Semantically underinformative utterances trigger pragmatic inferences

Ekaterina Kravtchenko (eskrav@coli.uni-saarland.de) and **Vera Demberg (vera@coli.uni-saarland.de)**
Computational Linguistics, SFB 1102; Saarland University
Campus C7.4; 66123 Saarbrücken, Germany

Abstract

Most theories of pragmatics and language processing predict that speakers avoid informationally redundant utterances. From a processing standpoint, it remains unclear what happens when listeners encounter such utterances, and how they interpret them. We argue that uninformative utterances can trigger pragmatic inferences, which increase utterance utility in line with listener expectations. In this study, we look at utterances that refer to stereotyped event sequences describing common activities (*scripts*). Literature on processing of event sequences shows that people automatically infer component actions, once a script is 'invoked.' We demonstrate that when comprehenders encounter utterances describing events that can be easily inferred from prior context, they interpret them as signifying that the event conveys new, unstated information. We also suggest that formal models of language comprehension would have difficulty in accurately estimating the predictability or potential processing cost incurred by such utterances.

Keywords: Psycholinguistics; pragmatics; redundancy; information theory.

Introduction

Many theories of language processing, pragmatics, and dialog contain constraints against utterances that add no new information to the discourse, or are informationally redundant (Cohen, 1978; Grice, 1975; Levy & Jaeger, 2007). Informationally redundant utterances, while typically viewed as infelicitous in the linguistics literature, have also been investigated in the dialog literature. These utterances might state information already in the common ground, echo information already stated, or provide information that was entailed (or strongly implied) by the preceding discourse (Walker, 1993). Baker, Gill, and Cassell (2008) note that redundant utterances are surprisingly frequent, and often used in response to signs of listener non-comprehension, when responding to listener questions, or when speaking to strangers. Similarly, Walker concludes that informationally redundant utterances are common in discourse, and that aside from addressing cognitive resource limitations, they often serve a narrative function.

It is, however, less clear how comprehenders process or interpret utterances that are informationally redundant with respect to their own discourse model – in other words, ones that do not address limitations in listener comprehension, make explicit inferences that may not otherwise be obvious, and so forth. We present evidence that certain types of informationally redundant utterances trigger pragmatic inferences, which increase their informational utility, presumably because they otherwise violate conversational norms with respect to utterance informativity (Grice, 1975). While these results, broadly, are predicted by formal models of pragmatic reasoning, it is less clear whether formal language models

that attempt to model utterance predictability, and the resulting processing cost (e.g., Smith & Levy, 2013), would correctly represent either: a) the predictability of these utterances, if they are in fact infrequent (c.f., Walker, 1993); and b) any difficulty comprehenders encounter in processing them.

In general, there have been no strong positions taken in the psycholinguistics literature on whether redundancy should impose processing cost on the listener. Nadig and Sedivy (2002) and Arnold (2008), in the context of referring expression complexity, argue that redundancy may only be suboptimal from a speaker's perspective, and if anything may facilitate comprehension. On the other hand, Smith and Levy (2013) point out that utterances which assert 'obvious facts,' or otherwise overly predictable information, may be easy to integrate semantically, but incur processing cost given their (presumably) low probability of occurrence. To our knowledge, however, this has not empirically been investigated.

Although our study does not directly address the processing cost incurred by overly redundant utterances, context-dependent implicatures are often assumed to incur some amount of cost, particularly when there is no explicit contextual support for the implicature drawn (Sedivy, 2007). Further, if these types of utterances are as infrequent as predicted by most pragmatic theories, they would be expected to incur cost on this basis. We therefore discuss the potential implications of this for later work. Finally, we argue in the Discussion section that any difficulty listeners *do* encounter in processing these utterances would, problematically, not be fully accounted for by formal models of sentence comprehension which don't incorporate pragmatic reasoning.

In our study, we focus on utterances that are uninformative[1], in that they explicitly describe events in common, stereotyped sequences that are normally automatically inferred (Bower, Black, & Turner, 1979). To our knowledge, the pragmatic interpretation of redundant event mentions has likewise not to date been investigated experimentally, nor addressed in the formal literature.

Processing of Event Sequences

In this study, we look at overt descriptions of events that are part of typical and stereotyped event sequences (*scripts*), such as *going to a restaurant*, or *using the subway*. Literature on processing of event sequences shows that people anticipate upcoming events or future states once a script is 'invoked', and have faster reading times when information is consistent with previous script knowledge (Zwaan, Magliano, &

---

[1]To clarify, we use the term *uninformative* to describe utterances that contribute no new information to the common ground.

Graesser, 1995). Likewise, when recalling stories based on a variety of scripts, people have significant difficulty distinguishing between actions that were implied by the script, but not mentioned overtly, and those that were actually mentioned in the story (Bower et al., 1979). This finding suggests that event occurrences that are strongly implied by a script are part of its conventional meaning (unless explicitly indicated otherwise), whether they are explicitly mentioned or not.

Certain types of information may be more or less central to a script - for example, going *grocery shopping* almost always entails *paying for your groceries*, but does not typically entail *buying apples* (although the latter is entirely consistent with the script). The utterances we are interested in refer to central events that are easily inferred in the context of the event sequence – for example, *entering a restaurant*, in the context of a restaurant visit:

(1) a. Yesterday, Marie went out to eat. She **entered the restaurant**. She ordered pasta, and read a book.
    b. Yesterday, Marie went out to eat. She ordered pasta, and read a book.

Presumably, readers would have automatically inferred, in the second variant, that *Marie* had entered the restaurant, although it was not mentioned explicitly, and would find the first variant slightly anomalous. Highly inferable events are occasionally used as temporal anchors (*After she entered the restaurant, she...*), and may be used to transition back from interruptions to the script (*She stopped to talk to Brad on the street. She then entered the restaurant...*). However, outside of these contexts, mutually known and highly inferable events in scripts are often not mentioned overtly (Bower et al., 1979; Regneri, Koller, & Pinkal, 2010).

Interpretation of Redundant Utterances

There is evidence that speakers avoid redundancy, outside constraints on grammar and meaning, at every level of linguistic processing (Jaeger, 2010). Likewise, comprehenders should expect producers to make choices that conform to these norms. At the utterance level, they approximate Grice's Quantity maxim, which dictates that speakers should provide neither more nor less information than is required to communicate a meaning in context (Grice, 1975). According to the latter, listeners would be predicted to draw (highly context-sensitive) implicatures from informationally redundant utterances, if they are to preserve the belief that the speaker is cooperative.

An example of an utterance that might give rise to a context-dependent implicature is the following, uttered by one speaker to another (assuming both are familiar with *John* and *John's* typical habits):

(2) "John went grocery shopping. **He paid the cashier!**"

In this example, the speaker is arguably flouting a conversational maxim, by stating something which can be clearly inferred from preceding information. The event described by the second, uninformative utterance is already strongly implied to have occurred by having invoked the *grocery shopping* script, and mentioning it overtly is therefore redundant. Such an utterance in itself contributes no extra information, and if it is to conform with listener expectations about speaker informativity, then the listener is likely to attempt to infer what *other* information the speaker attempting to convey.

The most obvious inference appears to be that this behavior is atypical for *John*; i.e. that it could not have been inferred, given what had already been stated, that the behavior described had taken place. In other words, a straightforward interpretation might be that *John* is not typically in the habit of paying at the grocery store. The only case in which the default inference(s) might not be made is if information was introduced into the common ground already suggesting that this was an atypical event – say, that *John* is a habitual shoplifter.

In summary, on the one hand, reading something that is very easy to infer, given world knowledge, intuitively should not cause any particular difficulty. However, making assertive utterances that contain no useful information clashes with listener expectations of speaker behavior, and presumably occurs only infrequently (Smith & Levy, 2013). Listeners are therefore predicted to attempt to 'make sense' of the speaker's behavior by coming up with an alternate (informative) meaning for the utterance.

Methods

Participants

200 participants, located in the US, were recruited on Amazon Mechanical Turk.

Materials

24 stimuli, written as brief stories/narratives, were constructed, based on distinct scripts or activities. Each appeared with 2 initial contexts (*typical* vs. *atypical*), and assertive utterances described 2 types of events (*predictable* vs. *optional*), for a total of 4 conditions[2].

Predictable events ([4a] in examples (3)) would typically be inferred from the 'speaker' having invoked the script, while *optional* events ([4b]) could not be inferred. To clarify, we are using the term *predictable* to specify that the event can typically be inferred from the script, although this is, as we show, dependent on context. Initial discourse context was either *typical* ([1a], in (3a)), or *atypical* in that it implied the *predictable* event was in fact unusual ([1b], in (3b)):

(3a)   Typical context

[1a] John often *goes to his local supermarket, as it's close by*$_{typical}$.

[2] Today he entered the apartment with his shopping bags flowing over. He ran into Susan, his best friend, and talked

---

to her about his trip. Susan then wandered over to Peter, their roommate, who was in a different room.

<first (baseline) set of questions>

[3] She commented: "John went shopping. [4a/b] He {*paid the cashier₍a-predictable₎* | *got some apples₍b-optional₎*!} [5] I just saw him in the living room."

<second (final) set of questions>

In contrast, the following context manipulation was used in order to render the *predictable* event optional, or unusual:

(3b)   Atypical context

[1b] John often ***doesn't pay at the local supermarket, as he's usually broke₍atypical₎***.

[2] Today he entered the apartment with his shopping bags flowing over. He ran into Susan, his best friend, and talked to her about his trip. Susan then wandered over to Peter, their roommate, who was in a different room.

<first (baseline) set of questions>

[3] She commented: "John went shopping. [4a/b] He {*paid the cashier₍a-predictable₎* | *got some apples₍b-optional₎*!} [5] I just saw him in the living room."

<second (final) set of questions>

The story was presented incrementally. Participants first saw the first two paragraphs (without numbering or formatting), and the first set of questions. These always queried how often the subject of the discourse engaged in the *optional* activity, the *predictable* (inferrable) activity, and two scenario-relevant distractor activities (not shown):

1. How often do you think John usually *pays the cashier*, when at the grocery store?
2. How often do you think John usually *gets apples*, when at the grocery store?

Each question could be responded to on a continuous sliding scale of 'Never' to 'Always' (see Figure 1). The slider was not visible until the participant clicked on the point on the scale that they thought was most appropriate, to avoid having people default towards a particular value.

After responding to each question, the participant could submit their answers. Once they did, the first question set disappeared, and the rest of the text was revealed, as well as the second set of questions. At this point the entirety of the text was visible to participants. The questions referred to the same activities, but asked participants to update their estimates – "*Now how often do you think...?*". Participants could not see their previous answers.

12 of the stimuli included 3 discourse participants, one of whom engaged in the activity, the second who learned from that participant that they engaged in it, and the third to whom the second communicated this fact (as well as the relevant specifics). The other 12 only included two – the discourse subject who engaged in the activity, and the second participant to who they communicated this fact. Compared to the ex-

ample above, for instance, *John* might be communicating directly to *Susan*: "*I went grocery shopping. I paid the cashier! I just got back from the store.*".

The construction of these stimuli was constrained in several ways. The scripts were sufficiently complex to contain multiple candidate activities; there were central as well as optional components; and the central components were such that the activity could still conceivably occur without them. For example, one arguably cannot play *tennis* at all, without using a *racket*. There was also established common ground between all discourse participants, so that all were plausibly (from the point of view of the reader) aware of the discourse subject's typical habits, particularly with regard to the activity described.

Further, the activities were sufficiently stereotyped and (relatively) culturally invariable, so that all participants could be expected to agree on what a script entailed, what was or wasn't obligatory, etc.. Many of the scenarios, as well as estimates of 'centrality' of the events for the script, were taken from data provided by the authors of Raisig, Welke, Hagendorf, and van der Meer (2009), who looked at retrieval of script knowledge from memory. Additional scenarios were created where participant norming suggested that some scripts did not meet our criteria, were insufficiently familiar, or excessively culturally variable.

Procedure

Each participant was asked to read 4 stories randomly selected out of the total of 24, with each condition only presented once (given the very small number of items, there were no fillers). They initially saw the first part of the story only ([1] and [2] in (3a/b)), and the first set of questions presented in random order. After responding to all questions by selecting a point on the *Never* to *Always* scale (Figure 1), they were able to proceed. The first set of questions, including the responses, was then hidden, and the rest of the story ([3] through [5] in (3a/b)) was presented. This was followed by the second series of the questions, again presented in random order, asking for updated ratings. When answering the second set of questions, the entire story was visible to participants.

Never   Sometimes   Always

Figure 1: A sliding scale as used in the experiment.

Prior to seeing the experimental items, participants were initially given several practice questions, unrelated to the stimuli above, also using continuous sliding scales ranging from *Never* to *Always* (or similar). Unlike the experimental stimuli, these questions had 'correct' answers – such as *How likely is a fair coin to come up heads* {*never* | *twice* | *5 times*}*, if flipped 10 times*. If participants provided responses that could not be judged reasonably accurate, they were asked to re-read the instructions, and respond again, before they were able to proceed. This ensured that they were able to follow instructions, and were less likely to guess randomly throughout the

experiment. There were no 'accurate' answers, or 'accuracy' checks, in the actual experiment.

Predictions

We predicted that participants' final estimates of how often the discourse subject engaged in the relevant activity (*paying the cashier*) would decrease from baseline following the *predictable* event utterance (*He paid the cashier*), in the *typical* context (which only invokes the script). Here, participants are expected to infer that *John paid the cashier* was mentioned because it was 'informative'; i.e., unpredictable, or atypical.

In contrast, the same utterance (*He paid the cashier*) would not be expected to generate a similar inference in the *atypical* context, as that context already establishes the *paying* event as atypical, or 'informative.' There, participants' final estimates of how often the discourse subject (*John*) engaged in the relevant activity (*paying the cashier*) would either not change from baseline, or would increase slightly (as the critical utterance now provided evidence that *John* at least sometimes paid the cashier).

Likewise, the *optional* event utterance (*He got some apples*) would be expected to either not change, or slightly increase estimates of how typically the discourse subject (*John*) engages in this activity. Prior to having evidence that *John* at least sometimes *gets apples*, participants would have no evidence regarding the usual activity frequency, and might give low estimates initially. This should occur regardless of context, which provides no information regarding the baseline typicality of the *optional* activity.

Results

Baseline Activity Typicality

The responses to the first set of questions, aside from setting a baseline, also provide a norming measure for participants' prior expectations regarding how likely it is that the discourse subject habitually engages in a particular activity. As would be expected, the average baseline typicality rating of the *predictable* (typically inferable) activity in the *typical* context (e.g., *paying the cashier*, when shopping) was rather high, at 79.67 (1-100), while the baseline estimate for the same activity in the *atypical* context (where *John* is immediately introduced as someone unlikely to pay) was relatively low, at 41.89. The distribution of estimates can be seen in Figure 2.

The average baseline typicality ratings for the *optional* (non-inferable) activity (e.g., *getting apples*) were also comparatively low regardless of context, at 38.06 and 35.68. These numbers rule out the possibility that participants were simply responding randomly, or in a manner counter to the intended effects of context or activity 'inferability.'

Change in Activity Typicality Estimates

The variable of main interest was the change in rating, from baseline estimate to final estimate, of how often the participants thought the discourse subject typically engaged in the activity mentioned in the critical utterance ([4a/b] in the stimuli above). For the analysis, all factors were effect/sum coded.
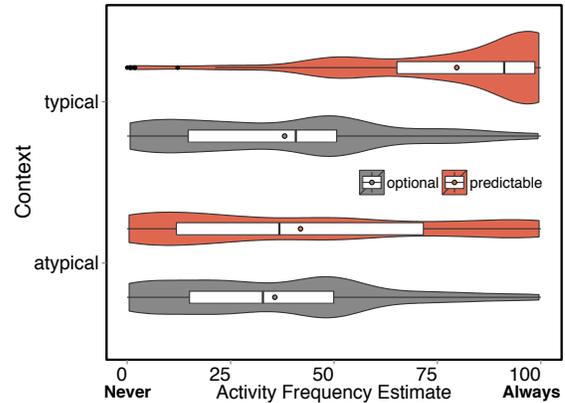

Figure 2: This plot shows the baseline activity typicality estimates, on a scale of 0–100. The violin plots (overlaid on box plots) show the distribution of estimates. The hollow circles represent mean values.

A linear mixed effects regression analysis, fitted using χ^2-test model comparison, showed that speakers are more likely to decrease their estimates of activity typicality in the conditions with the *typical* context (β = -5.06, p<.001), and those in which the critical utterance describes the *predictable* (inferable) activity (β = -6.98, p<.001) (see Table 1). We used the maximal model, with by-subject random intercepts and slopes for context and event predictability, as well as by-item random intercepts and slopes for both factors and their interaction. Critically, there was a significant interaction, driven by participants being significantly more likely to lower estimates of activity typicality in the condition where the utterance described a *predictable* (inferable) activity in a *typical* context (β = -5.08, p<.001). This is shown in Figure 3.
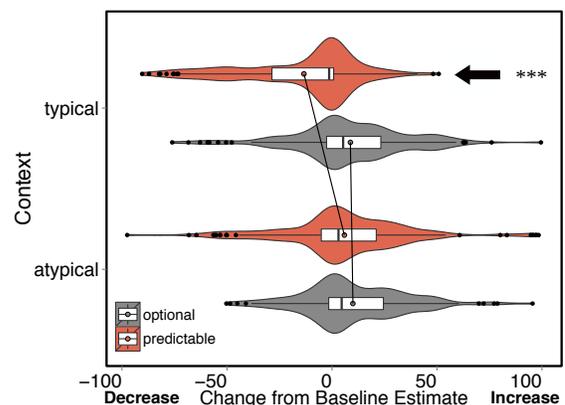

Figure 3: This plot shows the change from baseline estimates, following the *predictable* or *optional* event utterances.

The only condition in which estimates decreased from the baseline was in the *typical context–predictable event utterance* condition. Numerically, there was in fact a slight increase in estimates following all other conditions.

Table 1: This table shows the beta coefficients associated with each main effect in the model, as well as corresponding standard errors, *t*-values, and significance levels.

	Coef β	SE(β)	t	p
Intercept	2.85	1.87	1.52	0.14
Context: Typical	−5.06	0.97	−5.21	<**.001**
Event: Predictable	−6.98	1.53	−4.56	<**.001**
Context * Event	−5.08	1.13	−4.49	<**.001**

One potential problem with the design is a 'regression to the mean' effect in the *typical context–predictable event utterance* condition. The baseline estimates in this condition are relatively high (79.67), and many participants gave this condition a rating above 90 (1-100), with the median rating at 91.13. Given the predictability of the activity given the script, participants had more room to re-adjust their estimate downward than upward on the scale, as their initial estimate was closer to 100 than to 1. This gives rise to the possibility that participants were simply inclined to change their initial estimates in one way or another, and that the decrease seen is an artifact of the task design.

However, the question regarding the *predictable*, or 'inferable' activity was also present in the *typical context–optional event utterance* condition, where a change in typicality ratings would not be predicted (as there was no mention of *John* '*paying for the groceries*,' or not). The same hypothetical tendency to alter one's initial estimate would presumably be present in this condition, as well; however, no change is seen (+0.36). This strongly suggests that the decrease in estimates from baseline, in the *typical context–predictable event utterance* condition, reflects a genuine change in belief, and is not simply a byproduct of the task.

Discussion

To our knowledge, these are the first experimental findings of uninformative event descriptions giving rise to pragmatic inferences, which bring the utterance interpretation more in line with listener expectations of speaker informativity. This would be predicted by a large number of linguistic theories which, on the one hand, propose that speakers should avoid excessive redundancy; and on the other, predict that listeners, acting on the expectation that speakers will be informative, will draw pragmatic inferences from apparently uninformative utterances. Although intuitively, information consistent with previous world knowledge should be easy to process, comprehenders appear to treat these utterances as anomalous.

However, in the case that the anomaly of these utterances results in some amount of processing difficulty on the part of listeners, it is unclear whether formal models of language comprehension would predict this difficulty, or accurately estimate the presumed difference in their predictability (cf. Smith & Levy, 2013), when compared to the same utterances in an *atypical* context (as in our experiment). The utterances

in question (the *predictable* utterances in the different contexts) represent exactly the same string, and the preceding context is identical stretching over multiple sentences. Thus n-gram models, which can't represent long-distance dependencies, would not predict any difference in predictability (and consequently processing difficulty). More sophisticated models incorporating syntax or semantics, likewise, would not predict a difference, as there are no meaningful differences in syntactic structure, and semantic models wouldn't have access to the relevant event-based information that distinguishes the utterances.

Models of event sequences, which estimate *event* (vs. string) probability, may be able to estimate differences in predictability (and, consequently, processing) between utterances describing script-congruent and script-incongruent events. However, the general prediction of such models is that the more congruent an event is with an invoked script, the more predictable utterances describing that event should be. There is no principled way, within this framework, to divide events into different grades of predictability, such that utterances describing *predictable* events are easier to process than those describing *not-so-predictable* events, yet those describing *very predictable* events incur difficulty. In light of this, we suggest that to predict any difference between the two utterances, formal models of language comprehension would need to incorporate some form of pragmatic reasoning.

It may therefore be fruitful to further investigate these effects using online measures such as eye-tracking or self-paced reading. As typical speaker behavior is to elide utterances containing very predictable or easily inferable information, the local predictability of uninformative utterances (at the utterance end) should be comparatively low. Online measures such as reading times have been argued to reflect the predictability of linguistic events, rather than the ease of integrating information into one's mental model of a situation (Smith & Levy, 2013). Further, context-dependent pragmatic inferences are often presumed to incur a processing cost (Sedivy, 2007). In contrast, semantic integration accounts would predict no particular difficulty in processing information consistent with one's prior beliefs about a typical event sequence (Hagoort et al., 2009). Ease of semantic integration is typically highly correlated with predictability, making their effects difficult to disentangle (Smith & Levy, 2013), and this is a rare case where the two are at odds. Online measures alone, however, provide little information regarding how speakers interpret, or make use of, such utterances, which this study addresses.

However, as Walker (1993) points out, informationally redundant utterances may not be so anomalous, or rare, as might appear at first glance. Most theories predicting their scarcity assume a) unlimited working memory on the part of listeners, and b) ability to easily or quickly generate all inferences entailed by the discourse. They can serve narrative function, and are, for example, often used to draw attention to salient aspects of a discourse, or to showcase evidence for

certain beliefs. Further, as shown in this study, they can be reinterpreted by listeners to reflect new or useful information, even when not contextually supported. Overall, the inferences drawn from apparently redundant utterances may also be of a wider variety than tested here. However, uninformative utterances seem to share the property of necessarily provoking inferences about what is meant aside from the literal meaning, in order to remain felicitous. Where no obvious non-literal interpretation is available, comprehenders may simply assume that the speaker is uncooperative, having production difficulty, or has unconventional speaking patterns. It should should be noted, that in the context of a grocery trip, an 'uninformative' segment like *John paid the cashier*, followed by *with euros instead of dollars*, would not be so anomalous. The literal string itself is not redundant here, as it's part of a larger utterance that contributes previously unknown/uninferred information.

There are several additional avenues for follow-up. Although readers, with some consistency, obtained the interpretation that explicit mention of an overly predictable event implied its atypicality, it is possible that some readers reached other interpretations about what the speaker intended. Further, it's unclear whether those participants who didn't draw the intended inference were drawing another inference, detecting no abnormality in the utterance, or detecting an abnormality but not attempting to resolve it. Directly querying participants regarding non-literal interpretation of an utterance may, however, encourage the drawing of inferences they might not otherwise have been inclined towards. Similarly, the types of questions we used could have encouraged readers to preferentially look for, and make the specific inferences we intended them to draw. In that case, however, we would also have expected to see changes in activity typicality ratings in conditions where none were expected – which we don't.

Summarizing, this study shows that comprehenders attempt to compensate for the uninformativeness of redundant event mentions by assigning a more 'informative' pragmatic meaning, thus increasing their utility. Listener expectations regarding the information content of speaker utterances are consistent with theories of language processing and production, such as Uniform Information Density (Jaeger, 2010), as well as standard pragmatic accounts of what constitutes 'cooperative' speaker behavior (Grice, 1975). This study provides what is to our knowledge the first experimental account of how speakers interpret utterances that are informationally redundant given knowledge of stereotyped event sequences, and contributes to what is still a small body of experimental literature on context-sensitive conversational implicatures. Further, it points to potential problems for estimating utterance predictability in formal models of language comprehension.

Acknowledgments

References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495-527.

Baker, R., Gill, A., & Cassell, J. (2008). Reactive redundancy and listener comprehension in direction-giving. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue* (p. 37-45). Columbus, Ohio.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.

Cohen, P. R. (1978). *On knowing what to say: planning speech acts*. Doctoral Dissertation, University of Toronto.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (p. 41-58). New York: Academic Press.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (p. 819-836). Cambridge, MA: MIT Press.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23-62.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19*. Cambridge, MA: MIT Press.

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in childrens on-line reference resolution. *Psychological Science*, *13*, 329-336.

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2009). Insights into knowledge representation: The influence of amodal and perceptual variables on event knowledge retrieval from memory. *Cognitive Science*, *33*(7), 1252-1266.

Regneri, M., Koller, A., & Pinkal, M. (2010). Learning script knowledge with web experiments. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (p. 979-988).

Sedivy, J. C. (2007). Implicature during real time conversation: A view from language processing research. *Philosophy Compass*, *2*(3), 475-496.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Walker, M. A. (1993). *Informational redundancy and resource bounds in dialogue*. Doctoral Dissertation, University of Pennsylvania, Philadelphia, PA.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386397.