

CHAPTER 1
INTRODUCTION

Towards the end of my undergraduate program in linguistics I attended a particularly memorable lecture. During that lecture, it was revealed to me that in all my years of studying language I had only been looking at half the picture. It had escaped my notice that when people talk they typically move their hands and arms around as well, or *gesture*. Speakers link spaces in front of their bodies to topics in the discourse, point to elements that are not physically present, enact scenes they are describing, and use their hands in countless other ways that are extremely closely connected to the speech they are simultaneously producing. Because of this tight connection, many researchers have come to believe that gesture is part of the language system—that language is not just speech or sign, but speech or sign *plus gesture*. After that lecture, I became such a researcher.

Once language has been broadened to include gesture, discovering exactly how the two modalities coordinate becomes a major theoretical issue. In the past few decades there have been many significant discoveries on this front. Research on the timing of speech and gesture has shown that the meaningful portion of a gesture typically occurs with the most heavily-stressed segment of an utterance (Nobe, 2000). This pattern indicates that both speech prosody and gesture are jointly highlighting a particular portion of an utterance. Gestures also appear to express meaning differently depending on the grammatical structure of the speaker's language (Duncan, 1996; Kita & Özyürek, 2003; McNeill & Duncan, 2000). That is, differences in how languages lexically or grammatically encode information are correlated with differences in the information that tends to appear in gesture.

Despite an ever-expanding body of research, it nonetheless remains difficult to formulate precise predictions about how speech and gesture will pattern in a given situation. This is because language is massively shaped by context. Any individual utterance-gesture pair is the outcome of a series of choices on the part of the language user. One can see this by comparing the behavior of two people describing the same event: while there will be plenty of similarities in their descriptions, there will also be enormous variability. In describing the cognitive processes involved in language production, one can generalize across the similarities one observes, but one should also attempt to account for the variability. The work presented here is an effort to do just this.

In this dissertation, I select a particular source of variability, *speaker focus*, and attempt to describe its impact on language production. By speaker focus, I mean

the tendency to attend to a particular element while formulating an utterance. Two people describing the same event may attend to different elements while planning their descriptions, which leads to corresponding differences in their speech and gesture. I use a series of experiments to demonstrate that predictions can be made about how a narrator will encode her mental representation in speech and gesture depending on which element she has been induced to focus on. In carrying out these manipulations, I am attempting to recreate in a controlled way the kinds of choices speakers naturally make during language production.

In the experiments to be described, I show participants a cartoon motion event involving two elements, then manipulate which they encode as the syntactic subject of the utterance that describes that event.¹ The use of a cartoon motion event is motivated by the fact that there is a large literature on how people speak and gesture when describing motion, making it possible to formulate predictions. But why do I manipulate the choice of a sentence subject? This question will be discussed in more detail in chapter 2, but I will sketch the basic premise here. First of all, manipulating the choice of sentence subject is a way of getting two groups of participants to produce very minimal and systematic differences in their descriptions of an event. English has fairly rigid SVO (subject-verb-object) word order—the typical English sentence is thus a subject followed by a predicate. As a consequence, once a sentence subject has been chosen, the rest of the sentence has to be arranged to accommodate that choice. The end result is that two groups can be induced to produce different

¹ Throughout this dissertation, I will use *participant* to refer to people who took part in the experiments, thus *subject* is always going to refer to a grammatical element. Nonetheless, I will try to use terminology like *utterance subject*, *syntactic subject* or *sentence subject* (used interchangeably) to avoid confusion.

sentence structures that have the same propositional content. For example, *the mouse eats the cheese* and *the cheese is eaten by the mouse* have different subjects and syntactic structures but are semantically identical. The manipulation of sentence subject can thus be seen as a means to an end, the end being different sentences that say basically the same thing. As it happens, I also have a reason for specifically wanting sentence subjects to differ across my groups. To wit, the element that a person's attention is focused on while she is formulating her utterance tends to be encoded as the subject of that utterance (Tomlin, 1985, 1995, 1997). In other words, speaker focus and sentence subject are correlated. A difference in sentence subject thus provides strong evidence for an underlying difference in speaker focus.

I use two paradigms to achieve the manipulation of sentence subject. The first, presented as Experiment 1, uses *syntactic priming*. Syntactic priming is an exploitation of the fact that hearing a syntactic construction makes us more likely to use it ourselves. By priming my participants with constructions in which one or another element appears as sentence subject, I can induce participants to produce those structures themselves. The second paradigm, presented as Experiment 2, is a visual manipulation of attention. Directing attention to an element has been shown to increase the chance that it will be encoded as the subject of a sentence (Tomlin, 1985, 1995, 1997). For both these experiments, predictable differences emerge in participants' gestures depending on which element appears as the subject of their utterances.

Having successfully shown that speech and gesture can pattern in predictable ways for one motion event, the second component of the dissertation asks whether the

pattern can be replicated with other events. I apply the priming and attention manipulations to three additional events, presented as Experiment 3. These events differ according to certain parameters which appear to affect the success of the manipulation, namely agency, animacy, and the visuo-spatial properties of the event.

The organization of the dissertation is as follows. The next chapter will set the stage for the experiments to be described. I first discuss previous work on how speech and gesture pattern in cross-linguistic studies of motion events, in order to explain why I expect to see particular verbal and gestural behaviors in my data. I then provide justification for manipulating the syntactic subject of my participants' utterances.

This will require me to elaborate on the relationship between speaker focus, sentence subject and *sentence topic*, and *discourse topic*. By discussing these constructs, I hope to clarify how a speaker's verbal and gestural output are influenced by general discourse structuring principles, by the syntactic structures available for encoding a particular mental representation, and by the speaker's use of attention during utterance formulation.

Chapters 3 and 4 present the results of the first two manipulations of speaker focus (Experiments 1 and 2). Chapter 5 presents the results of the third experiment, which applies the same two paradigm to three additional events. Chapter 6 discusses the implications of my results for models of language production.