such that less salient inanimate entities tended to be described with passive relative clauses. There were also substantial individual differences in structure choices for inanimate target nouns. Speakers with more variability in syntactic structure of their utterances exhibited markers of increased production difficulty in their inanimate vs. animate trails. Implications for theories of grammatical encoding are discussed.

Keywords: Sentence Production; Picture Description; Relative Clauses; Visual Salience; Competition.

Introduction

Language production, as with many other sub-fields of psychology and cognitive science, has typically investigated central tendency as a means to understand certain behaviors or phenomena. This strategy has been highly successful, and we have come to know a great deal about the sentence production process: what factors affect production choices, and why these factors assert the influence they do. Another approach that may be equally informative is to investigate the variability in individual performance—essentially central tendencies at smaller grain sizes, across both items and contexts as well as across individuals. An understanding of this variability may give us additional insight into the underlying mechanisms behind language production processes.

Montag & MacDonald (in press) investigated the role of a joint visual and linguistic context on the form of referring expression, in contrast to some previous picture description studies that focused on the role of visual properties on production choices, with no explicit linguistic context (Gleitman, January, Nappa & Trueswell, 2007; Tomlin, 1995; 1997). It is reasonable to hypothesize that different linguistic contexts, which provide a specific task goal (such as answering a question) may interact with properties of the visual scene. Previous work has shown that that visual gaze patterns in a language production task are task-specific (Kuchinsky, Bock & Irwin, 2011), and so effects of visual properties, such as visual salience of elements to be referred to in a scene, may be task-specific as well. Indeed, Montag and MacDonald showed that in the context of a spoken question that guides visual search of a scene and sets up about a communicative goal for the utterance, visual properties of the scenes affected production choices for referring expressions. In that study, adult native English speaking participants viewed scenes and answered spoken questions such as "Who is wearing white?" or "What is red?" about animate and inanimate entities being acted upon in the scene. The pictures depicted several people and objects, so speakers' descriptions of these target entities typically contained object or passive relative clauses, such as those in Table 1.

Table 1: Sample object and passive relative utterances

Object Relative Clauses
Animate: The man (that/who) the girl is hugging
Inanimate: The toy (that) the girl is hugging
Passive Relative Clauses
Animate: The man (that/who is) being hugged (by the girl)
Inanimate: The toy (that is) being hugged (by the girl)

The animacy of the target entity affected production choices, such that participants produced almost exclusively passive relative clauses when describing animate targets and mix of object and passive relatives when describing inanimate targets. The key finding was that for inanimate entities, the visual salience of the target entity affected utterance choices. We found that less visual salient entities were more often described with passive relatives than more salient entities. This effect may seem counter-intuitive, given previous results that suggest that more salient entities tend to take more prominent positions in sentences (Gleitman, January, Nappa & Trueswell, 2007; Tomlin, 1995; 1997), because the less-salient inanimate entities were

described with passive relatives, patterning with the inherently-salient animate entities. We attributed this effect to the to conjunction of visual salience and the linguistic contex (answering a question) in which the utterance is produced. Less salient scene elements require longer visual search, as evidenced through longer initiation latencies. This increased search time caused the speaker to notice or attend more to the non-target competitors in the picture. We argued that speakers' additional focus on competitor objects, in conjunction with the linguistic context asking a question about a particular entity, encouraged the speaker to be more explicitly contrastive—to distinguish the target item from the other similar item in the picture, rather than to just describe one of many items in a scene. This suggests that one means by which visual salience can affect production choices is by subtly changing the desired message of the utterance. Thus, the role of visual scene properties in production choices is not uniform; it can depend on the linguistic context in which the utterance is formed. Thus we should not expect visual properties to play the same role in every context or experimental design.

Montag and MacDonald (in press) reported only central tendencies in their analysis of speakers' referential forms, however, there were substantial individual differences in the nature of the of visual and linguistic contexts. These individual differences could be informative about how producers settle on a referential form. Here we investigate individual variability in production choices and production difficulty, the greater production difficulty that may be associated with some choices, in order to better understand the mechanisms that underlie sentence production.

In the case of choices for words to convey a message, it appears that alternative forms (e.g., *sofa*, *couch*) compete for activation (Abdel Rahman & Melinger, 2009; Levelt, 1999; Peterson & Savoy, 1998; cf., Mahon, Costa, Peterson, Vargas & Caramazza, 2007), with the competition winner entering the utterance plan. In most cases, evidence for competition comes from measures of production difficulty, typically the latency to initiate speech, which is longer in cases of competition between alternatives relative to the situation in which there is a single dominant response (Dell & O'Seaghdha, 1991; Peterson & Savoy, 1998). Some researchers have suggested that production choices at the syntactic level also involve competition between alternative forms (Cook, Jaeger & Tanenhaus, 2009; Haskell & MacDonald, 2003; Stallings, MacDonald, & O'Seaghdha, 1998). This approach is more controversial than in the lexical competition case because sentence structures are often thought not to be stored and activated but are instead generated from syntactic rules (e.g., Jackendoff, 2002; Pinker, 1984), so that there are not alternatives available to compete for activation. Moreover, there is extensive evidence that sentence structures develop incrementally during production planning as easily accessible words and phrases are placed earlier in an utterance plan (Bock, 1986; 1987; Bock & Irwin, 1980; Prat-Sala & Branigan, 2000; F. Ferreira, 1994; McDonald, Bock & Kelly, 1993). On this view, the structure is emergent from other choices of words and word orders, and not a choice itself. Consistent with this incremental structure-building approach, several studies have suggested that the availability of several structural options, rather than being a source of competition, leads to increased fluency (V. Ferreira, 1996; van Gompel, Pickering, Pearson & Liversedge, 2005).

A common method to evaluate claims for competition is to compare production in a situation in which there are two roughly equally viable utterances (e.g. *sofa*, *couch*), versus one in which a single response is dominant. Longer initiation latencies in the multiple-option than single-option conditions have been interpreted as evidence for competitive processes at work in the multiple-option condition (Haskell & MacDonald, 2003; Spalek & Schriefers, 2005; Stallings et al., 1998). The absence of cost, or even an advantage to multiple options, has been interpreted as suggesting that there is no competition between alternative structures, and syntactic alternatives are constructed incrementally (V. Ferreira, 1996). The goal of the present study is to further test claims for and against competition in sentence production and thereby increase our understanding of sentence production processes. To accomplish this, we perform additional analyses on the picture description data reported in Montag and MacDonald (in press), which provide an ideal arena to test these competition hypotheses, in that one condition yielded almost no variability in utterance form (thus presumably no competition) while a second condition produced substantial variability in production choices across individuals, potentially indicating competition between alternative referential forms. We can therefore look at the consequences of the availability of production alternatives on production difficulty.

Method

Participants

Sixty-eight undergraduates at the University of Wisconsin-Madison participated in exchange for pay or for extra credit in an introductory psychology course. All were native speakers of American English.

Materials

Twenty verbs that can each take both an animate and inanimate grammatical object were selected. A color cartoon picture was created for each verb or adapted from pictures used by Gennari et al. (2012). Each picture contained two depictions of events named by the verb, one acting upon an animate grammatical object and once acting upon an inanimate grammatical object. For example, the pictures for the verbs *throw* and *hug* are shown in Figure 1. Each picture shows an animate entity—in these examples, a man—as the object of the action, and also an inanimate direct object—a ball in the case of *throw* and a toy for the *hug* picture. The animate and inanimate objects of the action were the target items in the experiment. Each picture also contained other elements, always including one or more

additional inanimate and animate elements matching the target items. Thus the *throw* picture in Figure 1 includes a second ball and several other men in addition to the ball/man being thrown, and the *hug* picture contains another man and another toy in addition to the ball/toy being hugged. These extra elements increased the specificity of speakers' descriptions of the target elements in order to distinguish them from other similar elements in the pictures. The two examples in Figure 1 also exhibit the variability in visual salience amount the inanimate target items. In the "throw" picture, the target ball is small and in the corner, while in the "hug" picture, the target toy is large and in the foreground of the scene.

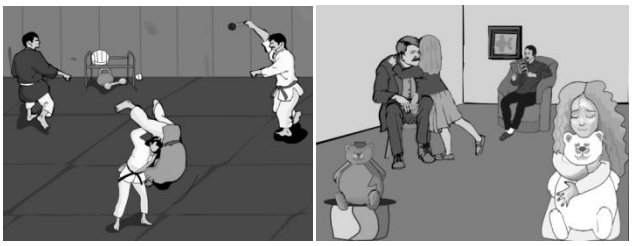

Figure 1: Test pictures for the verbs "throw" and "hug." Experiment participants saw the pictures in color.

To elicit speech, spoken questions were recorded for each picture, and the participants' task was to answer the question presented with the picture. Questions for experimental trials asked participants to describe a particular target person or object in the picture. For example, questions corresponding to Figure 2 would be "Who is wearing orange?" to elicit a description of the animate 'man' target wearing an orange jacket and "What is red?" for the inanimate 'ball' target being thrown by a man. There are multiple men and balls in the picture, so participants had to further describe the target item in order to identify it.

Forty-three filler pictures were included to reduce strategic effects and structural priming (the repetition of utterance sentence structure from one trial to the next; Bock, 1986). For filler trials, participants were asked to describe what a particular person was doing or identify a particular object; these items were designed to elicit simple sentences without relative clauses.

Procedure

In an initial pre-training task, participants were exposed to segments of the test pictures that depicted only the action to be described in the main task. After two seconds, a verb describing that action appeared and the participant was instructed to read that word aloud. This pre-training was done to encourage uniform verb usage across participants (e.g., "hold" vs. "carry"). Participants viewed, in random order, the two target-action picture segments for experimental pictures (one showing the action on an animate entity and the other showing the inanimate) and one picture segment for filler pictures. Fillers were included so that all pictures in the main experiment would have had some pre-exposure.

After the pre-training, participants performed the main experiment, using a variant of the task developed by Gennari et al. (2012). Participants were told that they would view pictures and answer questions about them. In a cover story, participants were told that their responses (picture descriptions) would be shown to a later group of participants, who would guess which picture entities they were describing. They were told that superficial changes would be made in the pictures when they were shown to the new participants, and so to be clear in their descriptions, participants should describe the actions in which the pictured people and objects were taking part. This cover task, in addition to the distracter items (e.g., the non-target ball) elicited a high rate of relative clause productions without explicit instruction or examples.

At the beginning of each trial, a picture appeared on the screen and remained for the duration of the trial. Three seconds after the picture appeared, participants heard a recorded question asking about a target person or object in the scene, such as "What is red?" Participants answered by speaking into a microphone; initiation latencies and all responses were digitally recorded for later analysis.

Animate and inanimate targets for experimental items were counterbalanced across participants so that each participant saw each picture only once and received 10 trials with a question about an animate patient and 10 trials with an inanimate theme question. Test and filler trials were pseudo-randomized such that there were always at least two fillers between any two test trials.

Visual Salience Measures. Three different measures were used to assess the visual salience of the inanimate target items (animate items were found to be all highly salient). First, a new group of participants provided an explicit rating (1-7) of the visual salience of the inamate targets. A second group provided an explicit rating of the animate agents acting on the inanimate targets, on the view that the visual salience of these agents could affect time to find and identify the inanimate targets. A third group of participants provided data for an implict measure, the latency to locate and name the inanimate targets when viewing the scene. A composite of these three measures was used in visual salience analyses below. See Montag and MacDonald (in press) for details of these measures.

Results

In Montag and MacDonald (in press), we found that participants produced almost exclusively passive response when describing animate targets (98.7% passive, SD = 4.1) and a mix of active and passive responses when describing inanimate targets (47.3% passive, SD = 38.6). In addition, we replicated a number of well-established measures of production difficulty, including initiation latencies, which are relevant to the present study. We found longer initiation latencies (F. Ferreira, 1991; V. Ferreira, 1996) to describe

the less-accessible inanimate versus animate (McDonald, Bock & Kelley, 1993; F. Ferreira, 1994) targets as well as the less-salient inanimate items as measured by composite salience measure. Thus, we attained a measure of production difficulty that may be used to investigate the effect of multiple production alternatives (competition) on production difficulty.

With one condition (animates) yielding uniform structure agreement and the other (inanimates) producing variability in structure, as well as a valid measure of production difficulty, we obtained conditions conducive to testing hypotheses for competition during sentence production, that multiple-structure situations will yield higher production difficulty than dominant option ones.

A positive correlation between the measure of production difficulty (initiation latency) and variability of a speaker's responses would be consistent with prior studies arguing for competition between alternative forms (Haskell & MacDonald, 2003; Stallings et al. 1998). While the central tendency was that participants produced approximately equal number of active and passive relative utterances when describing inanimate targets, there was enormous variability among participants. Figure 2 shows individual participants' choices of utterance forms for Animate and Inanimate targets. The figure shows that participants produced almost entirely passive forms for animate targets but wide variation for inanimate targets. Fifteen participants produced exclusively active utterances to inanimate targets (shown in 0 passive responses at bottom left of the figure), 16 participants produced exclusively passive utterances while the remainder (34) produced a mix of actives and passives. This variability is interesting in itself, though at this point it is unclear why participants made such enormously different choices on an identical task. We return to this issue in the General Discussion.

To quantitatively assess variability of utterance form across animate and inanimate targets, a structure variability score was computed for each participant. This value ranged from 0 (speakers produced only passives in both conditions—the participants shown in the upper right of Figure 2) to 1 (only actives for inanimate and only passives for animates—the participants at the left of the figure). We hypothesize that the structure variability score reflects the degree of competition a speaker experienced when producing sentences in this task. A speaker with a low value on this measure would be expected to have little competition, because they chose passive utterances for virtually every relative clause they uttered in the experiment. A person with a value near 1 would be expected to have high competition, because their responses are split between active and passive utterances, presumably reflecting high availability of both structures during production.

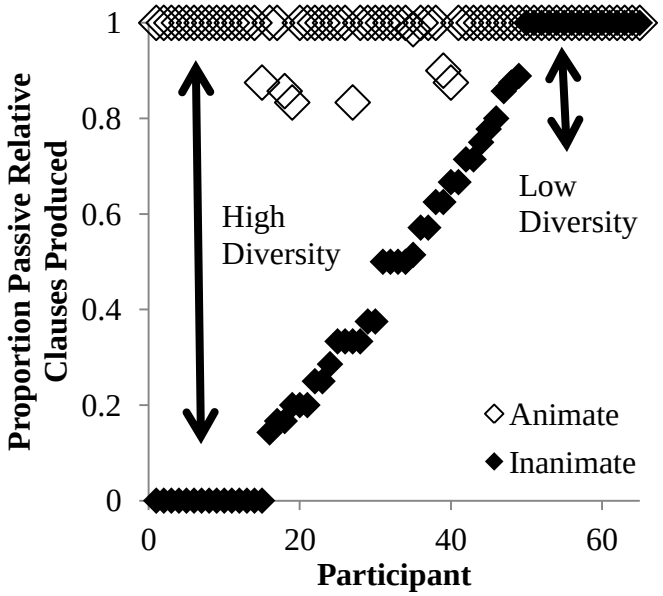

Figure 2: The proportion of passive relative clauses describing animate and inanimate targets, by participant

To test for the presence of competition, we examined the relationship between our production difficulty measures (initiation latency) with a participant's structure variability score. We had previously found that initiation latencies reliably varied with the number of words in the first noun phrase, so we analyzed only trials in which the first noun phrase consisted of two words (711 trials from 64 participants). A simple test for a main effect of animacy on difficulty measures would not be informative, however, because any differences might be due to features of the pictures rather than to availability of alternative structures. For example, the inanimate target *ball* is smaller than the animate target *man* in Figure 1, which could affect initiation latency to answer a question about these targets. For this reason, we defined production difficulty as the difference in initiation latency between animate (passive only) and inanimate trials. Greater production difficulty in inanimate trials (where a structure choice may be present) would be realized as a greater difference in initiation latency between animate and inanimate trials across a speaker's utterances.

To test for the presence of competition, we examined the relationship between this production difficulty measure and participant's structure variability score. We found a positive correlation between a participant's variability score and the difference between animate and inanimate initiation latencies (r=0.25, $p<0.05$), suggesting that latencies to inanimate targets were disproportionately slower in participants who frequently chose between alternate structures. This relationship is illustrated in Figure 3.

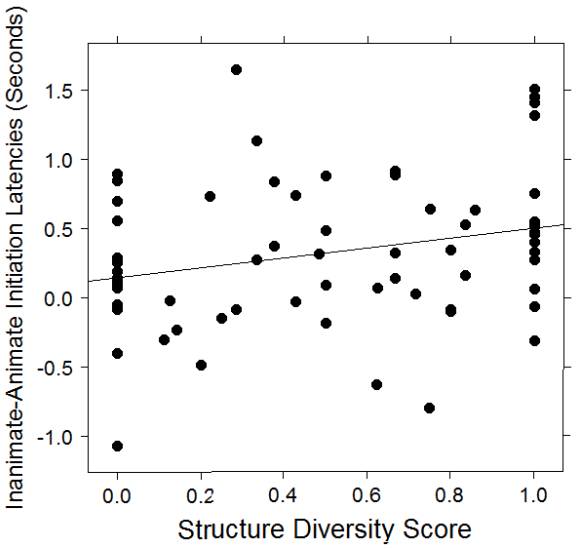

Figure 3: Scatterplot of participants' structure variability scores and the difference between initiation latencies to inanimate and inanimate target items.

Structure variability scores did not correlate with latencies within animate or inanimate trials, indicating that participants there was no relationship between overall speed and variability of form. Participants with a more viable second option to inanimate utterances exhibited differentially longer initiation latencies to inanimate-headed utterances, where a structure choice could plausibly be made[1]. This suggests that those participants who *were* making an active/passive choice on inanimate trials had a more viable second alternative, leading to an effect of competition. Thus at least in this task, the availability of options was costly to speakers, supporting a competitive mechanism in structure choice.

Discussion

The goal of this study was investigate variation in the choice of utterance form in referential expressions containing relative clauses. We found a significant correlation between participants' variability, as assessed by their structure diversity score, and the initiation latency difference between descriptions of animates, where the referential form is near-uniformly passive relatives, and the inanimate target condition, where productions are highly variable and competition between alternative forms could potentially exist. This result suggests that sentence production can proceed competitively.

[1] We also found that a participant's diversity score correlated with relative pronoun use in passive sentences (r=0.26, p<0.05) but because we fail to see an effect in active utterances, we believe that relative pronoun use in passive utterances was primed by active utterances, where these optional words are more common.

In light of this apparent competition, it is important to understand why people are variable and what exactly is competing. This study was not designed to directly answer these questions, but we consider some possibilities below.

Speaker Variability. One possible account of the variability in Figure 1 is that speakers in this task varied in their sensitivity to message/semantic vs. structural factors in utterance planning. More variable participants appear to have been strongly affected by the animacy of the element to be described, such that they produced passives for animate but not inanimate targets. By contrast, the less variable participants may have been less sensitive to animacy and comparatively more influenced by the tendency to repeat sentence structures over time (Bock, 1986). These participants produced passives for the animate targets and continued to use passives for the inanimate ones, independent of what was to be described. On this view, longer initiation latencies for the variable participants may owe to their weighing of more constraints, where structural priming from the animate trials promotes passive forms, and other factors promote active forms; see MacDonald (2013) for some discussion of how multiple constraints on utterance form can create conflict in utterance choices. This study obvious next step is to investigate variability with and without the presence of a strong structural priming factor such as the animate condition here.

The nature of competition. As the above discussion suggests, longer initiation latencies in the face of variable utterance forms may reflect the influence of competing factors, but it does not therefore follow that two syntactic structures are active and competing. That is, the nature of competition may lie at other levels. One possibility is that after a verb is selected, the alternative forms in which it can appear (active or passive, dative or double-object) become activated and compete with each other. This would attribute competition to the simultaneous activation of alternative verb forms when a particular verb is selected. Consistent with this view, Stallings et al. (1998) found that variability in the use of verb forms predicted initiation latencies, such that sentences containing verbs that were associated with more forms had longer initiation latencies.

Similarly, the competition could be attributed to alternative noun frames that are activated when a noun is selected or planned. When a sentence contains multiple nouns, speakers have a choice of which noun to make the grammatical subject or grammatical object. Perhaps these alternative candidates compete when a particular role is being assigned. Just as verbs can appear in alternative structure forms, nouns can appear in either subject or object frames in a sentence, and perhaps the choice of which noun to assign a particular frame can account for the observed competition.

This admittedly speculative discussion suggests that the existence of competition in structure choice need not entail competition between abstract syntactic structures themselves, nor does it suggest that all sentence production is purely competitive. Both competitive and incremental

processes can exist. Speakers do seem to produce utterances as soon as they are planned and continue to plan utterances as they are speaking. However, in the case of multiple viable alternatives, planning becomes more difficult and this pattern of incremental production is disrupted. The plan-as-you-go pattern of incremental production would then be disrupted when a speaker faces multiple structure options which compete with each other

Whether production processes are competitive, incremental, or a combination of the two has many implications for the mechanisms of sentence production. One implication of this data is the extent to which lexical choices are dissociable from syntactic choices which has consequences for how structure may (or may not) be represented and planned by the speaker. Further investigation competitive and incremental processes in sentence production will have many implications for the cognitive mechanisms of production and the representation of structure choices in the mind of the speaker.

Acknowledgments

Jessica L. Montag and Maryellen C. MacDonald, Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706. This research was supported by the National Institutes of Child Health and Human Development [Grants T32 HD049899 and R01 HD047425]; the National Science Foundation [Grant number 1123788]; and the Wisconsin Alumni Research Fund.

References

Abdel Rahman, R. & Melinger, A. (2009). Dismissing lexical competition does not make speaking any easier: A rejoinder to Mahon and Caramazza (2009). *Language and Cognitive Processes, 24*, 749-760.

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

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

Cook, S. W., Jaeger, T.F. & Tanenhaus, M. K. Producing less preferred structures: more gestures, less fluency. *Proceedings of the Cognitive Science Society*, Amsterdam, The Netherlands. July, 2009.

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

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

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*, 715-736.

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

Gleitman, L.R. January, D., Nappa, R., & Trueswell J.C. (2007). On the give and take between event apprehension and utterance formulation. Journal of Memory and Language, 57(4), 544-569.

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

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A. & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition, 33*, 503-535.

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

Montag, J. L. & MacDonald, M. C. (in press). Visual salience modulates sentence choice in relative clause production. *Language and Speech.*

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production Evidence for cascaded processing. *Journal of Experimental Psychology: Language, Memory and Cognition, 24*, 539-557.

Pinker, S. (1984). Language Learnability and Language Development. Cambridge: Harvard University Press.

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

Spalek, K., & Schriefers, H. J. (2005). Dominance affects determiner selection in language production. *Journal of Memory and Language, 52*, 103-119.

Stallings, L. M., MacDonald, M. C. & O'Seaghdha, P. G. (1998). Phrasal ordering constraints on sentence production: Phrase length and verb disposition in Heavy-NP shift. *Journal of Memory and Language, 39*, 392-417.

Tomlin, R. S. (1995). Focal attention, voice, and word order. Word order in discourse, ed. by P. Downing and M. Noonan, 517–52. Amsterdam: John Benjamins.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162-189). Cambridge: Cambridge University Press.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J. & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*, 284-307