GREEBLES greeble greeb. On reduction in speech and gesture in repeated references

Hoetjes, M.W.; Koolen, R.M.F.; Goudbeek, M.B.; Krahmer, E.J.; Swerts, M.G.J.

Published in:
Proceedings of the 33rd Annual Conference of the Cognitive Society

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

GREEBLES Greeble greeb
On Reduction in Speech and Gesture in Repeated References

Marieke Hoetjes (m.w.hoetjes@uvt.nl)

Ruud Koolen (r.m.f.koolen@uvt.nl)

Martijn Goudbeek (m.b.goudbeek@uvt.nl)

Emiel Krahmer (e.j.krahmer@uvt.nl)

Marc Swerts (m.g.j.swerts@uvt.nl)

Tilburg centre for Cognition and Communication (TiCC), Faculty of Humanities, Tilburg University,
PO Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

Previous research has shown that predictable information in speech is often reduced. The present study aims to find out more about reduction with regard to speech but mainly whether reduction also occurs in speech-accompanying gestures. To this end, a director-matcher task was set up in which speakers of Dutch took part. In this task the director had to refer to the same abstract object several times. The repeated references thus obtained were analysed for their reduction in speech and gesture. Speech results show that the number of attributes is reduced and gesture results show that gestures are reduced with regard to the number of hands that are used, their size and their precision. Implications for existing gesture models are discussed.

Keywords: speech; gesture; reduction; repeated references

Introduction

Reduction in repeated references

When speaking, people often produce referring expressions to describe objects in the world, mostly in order to identify a target referent to an addressee. For example, a speaker might point out someone to an addressee by saying "that tall guy with the glasses" while producing a pointing gesture. The interaction between the speaker and the addressee is multimodal: the speaker may use both auditory (speech) and visual cues (gestures) to ensure that the addressee knows what he or she is talking about.

In a conversation, it is likely that a speaker refers to the same object more than once, leading to the production of repeated referring expressions. From previous research, it is known that these repeated references lead to reduction in speech (e.g. Bard et al., 2000; Brennan & Clark, 1996; Lam & Watson, in press), while reduction in gesture has remained unexplored.

The Uniform Information Density (UID) hypothesis (Frank & Jaeger, 2008; Jaeger, 2010; see also e.g. Aylett & Turk, 2004) states that reduction in language production serves to optimize the interaction between the speaker and the addressee. More specifically, the UID hypothesis predicts that speakers are flexible in the sense that they lengthen the elements of an utterance with a relatively high information density, and shorten the elements with a low information density, which are less important for successful communication. This process makes the amount of information that is transmitted more uniform and optimal for both speaker and addressee. In this way, the UID hypothesis is in line with Grice's (1975) Maxim of Quantity: "Make your contribution as informative as is required (for the current purpose of the exchange)".

Arguably, reduction in repeated references follows a similar pattern: it could be the case that when speakers repeatedly refer to an object, they tend to only reproduce those auditory and visual cues of the initial referring expression that help the addressee to easily identify the target referent, while the less informative cues are omitted. In short, this implies that speakers reduce predictable information, but do not reduce important information.

Reduction in speech

Consistent with the UID hypothesis, previous research has indeed shown that repeated references lead to reduction in *speech* in at least two ways. Firstly, several studies have found that repeated references contain fewer words than initial references to the same target (e.g. Clark & Wilkes-Gibbs, 1986). Brennan and Clark (1996) explain this in terms of people establishing conceptual pacts, which they define as temporary agreements between speaker and addressee about how a target is conceptualized. These conceptual pacts are usually determined by the common ground between the speaker and the addressee. Since these conceptual pacts are not established all at once, but during a series of successive references, as Brennan and Clark (1996) argue, speakers often shorten their repeated references to simpler and more efficient expressions (e.g. from an initial

reference where an object is described as "the purple one with the curly bit at the end that looks a bit like a snake" to a repeated reference where the same object is simply described as "the snake").

Secondly, among others, Aylett and Turk (2004) and Bard et al. (2000) have revealed that repeated references are also reduced acoustically. In their paper, Bard et al. (2000) measured the intelligibility of initial and repeated target descriptions, where they regarded utterances as highly intelligible when the addressee could easily identify the target. Their results showed that repeated references were less intelligible than initial ones, because these were uttered less clearly, even when there was no mutual visibility between speaker and addressee. In line with this, Lam and Watson (in press) found a similar effect for reduction in word duration.

Reduction in gesture

Considering the UID hypothesis (Frank & Jaeger, 2008; Jaeger, 2010) that predictable information is reduced, and given that communication does not only consist of verbal but also of nonverbal aspects, one may wonder whether a similar reduction process as discussed above also takes place for *gestures* produced in repeated references.

In gesture research, there is general consensus that speech and gesture are closely related (Kendon, 1980, 2004; McNeill, 1992). Most gesture models are based on the language production model by Levelt (1989) and agree that speech and gesture have a shared origin. The difference between the different gesture models lies in the point at which the two communication streams (speech and gesture) part. This means that different gesture models predict different results with regard to gesture reduction in repeated references. The gesture model as proposed by De Ruiter (2000), for example, where speech and gesture are planned together, would predict that speech and gesture *can* be reduced together. The gesture model as proposed by McNeill (1992), where there is an even closer link between speech and gesture and where gestures are seen as a direct "window into the mind", might predict that speech and gesture *have* to be reduced together.

Presently, as far as we are aware, no studies have been done on *gesture* reduction in repeated references. Earlier work on *speech* reduction has revealed that (a) repeated references are shorter than initial ones, and (b) that repeated words are less intelligible. Taking into account the close relationship between speech and gesture, we hypothesize that repeated references are accompanied by fewer gestures, and also that these gestures are less 'precise' than those accompanying initial references. A research question is to what extent gesture reduction during repeated references is similar to speech reduction in repeated references.

Even though we are not aware of any studies looking at gestures during repeated references, some circumstantial evidence for our hypothesis can be derived from studies looking at gesture and common ground. In these studies, common ground was established in repeated story narrations. Results from most of these studies (e.g. de Ruiter, Bangerter, & Dings, in press; Holler & Wilkin, 2009; Jacobs & Garnham, 2007) show gesture reduction due to common ground but results are mainly based on reduction in the *quantity* of the gestures. Reduction in gesture *quality* has presently not received a lot of attention, with the exception of two studies. One study, by Holler and Stevens (2007), found that speakers are less likely to represent specific information (in this case of size) in gesture when there is common ground between interlocutors. Gerwing and Bavelas (2004) found that, overall, gestures produced when there was common ground were less complex, precise and informative than gestures produced when there was no common ground. However, in this study gestures were not studied pairwise but overall across story narrations.

Several open questions from above mentioned previous research can be answered by studying gesture reduction in repeated references. Firstly, gesture reduction in repeated references can tell us more about the details of the relationship between gesture and speech. Secondly, since previous research has found reduction in *speech* in repeated references, it is interesting to see whether a similar process takes place for *gesture*. Furthermore, the previous work on gesture reduction is only based on common ground, not on repeated references, and has focused on gesture *quantity*. Presently, no consensus has been reached on how to study *qualitative* reduction of gestures.

The present study

In the present study we will look at speech and gesture reduction in repeated references. Since there is no consensus yet on how to study qualitative gesture reduction, one of the aims of the present study is to develop a methodology with which to objectively quantify qualitative gesture reduction. We expect that gestures will become reduced in number and in appearance, or, in other words, that there will be quantitative and qualitative gesture reduction in repeated references. We also want to compare reduction in speech with reduction in gesture.

To study these aspects of speech and gesture, we conducted a director-matcher task and a perception test, details of which will be discussed below.

Experiment I: Data collection and analysis

Method

In order to study reduction in repeated references a data set was created which consisted of video recordings of participants taking part in a director matcher task. In this task, the director had to describe an abstract figure (a "Greeble") to the matcher in such a way that the matcher could identify the correct figure from a range of similar

looking figures. In the stimuli, there were several figures that had to be described two or three times, leading to repeated references to the same item.

Participants

Participants were 106 undergraduate students (31 male, 75 female, age range 18-29 years old, *M* = 21 years and 7 months), who took part in pairs as partial fulfilment of course credits. From these pairs, data from 5 pairs was left out because there were technical problems or because the participants had not understood the experiment, leading to a data set consisting of data from 48 pairs of participants (48 directors and 48 matchers).

Stimuli

The stimulus material consisted of pictures of Greebles[1], which are intended to be hard to describe, small yellow objects, initially designed so as to share characteristics with human faces. These Greebles vary in terms of their main body shapes ("Samar", "Galli", "Radok", "Tasio"), their gender ("Plok", "Glip"), the different types of protrusions that they have ("Boges", "Quiff", "Dunth") and in terms of the shapes and sizes of these protrusions (see Figure 1 for an example Greeble and see Gauthier & Tarr (1997) for a more detailed description of the Greebles and their properties).

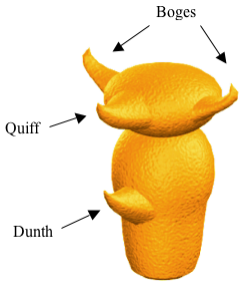

Figure 1. Example Greeble, in this case with the main body shape "Tasio" and of the gender "Glip".

The Greebles were selected for their abstractness, their similarity to each other, and because they differed mainly with regard to their shape, making it likely that the director would give detailed target descriptions and that many gestures would be used. In order to avoid associations with small puppets (which would make the pictures less abstract), the Greebles were turned upside down.

Two picture grids containing 16 Greebles were used. Each picture grid was used for 15 trials, which made a total of 30 trials. In each trial, there was one target object (marked by a red line), which was surrounded by 15 distractor objects. An example of a picture grid can be seen in Figure 2.

The order in which the participants were presented with the picture grids was counterbalanced over participants. The crucial manipulation in the task was that several Greebles had to be described repeatedly: in each of the picture grids, two Greebles had to be described twice, and two Greebles had to be described three times. Repeated references to the same object were never straight after one another.

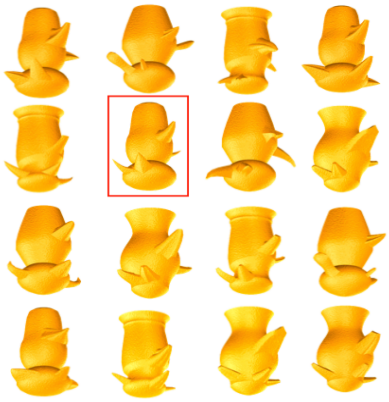

Figure 2. Example of one of the picture grids presented to the director. The object with the red square surrounding it is the target object of that particular trial.

Setup

The experiment was performed in a lab, where the director and the matcher were seated at a table opposite each other (see Figure 3 for the setup).

Figure 3. Setup of the experiment, matcher sits on the left and director sits on the right.

Both participants were filmed during the experiment: one camera was positioned behind the matcher (filming the director) and another camera was positioned to the side of the director (filming the entire setup). The participants were given written instructions and had the opportunity to ask questions, after which the experiment started. The *director* was presented with the trials on a computer screen, and was asked to provide a description of the target in such a way that it could be distinguished from the 15 distractor objects. The *matcher* had a box filled with 16 cards in front of her, which was not visible to the director. The cards in the matcher's box showed the same objects as on the director's screen, but these objects were ordered differently for the director and the matcher. Based on the director's target description, the matcher had to pick the corresponding card

[1] Images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University. URL: http://www.tarrlab.org/

from the box in front of her. Once the correct card was found, the director went on to the next trial. The matcher was allowed to ask for more information or to repeat information that had already been given but there was no free conversation between the director and the matcher. After 15 trials, the director was shown a second grid containing 16 new objects, and the matcher was presented with a new box filled with stacks of cards of these objects.

Data analysis

Speech and gesture analysis has taken place for the first and third references to the objects that had to be described three times The statistical procedure consisted of repeated measures ANOVAs.

Speech analysis We looked at aspects that would tell us more about reduction in speech, such as speech duration, number of words and number of attributes. The number of attributes shows how repeated references are reduced in terms of their content. An attribute can be defined as a property that can be assigned to a target referent. When constructing the trials, we made sure that all targets could be distinguished by means of 4 attributes. We designed an annotation scheme containing over 20 attributes that speakers could potentially mention when describing a Greeble. This scheme was based on the basic characteristics of Greebles (main body shape, gender, protrusions) and was expanded with attributes describing all other properties that they can possibly have (mainly concerning the protrusions' shapes, locations and sizes). The attributes in the annotation scheme were based on the participants' descriptions.

Gesture analysis For the gesture analysis we not only looked at the quantity of gestures (in raw numbers and proportional use), but also developed several ways to analyse reduction in the quality of the gestures. For these qualitative analyses we created a subset of the data. This subset consisted of all directors (23 out of 48) who produced at least one gesture for each initial and repeated reference. We took the following qualitative aspects into account. We measured the duration of the stroke (main part of the gesture, see Kendon, 2004 and McNeill, 1992) in seconds. We measured the visibility of the gesture by coding whether the gesture was visible for the addressee (code 1) or not (code 0). We measured the amount of repetition within the gesture by counting the number of repeated strokes. This was done because the data set contained many gestures with repetitive strokes. We measured the size of the gesture by coding whether the gesture was produced with a finger (code 1), the hand (code 2), the forearm (code 3) or the entire arm (code 4). We measured the proportion of gestures that were produced using two hands. Finally, we studied whether the gesture was perceived to be more or less precise in a perception test (discussed in Experiment II).

Results

Reduction in speech As can be seen in Table 1, compared to initial references, repeated references showed reduction in speech in several ways. Firstly, it was found that it took the participants significantly less time (in seconds) to refer to the target in a repeated reference (M = 27.2, SD = 1.9) compared to when they referred to the target in an initial reference (M = 42.5, SD = 1.5) ($F_{(1,47)}$ = 40.752, p < .001). Note that this effect cannot be explained through a general reduction of descriptions over time, since *non*-repeated references later in the grid were not shortened compared to the initial references. Also, the number of words that speakers used to describe a target, was significantly lower for repeated references (M = 59.8, SD = 5.2) than for initial references (M = 91.9, SD = 3.4) ($F_{(1,47)}$ = 37.283, p < .001). The total number of attributes that were mentioned in initial and repeated references shows how repeated references are reduced in terms of their content. We found that significantly fewer attributes were mentioned in repeated references (M = 7.4, SD = .3) compared to initial references (M = 9.4, SD = .3) ($F_{(1,47)}$ = 40.614, p < .001).

Table 1. Overview of overall mean results for dependent variables in speech, for initial and repeated references.

	Initial (*SD*)	Repeated (*SD*)
Duration ***	42.5 (*1.5*)	27.2 (*1.9*)
Words***	91.9 (*3.4*)	59.8 (*5.2*)
Attributes***	9.4 *(.3)*	7.4 *(.3)*

*** p < .001

Reduction in gesture As can be seen in Table 2, compared to initial references, repeated references showed reduction in gesture in several ways. For the quantitative analyses, results show that speakers produce significantly fewer gestures during repeated references (M = 3.9, SD = 0.42) compared to initial references (M = 5.5, SD = 0.46), ($F_{(1,47)}$ = 22.750, p < .001). Since this could also be due to the reduction in the number of words in repeated references, we also looked at the proportional use of gestures (G/W, number of gestures divided by the number of words) and here we see a significant *increase* in the proportional use of gestures for the repeated references (M = .094, SD = .02) compared to the initial references (M = .067, SD = .02) ($F_{(1,47)}$ = 4.037, p = .05). In other words, the reduction in repeated references is smaller for gestures than for the number of words.

The qualitative analyses revealed that gestures produced during repeated references were smaller ($F_{(1,22)}$ = 5.419, p < .05) than the gestures produced during initial references. Gestures produced in repeated references were relatively more often produced with the smaller articulators (finger and/or hand) whereas gestures produced in initial references were relatively more often produced with the larger

articulators (forearm and entire arm). Moreover, we found a reduction in the percentage of two handed gestures, meaning that even though overall, most gestures were produced with two hands, there were fewer two-handed gestures in repeated references (M = .075) than in initial references (M = 0.85) ($F_{(1,22)}$ = 8.209, p < .01). The other variables taken into account (gesture duration, gesture visibility, repetition within the gesture) were found to be not significant.

Table 2. Overview of overall mean results for dependent variables in gesture, for initial and repeated references.

	Initial (*SD*)	Repeated (*SD*)
Number of gestures***	5.5 (*.46*)	3.9 (*.42*)
Gesture proportion (G/W)*	.067 (*.02*)	.094 (*.02*)
Gesture duration (sec)	1.6 (*.09*)	1.4 (*.14*)
Gesture visibility (0-1)	.93 (*.03*)	.93 (*.03*)
Gesture repetition	.37 (*.09*)	.38 (*.08*)
Gesture size (1-4)*	3.7 (*.07*)	3.5 (*.11*)
Two-handed gestures (%)**	0.85	0.75

*** p < .001, ** p < .01, * p < .05

Experiment II: Perception

As part of the qualitative gesture analysis, the aim was to measure gesture precision. It is difficult to define objective measures with which to code gestures' precision, therefore a perception test was run where participants had to judge gestures' precision.

Participants

Twenty three participants (20 female, 3 male, age range 19-65 years old, M = 26 years old) who did not take part in the Greebles experiment took part in the perception test. All participants were academically qualified but had no previous knowledge about the Greebles experiment.

Stimuli

For this perception test, 66 pairs of video clips were selected from the qualitative data subset obtained in Experiment I. These pairs of video clips contained two gestures (one in each video clip), produced by the same director, showing the main shape of the same stimulus item. One video clip showed a gesture produced in an initial description of a stimulus item, the other video clip showed a gesture produced during a repeated description of the same stimulus item. The order in which the initial and repeated gestures were presented was counterbalanced.

Setup

The participants were presented with the 66 pairs of video clips. For each pair of video clips, they had to decide in which video clip they thought the gesture was "the most precise". Participants were not aware that the gestures were produced in descriptions to the same object. No instructions were given with regard to what they should base this precision judgment on. The test was a forced choice test and participants were asked to go with their fist intuition. Each gesture that was considered to be the most precise received one point.

Results

A two-tailed t-test showed that overall, gestures produced during repeated references (M = 9.23, SD = 6.9) were considered to be significantly less precise than the gestures produced during initial references (M = 13.77, SD = 6.9), ($t_{(65)}$ = 2.658, p < .05). The proportional results, where not just the gesture that received the most points but the gesture's exact scores were taken into account (each gesture that was considered to be the most precise received one point from each participant), also show that gestures produced in an initial reference were considered to be more precise (M = 0.60, SD = .30) than the gestures produced in repeated references (M = 0.40, SD = .30). The perception test has shown that gestures produced in repeated references are seen as significantly less precise than those produced in initial references.

Discussion

The results of the two experiments presented in this paper have shown that speakers reduce their speech and gestures in several ways when they repeatedly refer to the same target. These results are in line with our expectations, and fit the Uniform Information Density hypothesis (Frank & Jaeger, 2008; Jaeger, 2010). As we have seen in the introduction of this paper, the UID hypothesis holds that predictable information is reduced in language production, and that elements with low information density are likely to be omitted from an utterance. Our results suggest that the same goes for repeated references: speakers reduce repeated descriptions in terms of both their speech and their gestures.

With regard to the *speech* that speakers produce, our results show that reduction in repeated referring expressions occurs in terms of the number of attributes that speakers mention when describing a target: significantly fewer attributes were mentioned in repeated references as compared to initial ones, showing that repeated references are less informative than initial ones. Besides this, our results also reconfirmed the findings of, among others, Clark and Wilkes-Gibbs (1986), by revealing that repeated referring expressions contain fewer words than initial ones.

With regard to *gesture* we can now claim two things. Firstly, we found that repeated gestures are reduced with regard to their overall number, their size, the percentage of two-handed gestures, and their perceived precision. These results show that gestures, as previous findings found for speech, are reduced in repeated references. There are parallels with reduction in speech, both in the reduction in the overall number of gestures, which can be compared to

the reduction in number of words and creation of conceptual pacts in speech, and in the reduction in gesture precision, which is arguably comparable to acoustic reduction in speech. Secondly, the results also showed that even though overall, speakers use fewer gestures in repeated references, the proportional use of gesture *increases*. In short, there is more reduction in words than in gestures. This finding, that speech and gesture are not necessarily reduced together, can be interpreted as evidence for the fact that speech and gesture production are, at least to some extent, separate processes.

The results showed an overall reduction in the duration of the descriptions. However, the reduction due to repetition was larger than the overall reduction, showing that the effect was not simply due to the director's overall description practise but mainly due to the communicative situation. The fact that the communicative situation seems to play such a large part in the experiment also suggests that the gestures were mainly produced for the listener. However, it could still be the case that the gestures serve a speaker internal purpose as well. In the present study almost all gestures were visible for the addressee. To be able to say more about whether the changes in gesture that we found are due to speaker internal or communicative reasons, we will conduct a follow-up study with no mutual visibility between interlocutors. Other future work will include further developing the present methodology with which to quantify the quality of the gestures. Aspects of the gestures that have not yet been taken into account but that might be relevant are, for example, the handshapes used in the gestures and the length of the gesture holds.

Conclusion

In conclusion, in this study on reduction in speech and gesture in repeated references we found for *speech* that repeated references are less informative than initial references. For *gesture* we found that gestures in repeated references are reduced in their overall *quantity* as well as in several aspects of their *quality*. We also found that proportionally, the gesture quantity *increases* for repeated references compared to initial references.

The present study contributes to previous research in two major ways. Firstly, the finding that speech and gesture are not necessarily reduced together indicates that their reduction might to some extent be two independent processes. Secondly, reduction in gesture in repeated references has not been studied before and to date there has been no consensus on a methodology with which to study qualitative gesture reduction. The present study is an important step in this direction.

Acknowledgments

We would like to thank Joost Driessen for help in annotating the data and Bas Roset for technical support. We received financial support from The Netherlands Organization for Scientific Research, via a Vici grant (NWO grant 27770007), which is gratefully acknowledged.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

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

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, 22*(6), 1482-1493.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture*. Cambridge: Cambridge University Press.

de Ruiter, J. P., Bangerter, A., & Dings, P. (in press). The interplay between gesture and speech in the production of referring expressions: Investigating the trade-off hypothesis. *Topics in Cognitive Science*.

Frank, A. F., & Jaeger, T. F. (2008). *Speaking rationally: Uniform information density as an optimal strategy for language production*. Paper presented at the 30th Annual Meeting of the Cognitive Society (CogSci), Washington.

Gauthier, I., & Tarr, M. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*, 1673-1682.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*, 157-195.

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

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Psychology, 26*(1), 4-27.

Holler, J., & Wilkin, K. (2009). Communicating common ground: how mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267-289.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291-303.

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

Kendon, A. (1980). Gesture and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal communication and language*. The Hague: Mouton.

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.

Lam, T. Q., & Watson, D. G. (in press). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition*.

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

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