Producing Less Preferred Structures: More Gestures, Less Fluency

Susan Wagner Cook (susan-cook@uiowa.edu)
Department of Psychology, E11 Seashore Hall
Iowa City, IA 52242 USA

T. Florian Jaeger (fjaeger@bcs.rochester.edu
Department of Brain and Cognitive Science, University of Rochester
Rochester, NY 14627 USA

Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain and Cognitive Science, University of Rochester
Rochester, NY 14627 USA

Abstract

Speakers often have choices about how to structure their utterances. However, even though multiple alternatives may be acceptable in theory, often one of them will be preferred over the others. The question we explored here was what happens when speakers produce less preferred alternatives. We developed a new experimental paradigm to reliably elicit the propositional or double object dative with varying degrees of preference. We then used this paradigm to investigate how, given properties of the message, an individual speaker's preference for a particular structure affects how that utterance is produced. Speakers gestured more and were more likely to be disfluent when they chose less preferred structures. Thus, having a choice per se does not guarantee more successful production. Instead, production is facilitated when speakers choose more preferred alternatives.

Keywords: speech production; gesture; disfluencies; dative alternation.

Introduction

Grammatical constraints restrict the structures speakers can use to convey their message – not all structures are grammatically acceptable. But grammar also provides speakers with some flexibility, by offering more than one grammatically acceptable way of encoding a proposition linguistically. Moreover, speakers appear to benefit from flexibility. When speakers can choose how to structure an utterance (e.g. when choosing between a double object, DO, or prepositional dative, PD to communicate a transitive event), they on average produce those utterances faster and with fewer errors in comparison with when they can not choose (Ferreira, 1996).

However, even though multiple alternatives may be acceptable in theory, often one of them will be preferred over the others. For example, in Ferreira's (1996) study, in the conditions where choice was available, speakers preferred (65-77%) one of the possible alternatives. The question we explored here was what happens when speakers choose to produce the less preferred alternative.

Preference and Production

Alternative theories of language production appear to make opposing predictions about why speakers produce less preferred structures, leading to different predictions about how these structures should be produced.

Availability-based accounts of production (e.g. Bock & Warren, 1985; Ferreira & Dell, 2000) argue that speakers produce less preferred alternatives because of the material that was more available at the moment of choice. Although the material consistent with the preferred choice is *generally* more likely to be available at the choice point, when it is not, speakers will instead produce the less preferred alternative, allowing them to take advantage of the material that is currently available. On this account, speakers should experience no more or less difficulty when producing less preferred or preferred structures, because these structures are produced in precisely those environments where they are beneficial to the speaker.

Competition-based models of production (e.g. Haskell & MacDonald, 2003; Solomon & Pearlmutter, 2004) assume that speakers produce less preferred alternatives because this structure is more active at the moment of choice, given interactions among all operating constraints. However, when the less-preferred structure is chosen, more often than not, the preferred structure will be more active than the less preferred structure will be when the preferred structure is chosen. Accordingly, speakers should be more likely to run into difficulty when producing less-preferred structures, due to greater ongoing competition from partially active preferred structures. Such difficulty may show in a higher rate of disfluencies and gestures, as well as slowed speech.

These accounts offer explanations into the mechanism underlying production difficulty during less preferred structures. Computational level accounts, such as accounts of rational production (in the sense of Anderson, 1990), also make predictions about the production of less preferred structures. Since less preferred structures have a lower probability of occurring in the context where they are less preferred, they necessarily encode more information. Accordingly, under the principle of Uniform Information

Density (Jaeger, 2006; Levy & Jaeger, 2007), we should expect speakers to provide more *signal* when producing less preferred structures. Additional signal could be provided by speaking slower, by inserting disfluencies, and by gesturing. Disfluencies increase the amount of time over which information is distributed and the signal a higher probability of a priori low-probability information (see discussion). Gestures provide additional signal and channel over which information is distributed.

To explore the relation between preference and production, we designed a new experimental paradigm to vary the degree of preference for one structure over another. This allowed us to both evaluate the competing predictions that availability and competition models make about difficulty when speakers produce a less-preferred structure and to explore whether production was consistent with the principle of Uniform Information Density. Based on previous work, we predicted that speakers would be less fluent and would gesture more as both production difficulty and information density increased.

Reliably Eliciting Preferred and Dispreferred Ditransitive Structures

One grammatical structure that affords choice of structure and that has been widely studied is the dative alternation. Speakers can use either the prepositional dative (PD) or the double object (DO) construction to convey a single message. For example, speakers can choose between saying "Simon gave the backpack to the man" (PD) or "Simon gave the man the backpack" (DO). This choice depends on a variety of factors, including the pronominality, animacy, definiteness, givenness, and weight of the two argument expressions (Bresnan, Cueni, Nikitina, Baayen, 2007), as well as verb bias (Stallings, MacDonald & O'Seaghdha, 1998) and syntactic priming (Bock, 1986).

Production experiments often have high exclusion rates. Those that succeed often use recall tasks to elicit particular structures (Bock, 1986) or use highly unnatural tasks (e.g. constructing sentences from words, Ferreira, 1986). Corpus studies, in contrast, have been criticized for lack of control, (but see Bresnan et al 2007; for discussion of corpus vs. experiment, see Jaeger 2009). Our goal was to elicit utterances under relatively natural conditions while achieving a high yield of the desired verbs and structures.

We developed experimental stimuli designed to elicit precisely the sort of linguistic variety known to predict dative alternation in corpus studies (Bresnan et al., 2007). We used animated videos to elicit verbs known to have varying amount of bias, and we included contrasting elements in some scenes to manipulate the weight of the noun phrases necessary to describe the elements in each scene.

Methods

Subjects

Sixteen college students participated in the current study; video data from one subject was lost and not included in the data analysis.

Stimuli

Our stimuli were computer-animated vignettes of a human-like actor interacting with other people and objects created using Poser software. The experimental stimuli were designed to elicit the following six ditransitive verbs: *give*, *hand*, *offer*, *show*, *take* and *throw*. These verbs were chosen because they were predicted to vary in verb bias based on previous research. The filler items were designed to elicit the following six transitive verbs: *carry*, *kick*, *lift*, *push*, *pull*, and *touch*.

A single animated figure ("Simon") was the agent in all of the vignettes. Repetition of the agent throughout the experiment was used to increase the likelihood of this role being the subject of the sentence, yielding more structures of the desired type (active ditransitive structures).

Experimental items included at least one additional animated human recipient and at least one inanimate theme. Unlike the agent, the themes and the recipients were never repeated across items. To manipulate argument expression weight (Bresnan et al., 2007), experimental items differed in whether the scene contained a contrast referent to the recipient and/or theme or neither. Contrast referents were always identical to the target referents involved in the main action, except for one distinctive visual feature. There were four stimulus conditions: No Contrast, Recipient Contrast, Theme Contrast, and Both Contrasts (see Figure 1). A given participant saw each item in only one of the possible conditions.

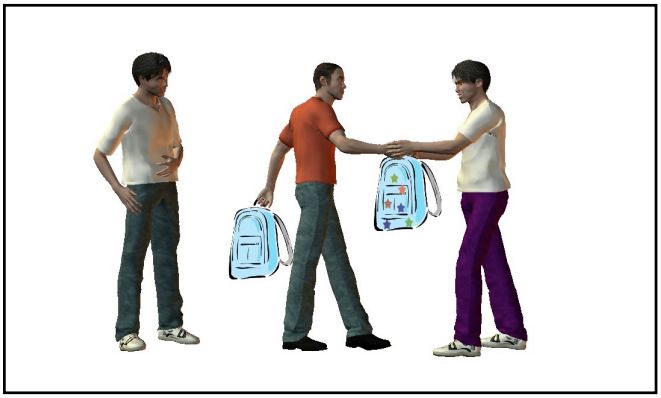

Figure 1: Still image from the animation designed to elicit contrast on both the theme and the recipient (as in "Simon gave a backpack *with stars* to a man *with purple pants*")

Procedure

Participants were instructed to describe "what Simon does in each vignette" to an experimenter "as if you were describing it naturally to someone who had never seen it before". An undergraduate experimenter listened attentively and provided naturalistic feedback. The entire procedure was videotaped for subsequent analysis.

Coding

All speech was transcribed and annotated for disfluencies. The disfluencies annotated were pauses, filled pauses, word repetitions, sentence restarts, fillers, and word lengthening. We plan to extend the annotation to include speech rate information, to assess whether speaker slow their speech rate when producing less preferred structures (as suggested by the corpus study presented in Tily, Gahl, Arnon, Kothari, Snider, & Bresnan, 2009).

The presence of hand gestures during target utterances was also coded. We excluded functional movements (scratching, or adjusting hair) and counted all remaining hand movements as gestures.

Results

The stimuli reliably elicited the intended verbs, contrast and structures from speakers, even though there were only general instructions and no pretraining on the items (77% usable trials). Speakers also produced referring expressions of varying weight during the target utterances (Table 1). Moreover, speakers reliably used a small set of verbs, providing enough data to estimate the structural bias for the included verbs during the experiment, as can be seen in Table 2.

Table 1: Variation in referring expressions across conditions.

	Theme Heavy	Recipient Heavy
No contrast	.19	.28
Theme contrast	.41	.23
Recipient contrast	.10	.78
BothContrast	.43	.74

Table 2: Verbs elicited in the study and their bias estimated when controlling for other factors known to predict dative alternation.

Verb	PD bias in our data
Show	0.27
Offer	0.32
Give	0.34
Hand	0.52
Toss	0.72
Throw	0.89
Take	0.98

As expected given previous work (Bock, 1986; Bresnan et al., 2007; Stallings et al., 1998), participants' choice of syntactic structure was affected by a variety of factors. We analyzed the data using multilevel logistic regression to predict the log odds of producing a PD, based on the verb, theme and recipient pronominality, theme and recipient weight (whether the argument expression was realized with a contrastive prepositional phrase, e.g. "backpack with stars"), as well as the syntactic structure of the immediately preceding experimental item. We also included random intercepts for subject and item effects. As can be seen in Figure 2, all factors had expected effects on speaker's preference for PD. Speakers were more likely to produce a PD when the preceding utterance was a PD (syntactic priming, β=1.74, z=4.33, p<.01), when the theme was referred to by a pronoun (i.e. given and likely to appear early in the utterance, β = 2.10, z=2.60, p<.0001) and when the recipient was realized with contrastive modification (heavier and likely to appear later in the utterance, β = 1.31, z=4.81, p<.00001). Conversely, speakers were less likely to produce a PD when the preceding utterance was a DO (β = -.67, z=1.98, p=.048), and when the theme was realized with contrastive modification (β =-0.72, z=2.60, p<.01). Recipient pronoun did not reach significance.

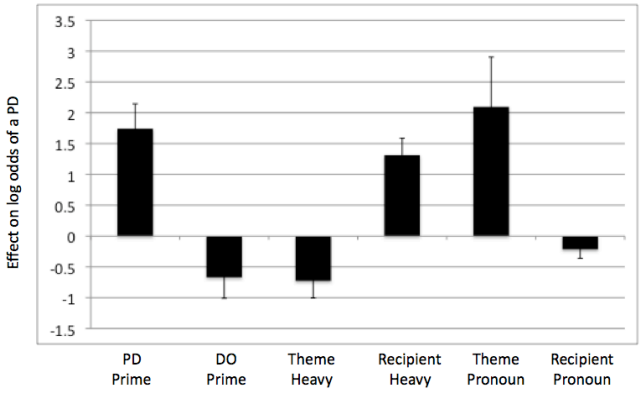

Figure 2: Effects on log odds of producing a PD.

Thus, the stimuli reliably elicited ditransitive structures that differed along the intended dimensions (complexity of the argument expressions and verb bias), leading to a widely spread distribution of PD bias across the entire experiment (.19 to .9995). We then used the PD bias estimate from our model to determine the preference for the actually produced structure as the probability of producing it (i.e. when a DO was produced, we subtracted the predicted probability of producing a PD from one to obtain the predicted probability of a DO). The preference for the structures that were actually produced ranged from .007 to .995. Note that the multilevel model accounted for random speaker differences. Thus, we are now in a position to investigate how, given certain properties of the message, an individual speaker's

preference for a ditransitive structure affects how that utterance is produced.

Preference and Production Difficulty

We conducted two separate analyses to investigate the relation between speakers' preference for a structure and the production difficulty experienced while producing it. The first analysis investigates the distribution of disfluencies, the second investigates the distribution of gestures.

Disfluencies can be seen as overt signs of production difficulty. Speakers are more disfluent before words that are unlikely (Beattie & Butterworth, 1979; Shriberg & Stolcke, 1996), and when they are planning longer and more complex utterances (Clark & Wasow, 1998).

Gesture and Speech Production

In addition to disfluencies, hand gestures may offer a privileged window onto speech planning because they can occur simultaneously with speech. Indeed, gestures are precisely timed with respect to the accompanying speech (McNeill, 1992). Moreover, gestures have been related to factors known to be important for speech production, and are associated with information that is conceptually more difficult. For example, gestures are more likely to occur with infrequent words, and with new information. Gestures may even function for speakers in these situations, facilitating lexical access (Morrel-Samuels & Krauss, 1992) and/or lightening cognitive load (Goldin-Meadow, Nusbaum, Kelly & Wagner, 2001).

We used a multilevel logistic regression to predict the probability of producing at least one disfluency during the target utterance, given the preference for the structure (i.e. the probability of the structure as estimated above), and controlling for the number of words in the utterance. We also included random subject and item effects. There was a significant and negative effect of preference on disfluency (β = -.28, z=2.61, p<.01, See Figure 3). The more less preferred the utterance, the more fluently it was produced.

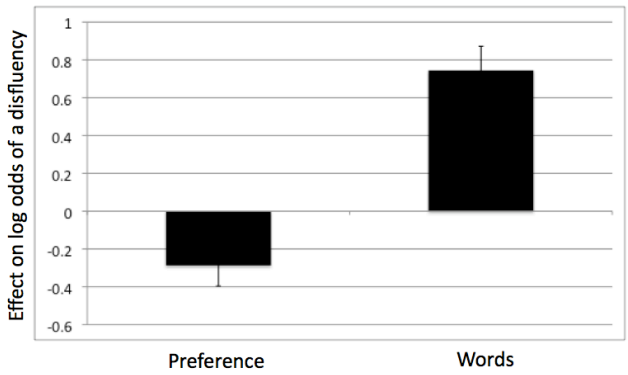

Figure 3: Effects of preference and the number of words on log odds of producing a *disfluency*.

As a second look at the relation between preference and production, we explored how the preference for each utterance was related to the probability of gesturing during that utterance. We again used a multilevel logistic regression to predict the probability of producing a gesture during the target utterance, given the preference of producing the structure, and controlling for the number of words in the utterance. There was a significant and negative effect of preference on gesture. The more preferred the structure, the less likely a speaker was to gesture during that utterance (β = -.41, z=2.43, p=.015, See Figure 4).

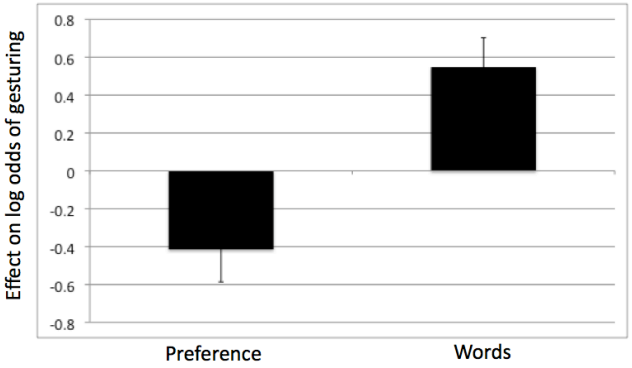

Figure 4: Effects of preference and the number of words on log odds of producing a *gesture*.

Thus, our findings indicate that, across the utterance as a whole, higher preference were associated with less difficulty during production, and less recruitment of multiple modalities during production. Thus, these findings are consistent with the claim that speakers experience greater difficulty producing less preferred structures. However, models of language production predict difficulty or facilitation at a particular point in a speaker's utterance. While competition-based models are compatible with a longer lasting effect of choosing a dispreferred structure, production difficulty should be highest during the noun phrase immediately following the choice point (the verb), because it should be relatively less active. In contrast, availability-based models suggest that this noun phrase should be readily produced, as it is the one being chosen *because* of its availability. Accordingly, we conducted follow-up analyses focusing only on the first noun phrase produced by speakers after the verb.

Limiting our analysis to the portion of the utterance where speakers commit to a particular structure did not change the pattern of findings. Speakers were again more disfluent during the first noun phrase when the structure was less preferred (β = -.32, z=2.78, p<.01), and they were more likely to initiate a gesture during this clause (β = -.69, z=2.12, p=.03). Thus, as speakers commit to a particular structure, production reflects the probability of that structure.

Discussion

These findings reveal that both fluency and gesture can be predicted from the probability of (preference for) the

syntactic structure that is in the process of being produced. When speakers produce more preferred structures, they gesture less and are more fluent. Thus, having a choice per se does not guarantee more successful production. Instead, production is facilitated when speakers choose the preferred alternative.

This result is unexpected for availability accounts (e.g. Bock & Warren, 1985; Ferreira & Dell, 2000), but consistent with competition accounts (e.g. Haskell & MacDonald, 2003; Solomon & Pearlmutter, 2004) of selection during syntactic production. The results are also unexpected for any account that consider the very notion of "producing a less preferred structure" to be ill-defined, because they are based on the assumption that whatever speakers produce in a given instance (by definition) *is* the preferred structure. Our results argue against this view. Speakers occasionally produce structures that they themselves usually disprefer in the given context and when they do so, they are more likely to be disfluent and to produce gestures.

One possibility is that these tendencies may arise in speakers' own production system, as suggested by competition accounts of language production. If producing less preferred structures is more difficult, then speakers are expected to be more disfluent. Disfluencies may also reflect speakers' accommodation of demands on listeners' processing systems. In comprehension, less expected material is processed more slowly, and speakers may use disfluencies as a signal to dynamically adjust listeners' expectations. Indeed, comprehenders are faster at processing expressions that introduce new discourse referents when the expression is preceded by a filled pause (Arnold et al., 2004). Whether speakers *intend* disfluencies as a signal to the comprehender remains an open question.

Like disfluencies, gestures may arise via demand on the speaker's production system, or as an overt signal to the listener's comprehension system. As a communicative signal, gesture may be a useful tool for revealing the particular alternatives that are being considered in a particular moment, both for experimenters interested in exploring the production process, and for listeners seeking to understand their conversation partners. Speakers often include information in their gesture that is not available in their speech (gesture-speech mismatches). Moreover, production of these gesture-speech mismatches reflects choice among various alternatives (Garber & Goldin-Meadow, 2002). Thus, one possibility is that, like disfluencies (Arnold et al., 2004), speakers' gestures may signal the presence of the less expected structure to listeners.

Alternatively, gestures may help speakers alleviate demands on their own production systems (Goldin-Meadow, Nusbaum, Kelly & Wagner, 2001; Melinger & Kita, 2007). Gestures have been hypothesized to facilitate both lexical and conceptual processing during speech production (e.g. Alibali, Kita & Young, 2000; Rauscher, Krauss & Chen, 1996). In the study reported here, disfluencies and gestures tended to co-occur, which is consistent with the hypothesis that they may arise via a common mechanism.

Additional work will be necessary to determine whether gestures and disfluencies are functional for the speaker or for the listener, or for both. Because some gestures are sensitive to the visual co-presence of a listener while other gestures are not (Alibali, Heath & Myers, 2001), further experiments varying the visual co-presence of the listener will allow exploration of whether speakers are designing the gestures and the associated disfluencies explored here for their listeners. If speakers intend disfluencies and gestures as communicative signals, speakers should increase production of disfluencies when gesture is unavailable to the listener, particularly for less preferred structures.

The findings reported here are also consistent with the hypothesis of Uniform Information Density (Jaeger, 2006; Levy & Jaeger, 2007). Uniform Information Density predicts that speakers prefer to distribute information uniformly across their utterances. This can be shown to be an optimal strategy with regard to communicative success (Shannon, 1948; cf. Aylett & Turk, 2004; Genzel & Charniak, 2002) and processing difficulty (Levy & Jaeger, 2007). Uniform Information Density, while compatible with an audience design interpretation, does not assume that speakers design their utterances so as to ease comprehension. Uniform Information Density can be derived to be a rational production strategy without reference to comprehension (Levy & Jaeger, 2007; for further discussion on the relation between Uniform Information Density and audience design, see Jaeger, 2006, in progress). Uniform Information Density hence provide a computational level account of what properties the production system should exhibit if it was structured in such a way as to support efficient production.

Because the information content of a particular word is inversely related to its probability, our findings reveal that the more informative a particular argument, the more disfluent the production and the more gestures speakers produce along with it. Thus, speakers may use disfluencies to distribute highly informative arguments over a greater amount of time, decreasing the information density and avoiding high information peaks. Indeed speakers are more likely to produce disfluencies before less probable words (Beattie & Butterworth, 1979; Shriberg & Stolcke, 1996) and in low probability syntactic environments (our results; see also Tily, Arnon, Bresnan, Kothari, & Snider, 2007). Disfluencies (and gestures) may then be seen as increasing the probability (and hence lowering the information content) of a priori low-probability words following them (cf. Jaeger, 2006, and ongoing work).

Consistent with the hypothesis of Uniform Information Density, our findings also suggest that speakers may utilize multiple channels in order to avoid information peaks. When speakers produce more informative arguments, they are more likely to distribute information across auditory and visual channels of communication via gesture. Previous

work exploring gesture production has similarly found that gestures are produced with informative speech (McNeill, 1992). Future work will be necessary to explore whether listeners use information from gestures as a visual cue to the information content of the associated speech.

Acknowledgments

This work was supported by NIH grant HD-27206 to MKT.

References

Alibali, M.W., Heath, D.C., Myers, H.J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Alibali, M. W., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15, 593-613.

Anderson, J. R. (1990) *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: disfluency and reference resolution. *Psychological Science, 15*, 578-82.

Aylett, M. & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America, 119*, 3048-3058.

Beattie, G.W., & Butterworth, B.L. (1979). Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech, 22*, 201-211.

Beattie, G. & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. British Journal of Psychology, 91, 473-491.

Bock, J.K. (1986) Syntactic Persistence in Language Production. *Cognitive Psychology, 18*, 355-387.

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

Bresnan, G., Cueni, A., Nikitina, T, & Baayen, R.H. (2007). "Predicting the Dative Alternation." *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts. Amsterdam: Royal Netherlands Academy of Science.

Clark, H. H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201–242.

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.

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

Garber, P. & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, *26*, 817-831.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 199-206.

Goldin-Meadow S., Nusbaum, H. C., Kelly, S. D., & Wagner, S. (2001) Explaining math: Gesturing lightens the load. *Psychological Science*, *12*, 516-522.

Haskell, T. R. & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, 48, 760-78.

Jaeger, T. Florian. 2006. Redundancy and Syntactic Reduction in Spontaneous Speech, Stanford University.

Levy, R., and Jaeger, T. F. 2007. Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, and T. Hoffman (Eds.), Advances in neural information processing systems (NIPS) 19, 849-856. Cambridge, MA: MIT Press.

Melinger, A., & Kita, S. (2007). Conceptualization load triggers gesture production. *Language and Cognitive Processes, 22*, 473-500.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.

Morrel-Samuels, P. & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 615-622

Rauscher, F. H., Krauss, R. M. & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 226-231.

Shannon, C. E. (1948) A mathematical theory of communication. *The Bell System Technical Journal, 27,* 379-423, 623-656

Shriberg, Elizabeth, and Stolcke, Andreas. 1996. Word predictability after hesitations: A corpus-based study. ICSLP 96.

Solomon, E.S. & Pearlmutter, N.J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49*, 1-46.

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

Tily, H., Arnon, I., Bresnan, J., Kothari, A., & Snider, N. (2007). What makes a construction predictable? Using semantic and contextual cues to better model phonetic reduction. CUNY Sentence Processing Conference.

Tily, H., Gahl, S., Arnon, I., Kothari, A., Snider, N., & Bresnan, J. (2009). Pronunciation reflects syntactic probabilities: Evidence from spontaneous speech. *Language and Cognition, 1*, XX-XX.