How Predictability and Givenness Produce Activation, and Acoustic Reduction

Jason Kahn

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology.

Chapel Hill

2011

Approved by: Dr. Jennifer E. Arnold

Dr. Peter Gordon

Dr. Neil Mulligan

Abstract

Speakers tend to use reduced pronunciation, e.g. shorter duration, when words are previously mentioned (i.e. “given”; Fowler & Housum, 1987), or predictable in context (Bell et al, 2009; Watson et al., 2007). Existing accounts of this phenomenon underspecify whether both giveness and predictability make independent contributions, and say little about the underlying cognitive mechanism. I propose and test the Activation Reduction Hypothesis (ARH), which states that any stimulus that activates representations used for language production should elicit reduced pronunciations. This unites givenness and predictability in a single plausible psychological mechanism, and makes novel predictions, which I tested in three experiments. The first experiment shows that linguistic stimuli elicit more reduction than non-linguistic stimuli, which also elicit reduction. The second shows that linguistic stimuli elicit reduction in the absence of strong predictability, suggesting a role for sheer activation. The third attempts to isolate this reduction at the conceptual level of representation, but shows little supporting evidence.

Acknowledgements

Many thanks to Kellen Carpenter, Samantha Handel, Leighanne McGill, Giulia Pancani, Jennifer Tate, and Alyssa Ventimiglia, who served as addressees and helped run experiments. Thanks to Florian Jaeger for comments on an earlier version of this paper.

Funded by NSF grant BCS-0745627 to Jennifer E. Arnold.

Table of Contents

List of Figures

List of Tables

Introduction

Suppose a psycholinguistics teacher wants to give a lecture on phoneme restoration. In doing so, she will have to use the word "phoneme" several times. On the first utterance, she will most likely pronounce it quite intelligibly, with extended duration, well-defined vowel targets, and relatively high pitch (Brown, 1983; Lindblom 1990; Terken 1985). Acoustically speaking, her pronunciation contains a great deal of information (Jaeger, 2006; Aylett & Turk 2004). The word "phoneme" itself, as well as its associated concept, is new and relatively unpredictable at the beginning of the lecture. But because any good lecture has a consistent topic and contiguous structure, she will most likely talk about phonemes again, and use the same word form, "phoneme." On these subsequent pronunciations, the acoustic properties of the word will change: the duration and pitch will decrease (Bell et al., 2003, 2009; Gregory et al., 1999, Jurafsky et al., 2001), the vowel targets will become less distinct (Aylett & Turk 2006), among other things, and the word's overall intelligibility will go down. Both the concept [phoneme] and the word "phoneme" have become given and predictable by the second time she mentions them. The current paper investigates why her pronunciation changes between the first and the second mention by focusing on how the predictability and givenness of phoneme and "phoneme" change.

Variation in linguistic form comes about at least in part due to two properties of a portion of an utterance: predictability and givenness. On one hand, the teacher's use of "phoneme" changes the *givenness* of both the word and its associated concept. The word enters the conversation, or discourse, with *new* status, but thereafter the word and the concept are both *given* as a part of the shared discourse. Similarly, the use of “phoneme” increases the *predictability* of both the word “phoneme” and the associated concept. Social expectation of the structure of a lecture suggests that the teacher will continue discussing phonemes. Many studies demonstrate that speakers tend to use reduced expressions to refer to given entities (Brown 1983; Fowler & Housom 1987; Fowler 1988) as well as predictable entities (Bell, Brenier, Gregory, Girand & Jurafsky, 2009; Gregory, Raymond, Bell, Fosler-Lussier & Jurafsky, 1999; Lieberman 1963; Jurafsky, Bell, Gregory & Raymond, 2001). Yet the mechanism that produces these effects remains obscure. The experiments reported here thus have two goals: the proposal of a cognitive mechanism for acoustic form variation, and the investigation of the possibility of separating predictability and givenness as determiners of that variation.

Givenness and predictability do not fully explain reduction themselves in part because the two co-occur. A referent’s becoming given increases its likelihood of reuse in a conversation (Arnold, 1998, in press; Givón, 1983). Indeed, some scholars define givenness **in terms of** predictability (see Prince, 1981, p. 226). Similarly, an entity cannot become predictable without also becoming given in some sense. Givenness can come about in several ways. For example, a word can be mentioned linguistically, or its concept can be evoked through other means, such as environmental salience. Either or both of these types of givenness might also make an entity predictable. Observing reduction on a particular word does not allow one to reliably attribute the effect to givenness or predictability, even with considerable knowledge about the context, linguistic or not.

Although many existing explanations for reduction include measures that attempt to account for the effects of both givenness and predictability, a fully satisfactory account will have to address several unanswered questions: 1) What is the common cognitive substrate or mechanism, if any, for the effects of givenness and predictability? 2) To what extent do givenness and predictability operate independently? 3) What sorts of information "count" for givenness and predictability, and therefore for reduction? Several proposals have been made in answer to question 1 (see Bell et al., 2009), and some recent work has addressed both question 2 (Lam & Watson 2010) and the portion of question 3 that deals with what I will call predictability source monitoring (Aylett & Turk 2006; Bell et al., 2009; Levy & Jaeger, 2007; Tily & Piantodosi, 2009, *inter alia*). [1]

As a step toward uniting and answering these three questions under a common framework, I propose the hypothesis that activation of the conceptual and linguistic representations associated with a word makes it a good candidate for acoustic reduction. I call this the Activation Reduction Hypothesis (ARH). More specifically, I claim that any stimulus or processing that activates representations associated with a to-be-produced word should lead to some measurable amount of reduction. Givenness and predictability in any form fall under this definition as potentiators of activation, as do anticipatory processing and simple associative priming as forms of facilitation. Importantly, this proposal effectively unites givenness and predictability under a single cognitive substrate. It also predicts that distinct sources of information should lead to additive reduction to the extent that they activate distinct representations in the production system, or increase the activation of already-activated representations. This graded-reduction aspect of the

[1] By predictability source monitoring, I mean the establishment of relationships among the various possible predictability measures associated with reduction. I will return to this issue momentarily.

proposal will allow me to provide at least partial answers to questions 2 and 3, with the answer to question 1 as a guide.

Linguistic forms can vary along several dimensions, including lexical form (Arnold 1998), morphological form (Frank & Jaeger 2007), vowel quality (Aylett & Turk 2006), word duration (both shortening and lengthening; e.g. Bell et al., 2009), consonant inclusion/deletion (Jurafsky et al., 2001), and others (for a review, see Bell et al., 2009). I focus on word-level durational reduction, which lends itself well to simple but powerful manipulations. The experiments described below hold syntactic, morphological, and phonological information constant while manipulating predictability and givenness. This approach has two primary advantages: holding higher-level linguistic variables constant allows for a direct comparison of the effects of only the factors of interest on reduction, without compromising the naturalness of the task too far (see Aylett & Turk, 2004 for a good example of an analysis restricted to particular syntactic environments); also, the use of duration as a dependent variable allows for indirect comparisons to a robust literature on these effects.

Here I report the results of three instruction-giving tasks designed to test the predictions of the ARH. In the first two tasks, one of a pair of participants (the speaker) told the other about the movement of objects on the speaker's computer screen, where I manipulated the givenness and predictability of these objects. In the third task, a single participant issued instructions with the ostensible goal of having them used in a future experiment, because the listener in the first two tasks was deemed irrelevant.[2] Overall, the experimental designs focused on creating tokens of the same utterance under different

[2] See below, in Experiment 2, for a discussion of a separate experiment that justifies this exclusion.

predictability and givenness conditions, to allow for a comparison of acoustic duration. Before explaining the manipulations in more detail, however, I will discuss the nature of acoustic reduction, theoretical accounts of the relationship between givenness, predictability, and acoustic reduction, and proposals for the underlying mechanisms.

What is acoustic reduction?

Speakers' pronunciation of a word falls along a continuum, acoustically speaking. Acoustically reduced (as contrasted with acoustically prominent) expressions tend to have shorter duration, lower pitch and vowel-target articulation, less pitch variation, and/or incomplete segmentation (Aylett & Turk 2004, 2006; Bell et al. 2003, 2009; Jurafsky et al. 2001, *inter alia*). The reduction of pitch height and pitch movement may be related to choices about whether to accent a word or not, but it has been demonstrated that variations on acoustic clarity do not depend entirely on accenting (Aylett & Turk, 2004; Bard & Aylett, 2004; Watson , Arnold, & Tanenhaus, 2007). All of these prosodic metrics contribute to an utterance's overall intelligibility (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon & Newlands, 2000), i.e. how easy it is for a listener to identify the word out of context. Intelligibility, and particularly duration, is a relative phenomenon. A speaker may hypoarticulate when producing an expression for his colleagues, and hyperarticulate when producing the same expression for his students (see Lindblom, 1990).

Givenness effects on reduction

There is substantial evidence that speakers' references to given information tend to have reduced forms, relative to references to new information. This research builds on theories of discourse structure, which emphasize the difference between **given** information, i.e. information that has already been mentioned or otherwise evoked, and **new**, or unmentioned information (e.g., Chafe 1976; 1994; Prince 1981; 1992). The given/new distinction is one way of characterizing the **information status** of discourse elements, which itself correlates with linguistic phenomena such as accenting (Venditti & Hirschberg, 2003), pronoun use (Chafe, 1976; Gundel, Hirschberg & Zacharski, 1993), and word order (Birner & Ward, 1998). Canonically, anything mentioned in the discourse counts as given information, precisely because it has been mentioned and is presumably in the participants' memory of the conversation. The teacher's introduction of phonemes to her students makes that information given, arguably for the remainder of the lecture. The word form "phoneme" thus becomes a good candidate for reduction, *ceteris paribus*.

Analyses of word pronunciations in running speech have demonstrated that references to previously-mentioned information tend to be shorter and less intelligible than references to new information (Brown 1983; Fowler 1988; Fowler & Housum 1987). For example, in a map description task, speakers who have already produced a reference to a landmark on the map tend to reduce its form on the second production (Bard & Aylett 2004, with data from Anderson, Bader, Bard & Boyle, 1991). This phenomenon appears to have more to do with the act of referring than the linguistic form itself, as non-co-referential repeated mentions do not get reduced (Bard, Lowe &

Altmann 1989; Fowler 1988). Comprehension experiments also demonstrate that listeners tend to prefer reduced forms for given information, and tend to associate reduced forms with given information (Arnold 2008; Baumann & Grice 2006; Bauman & Hadelich 2003; Dahan, Tanenhaus & Chambers 2002).

Many accounts of givenness have a strong textual or spoken dialogue focus, which allows them to make useful predictions about linguistic phenomena such as anaphoric reference (Gordon, Grosz, & Gillom, 1993), accent placement (Halliday, 1963), and choice of referring expression (Ariel, 1990; Gundel et al., 1993). Givenness leads to a greater number of pronouns as opposed to full lexical noun phrases for purposes of reference (Arnold, 2008), to deaccenting as opposed to accenting in prosody (Halliday, 1963), and to less explicitly specified referring expressions (Ariel, 1990; Gundel et al., 1993). Broadly speaking, the importance of givenness for these accounts lies in its contribution to the meaning or form (especially the prosodic form) of a text or utterance. The reverse also holds: accent placement and form, as well as choice of referring expression, lead to a partial determination of givenness, and therefore meaning (Pierrehumbert & Hirschberg, 1990; though see Terken & Hirschberg, 1994).

Other accounts have more to say about how givenness affects cognition and vice-versa. Chafe (1976, 1994; see also Ariel, 1990; Gundel et al., 1993), for example, distinguishes among types of givenness by referring to mental states: *new* information lies outside of the speaker's current attention, and is inactive in his or her mind, while *given* information can be either fully *active* or only *partially active*, depending on how closely the speaker attends to the information. The ARH builds on these views, but focuses instead on the relationship between givenness and linguistic *processing*. It explicitly states that givenness activates the processes and representations involved in language production, which has acoustic reduction as a behavioral outcome. Of course, I do not mean to imply that meaning-based accounts cannot inform processing.

Givenness accounts also make a distinction between linguistic and non-linguistic givenness. Most researchers claim that information can become given via explicit linguistic mention (i.e. linguistic givenness) and situational evocation (i.e. non-linguistic givenness; Clark & Marshall 1981; Lambrecht 1994; Prince 1981, 1992), where situationally evoked information becomes salient in an environment through whatever means (e.g. visual salience, task relevance). Whether these two types of givenness differ in their behavioral outcomes is a question left implicit in these theories. A framing pair of answers comes from comprehension work (Baumann & Hadelich, 1994) and from corpus work (Bard & Anderson, 1994). Baumann & Hadelich show that listeners prefer one type of accenting for linguistically given information and another type for non-linguistically given information. Bard & Anderson show that an entity's physical presence suffices to elicit a significant reduction in intelligibility.[3] Listeners distinguish between the two types of givenness, while speakers appear not to do.

The ARH takes inquiry about the contribution of givenness to reduction a step further. Among other things, it provides a framework for asking about the relationship between linguistic and non-linguistic givenness. The ARH predicts (along with other accounts) that non-linguistic information should matter for reduction, but it departs from earlier work in specifying both how and why. A discourse-status account, where an entity either counts as given or not, would say that both linguistic and non-linguistic

[3] For more recent experimental work, see Fukumura, Van Gompel & Pickering, 2010.

information lead to givenness, because both place an entity within a discourse as given. A processing-based account like the ARH says that non-linguistic information should lead to reduction to the extent that it activates representations associated with language production, which means that explicit evocation of an entity should elicit more reduction relative to a simple appearance as part of an environment. The ARH also predicts that linguistic and non-linguistic givenness should differ to the extent that the former activates more processes or representations than the latter, with the observable effect of a difference in reduction. It thus creates a principled, psychologically plausible explanation for why givenness creates reduction.

Predictability effects on reduction

As with givenness, speakers tend to use reduced forms to refer to predictable information. For example, a classic experiment by Lieberman (1963) demonstrated that the word “nine” has a longer and clearer articulation in the context “The next number is nine” than in a context that makes it more predictable, e.g., “A stitch in time saves nine”. *A priori*, a frequent word like *cat* is more predictable than an infrequent one like *cataphract*, and more frequent words tend to have shorter durations (Zipf 1929) and shorter onset-to-speak latencies (Griffin & Bock 1998). *N-gram* relationships among words (i.e. how likely a word is to follow or precede a particular *n* words) also appear to partially control reduction (Gregory et al., 1999, Jurafsky et al., 2001, Bell et al., 2003, 2009, Aylett & Turk 2004, 2006).

Predictability-based accounts of reduction use probabilistic relationships among linguistic variables to predict form variation. At the word form level, the predictability of a **word** in the context of the surrounding words leads to acoustic reduction (e.g., Bell et al., 2009; Jurafsky et al., 2001, 2003).. Reduction due to probability may occur via the activation of the word form representations, at least at the lemma level, and possibly phonological, and articulatory levels as well (see Jurafsky et al., 2001). These effects are not limited to predictability in the traditional sense, where words become probable because of prior events; words are also reduced when the **following** context makes them predictable (e.g., Bell et al., 2009). This is likely to occur because utterances are at least partially pre-planned (Ford, 1982). These effects suggest that co-occurrence probability matters because planning is facilitated when utterances are statistically consistent (i.e. redundant).

Accounts that focus on probabilistic explanations of reduction primarily use data from corpus analyses. The large data sets in these studies permit the effective calculation of several different types of probabilistic relations. The somewhat straightforward *n*-gram relationship mentioned above, and the conditional probability and mutual information measures derived from it, control a significant portion of the variation in durational reduction (Jurafsky et al., 2001, *inter alia*). A slightly more general approach calculates the **information** conveyed by a particular probabilistic relationship, where “information” here has a technical meaning (Jaeger, 2006; Levy & Jaeger 2007; Shannon, 1948). Information-based accounts claim that speakers attempt to modulate the rate at which they transmit information (i.e. the probabilistic relationships among the variables in their speech) by reducing words that have high probability (and therefore low information) in a context. Although the probability calculations in these accounts differ, the differences do not bear on my purpose here, and I collapse across them below.

In addition to the word level, the predictability of a **referent** also appears to affect acoustic reduction. Arnold (1998) demonstrated that certain discourse contexts correspond with an increased likelihood that a particular entity will be mentioned, irrespective of the lexical form used to refer (e.g., pronoun, name, description, etc.; see also Tily & Piantodosi, 2009). These are the same discourse contexts that are often proposed to make entities accessible in the discourse, e.g. grammatical subject position, or recency of mention (Arnold, 1998; 2008, in press). Arnold labeled this kind of predictability **expectancy,** which is critically different from word predictability in that it calculates the predictability of the referential event, not the word. This kind of discourse context has also been shown to influence acoustic reduction; for example, recency of mention (Bard & Aylett, 2004; Fowler & Housom, 1987), or grammatical subjecthood / parallel function (Terken & Hirschberg, 1994). Watson, Arnold, and Tanenhaus (2008) also found that task constraints on referential predictability influenced word durations.

As it did with accounts of givenness, the ARH unites and extends the predictions of predictability-based accounts. The probabilistic relationships within a language should create activation for representations associated information in a discourse (see Balota et al. 1989 for an account along these lines). Stronger relationships should create more activation, and thus lead to greater reduction. The ARH also makes three novel predictions. First, expectancy (i.e., referential predictability) should create activation of discourse-level representations of an entity, facilitating production at the conceptual level. Second, predictability should have an effect independent from givenness to the extent that the two sources of information activate different representations. Third, predictability should arise from relevant non-linguistic information, in addition to linguistic information. Predictability-based accounts can plausibly extend to include these predictions (that is, nothing I say here contradicts existing claims), but to my knowledge the ARH is the first hypothesis to make this set of predictions.

Speaker-centric versus Addressee-Oriented Reduction

Another candidate explanation for reduction, addressee-based accounts, focuses on the speaker-listener interaction (Bard et al., 2000; Halliday, 1963; Lindblom, 1990). Addressees can identify given and predictable referents more easily than new or unpredictable ones, with the consequence that the speaker can produce a less explicit linguistic form and still expect the addressee to identify the correct referent. In this case, the production system might include representations of the addressee's knowledge, perhaps in parallel with each of the levels of representation outlined above. Speakers and addressees often share a considerable amount of information during a conversation, such that speaker-centric and addressee-oriented processes make similar predictions in many situations. Indeed, in these experiments both speakers and addressees receive the same information about the objects on a trial, and I will not attempt to distinguish between these two routes to reduction (but see Bard & Aylett, 2004; Bard et al., 2000).

Testing the Activation Reduction Hypothesis

Many models of language production suggest that production occurs in several stages, as illustrated in Figure 1 (derived from Levelt, 1996). Activation at any of these levels has the potential to affect the production of the utterance, either via facilitation, anticipatory processing, or a monitoring mechanism that assesses the state of activation of the system. The ARH combines givenness and probabilistic information into a single metric that permits a comparison of their relative contribution. It also places no restrictions on the source of activation, such that linguistic and non-linguistic information, as well as givenness and predictability, can all play a role.

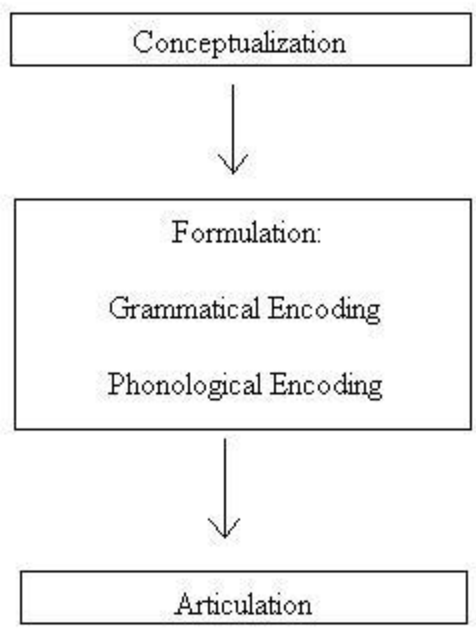

Figure 1 – The stages of speech planning (adapted from Levelt 1989)

Experiment 1 tests three major predictions of the ARH: 1) Does referential-event predictability and/or givenness 'count' for the purposes of reduction? 2) Does non-linguistic information 'count' for the purposes of reduction? 3) Do linguistic and non-linguistic information elicit different degrees of reduction?

Experiment 2 asks two questions that further clarify the relationship between givenness and predictability: 4) Does givenness operate independently of predictability for the purpose of producing reduction? 5) Does the same relationship between givenness and predictability hold for linguistic and non-linguistic information?

Experiment 3 sharpens the focus of question 4 and uses eye-tracking to address a final pair of questions: 6) Is simply conceptual givenness sufficient for reduction, or is predictability necessary also? 7) Can eye-tracking provide a finer-grained measure of how speakers use non-linguistic information for reduction?

I will begin by explaining how Experiment 1 addresses the first three questions, before describing the experiment itself and its results. I will revisit Experiment 2 and 3's questions in light of the results from each other experiment.

The first major question addressed here comes from the generality of the ARH: can referential events serve as a locus for givenness and/or predictability information, and thereby elicit reduction? Prior corpus work suggests an affirmative answer (Arnold, 1998; Tily & Piantodosi, 2009), as does experimental work (Lam & Watson, 2010; Watson et al., 2007). Although corpus analyses can only suggest correlations, not establish causal relationships, Arnold (1998) and Tily & Piantodosi (2009) establish that referential events affect form reduction. Information that increases the likelihood of a referential event also increases the likelihood that a speaker will use a pronoun as opposed to a full lexical noun phrase. Although experimental studies do not permit infallible inferences about causal relationships, they do permit direct comparisons between situations where entities become good targets for reduction *as referential events* and situations where no such felicity holds. Both Lam & Watson (2010) and Watson et al. (2007) provide good evidence that task-relevant information that increases the likelihood of reference to an object decreases the duration of the head noun in a reference to that object, and potentially other parts of the phrase. The experiments reported here are all designed to replicate and potentially extend these findings. In particular, Experiment 1 provides evidence that givenness affects not only the duration of the word for the referent, but also the latency to begin speaking and other words in the carrier phrase.

The question of whether non-linguistic information counts for reduction matters because the answer seems obvious – of course non-linguistic information should count. Unfortunately, much linguistic theory and psycholinguistic experimentation that relates givenness and predictability to reduction focuses on the text itself. Corpora, for example, typically contain very little information about the environment the participants occupy, and so cannot use non-linguistic information predictively. The notable exception of the HCRC map task (Anderson et al., 1991) includes information about the state of each participant's map, which allows for comparisons between visually-available and entirely new material (see Bard & Anderson, 1994). Indeed, they find a difference between visually-present and entirely new material, which confirms the intuition that non-linguistic information matters.

The task described below takes answering the second question one step further. In the map task, all landmarks had the potential to serve as referents. Their sheer presence therefore counted as non-linguistic information. The tasks below set out to determine whether providing priming non-linguistic information would elicit reduction or not. The explicit activation of the representations associated with the non-linguistically primed entities aligns more closely with my hypothesis. It also allows for a direct comparison with linguistic priming information. My tasks and the map task exhibit two different types of naturalness, both of which are relevant for understanding how non-linguistic information gets integrated into processing. The map task featured a rich environment with several objects, which became relevant as the subjects chose to use them in descriptive utterances. In my tasks, objects have relevance because of task demands, where their sheer presence does not count unless they get primed. Both types of conversational topic structure occur in everyday conversations.

The third question Experiment 1 addresses is whether linguistic and non-linguistic information differ in the amount of reduction they elicit. Taken strictly, meaning-based theories of givenness do not make a principled distinction between the two – given counts as given, regardless of the source. Existing processing-based accounts do not put forth any explicit claims about the possibility of behavioral differences based on types of givenness, but could plausibly be extended to include such a distinction. The mental states associated with partially active versus active states on Chafe's theory, for example, might map onto the difference between linguistic and non-linguistic information, to the extent that one activates an entity more than the other. The ARH makes this claim explicitly, and says that linguistic and non-linguistic information should differ in the amount of reduction they produce because they activate a different proportion of the representations associated with language production processes.

In order to answer these three questions, pairs of participants perform a variant of a referential communication task developed by Watson and Arnold (2005; see also Lam and Watson, 2010). They were provided with two types of priming information: linguistic givenness (e.g., hearing the word "airplane") and non-linguistic givenness (e.g., seeing an airplane). Speakers issued instructions to listeners about the movements of three different pictured objects from an array, as in Figure 2. For example, after seeing a picture of an airplane shrinking, the speaker would tell the listener “the airplane shrinks.” The listener would then make the airplane shrink on his or her screen. A second and then a third object movement and an instruction would follow, with the same movement-utterance-execution sequence performed for a different object. The critical manipulation was whether the objects were a) linguistically given (by having their name previously mentioned), b) visually given (i.e. non-linguistically given, by having their picture flashed), or c) completely new (as a control condition). Both experimental conditions also made the given entities predictable (c.f. Arnold 1998), but as a referential event.

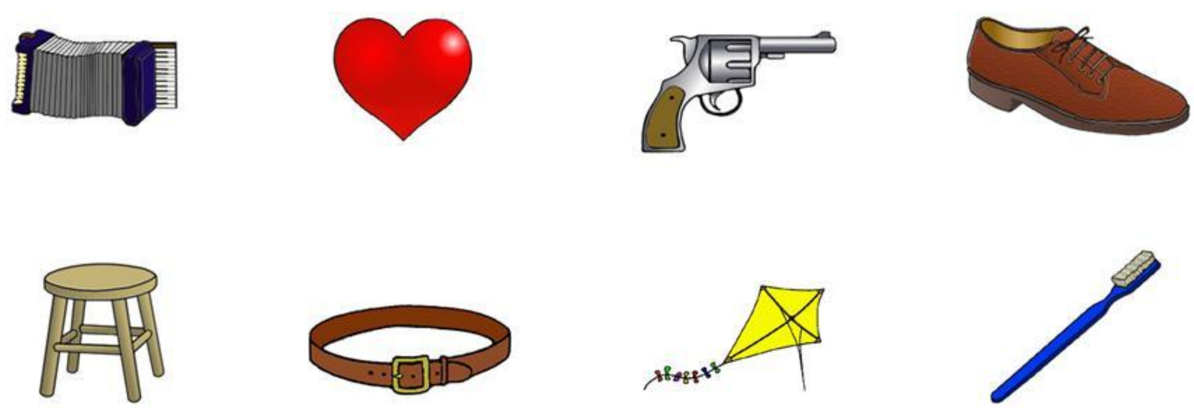

Figure 2 – An Example Object Array

The set of manipulations in Experiment 1 thus allows me to answer three of my questions. The priming information makes entities given, and reference to particular entities predictable, such that any observed reduction should revolve around referential events. Two comparisons between conditions, outlined in Figure 3 would provide answers to the second and third questions: 1) a comparison between utterances in the control and non-linguistic conditions will address whether non-linguistic information counts for durational reduction; 2) a comparison between the linguistic and non-linguistic conditions will address whether linguistic and non-linguistic information differ in producing reduction.

Any durational difference between the control and non-linguistic conditions should reflect the non-linguistic information's having primed representations that go unprimed in the control condition, which would allow for ease of activation or preplanning. The comparison between the linguistic and non-linguistic conditions is slightly more complex. The priming information provided in both of these conditions has two functions: it makes the three target objects given, either linguistically or non-linguistically, and it makes them predictable as referents in the task. Thus, the only theoretically relevant contrast between an utterance in each condition is the type of givenness, because at each of the first, second, and third instructions, both informational manipulations make the first, second, and third object predictable to the same degree. The important contrast for answering the third question is thus between linguistic and non-linguistic third instructions. Any observable difference in durational reduction between the two should in principle come only from the difference in givenness, because predictability is held constant.

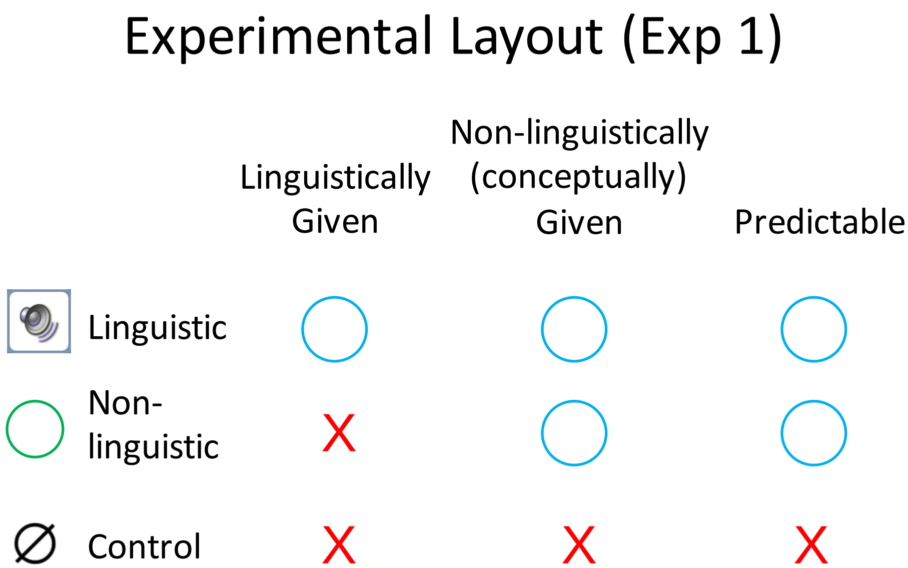

Figure 3 - The layout of Experiment 1, with conditions listed on the left and factors of interest listed across the top. Relevant comparisons are between 1) the Non-linguistic and control conditions, which differ on givenness and predictability 2) the Linguistic and Non-linguistic condition, which differ on linguistic givenness

Experiment 1: Linguistic vs. Non-Linguistic Givenness

Participants. 15 female undergraduates participated in an instruction-giving cooperation experiment as speakers for course credit. Females were not specifically recruited; however, in order to facilitate a pitch analysis (not reported here), females were asked to serve as speakers whenever possible. 3 were excluded due to equipment failure, for a total of 12 speakers. 9 additional undergraduates (2 female) participated in the experiment as listeners for course credit. The remaining 6 speakers were paired with a lab assistant.

Materials and Design. 221 colorized versions of the Snodgrass & Vanderwart (1980) line drawings served as the stimuli (Rossion & Pourtois, 2001). These drawings were normed by native French speakers for imageability, visual complexity, familiarity, as well as name agreement in French, although this latter measure was not used here. The appendix lists the names of the 221 experimental items. The English name of each picture, as reported in Rossion and Pourtois (2001), was recorded by the first author for use in the linguistic priming condition. All recordings were made on a Marantz professional solid-state recorder over an Audio Technica professional microphone headset**,** using citation format (i.e. no distinguishing inflection and clear pronunciation).

On each trial, participants saw an array of 8 objects. Three of these objects were the **target objects,** i.e. those that were to be mentioned by the speaker in spoken instructions (e.g., *The windmill shrinks, the tomato rotates left, the candle fades).* The third (last) object to be mentioned was termed the **experimental target** (here: the candle). A total of 24 objects served as experimental targets. These were divided into three 8-item lists (see Figure 4), which had approximately equivalent average values of log frequency, syllables, imageability, familiarity, and visual complexity. The experimental targets were the trials used in the primary analysis. The other 7 objects in the array were selected at random by the computer program from the remaining 197 objects. From these 7 objects, the program randomly selected two to serve as the target objects in the first two instructions for each trial, which are called the **non-experimental targets**.

Ling	Ctrl	Non-ling	Ctrl	Non-ling	Ling	Non-ling	Ling	Ctrl
Set 1	Set 2	Set 3	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3

Set 1	Set 2	Set 3
train	bee	shoe
ring	knife	couch
toothbrush	horse	sheep
nail	harp	bow
lamp	bell	needle
basket	broom	bus
snake	fly	whistle
finger	clock	deer

Figure 4- An example of the list design in Experiment 1. The three sets of words were randomized within each condition-set pairing. No condition could appear twice in a row, and a condition could repeat only after all three conditions had been seen the same number of times. (Ctrl = control condition, Ling = linguistic condition, Non-ling = non-linguistic condition)

The critical manipulation in the experiment was whether within the trial the target objects were linguistically given, non-linguistically given, or not given at all. On *linguistic* trials, the target objects' names were spoken aloud in random order, at a volume both participants could hear, before the instruction-giving portion of the trial. On *non-linguistic* trials, all three target objects flashed twice, simultaneously, on both participants' screens, before the instruction-giving portion of the trial. On *control* trials, nothing distinguished the target objects from the other objects, all of which stayed on the screen for 2.5 seconds (the approximate time required to say the objects' names, or make the objects flash) before the instruction-giving portion of the trial.

The experimental target (i.e., the last to be mentioned) had a special status because its status *as* the experimental target was completely predictable at this point in a trial, but only in the two priming conditions. In the control condition it was not especially predictable, beyond the expectation that one of the pictured objects would be mentioned.

Following Arnold & Griffin (2007) and Watson, Arnold & Tanenhaus (2007), the experiments used a modified Latin Square design, in which each participant saw each item in all three conditions. This allowed us to directly compare the production of a word across givenness conditions for each participant. The 24 experimental target items were organized into list of 8 items each, and each list occurred three times, once in each condition. Thus, the first three blocks represented a standard Latin Square design, where each item was only viewed once. A separate analysis of these blocks alone was performed to test that the effects are not the result of repeated experience with an item over the experiment.

There were a total of nine blocks, which included three of each condition type (e.g. three linguistic, three non-linguistic, three control). Six sequences in this format were created from Latin Square combinations of the target item sets and the conditions. See Figure 4 for a schematic example of a sequence. Each item list (and thus each item) appeared in each condition, for a total of three appearances. Conditions were organized into triads, such that each condition had to appear once before another could appear twice, and no condition could appear twice in a row. Two participants were run on each of the six sequences. Each speaker participated in 72 trials, for a total of 216 uttered instructions.

A customized Python script written by the first author controlled the experiment. The speaker's stimuli were presented on a laptop PC, and the listener's were presented on a desktop PC**.**

Procedure. Both the speaker and the listener sat in front of a computer, at a distance that permitted them to communicate but not see the other's screen. Both participants saw task instructions on their own screen. Participants were instructed that they would see arrays of objects, and were given an example of such an array (Figure 2). The speaker was told that she would be issuing instructions about the objects' actions to her partner, who would mimic those actions on his or her computer. Examples of the possible actions (expand, rotate left, rotate right, shrink, and fade) were then shown with random stimuli, and a sentence frame was provided. For example, the speaker's instructions might have said "You might see [an airplane rotating left], and you might say 'The airplane rotates left.'" The listener would have simultaneously seen "You might hear 'The airplane rotates left,' and you would click the airplane and then the 'Rotates Left' button."

The participants were told that the experiment would be divided into sets of trials, and that before each set they would know whether they would receive linguistic, non-linguistic, or no information about the target items. 6 practice trials followed these instructions, 2 from each information condition. The objects on these practice trials were randomly selected, and did not include any experimental items.

Analysis. Utterances were analyzed individually using Praat software (Boersma & Weenink 2009). I examined the visual waveform and listened to the sound file to identify the onset of each trial (indicated by a beep that co-occurred with each action onset), as well as the onset and offset of the determiner *the,* the object word, and the action phrase. Two measures of duration are presented here: latency to begin speaking (time from the onset of the beep to the onset of speech), and the duration of the noun. The latency to speak for each instruction reflects planning time, where shorter latencies indicate less complex and/or demanding plans. The duration of the object word most accurately reflects the effect of the manipulation, i.e. whether the primed object had in fact become a good candidate for reduction. Utterances were excluded if they were disfluent in any way (e.g. pauses longer than approximately a second, false starts, disfluent 'the') or if the speaker used the wrong object or action word.

For both of the duration measures, multilevel models were constructed to assess the contribution of givenness and predictability. All quantitative predictors were centered to reduce collinearity. Collinearity is a measure of the degree to which two variables in a regression correlate with each other, in addition to the relationship they share with the dependent variable. The primary problem with collinearity is that it increases the standard error associated with the affected variables. Reducing it thus cleans up the relationship between the independent variables and the dependent variable. The logarithm of each duration was taken to increase its linear relationship with the predictors. Subject-wise outlier exclusion was not necessary after this transformation. Models were constructed with the lmer function from R's lme4 package, which uses a Restricted Maximum Likelihood Estimator to obtain parameter estimates. Markov Chain Monte Carlo (MCMC) simulations were performed with R's pvals.fnc function, which runs 10,000 simulations to estimate the High Posterior Density intervals. Essentially these quantify the chance a parameter differs from zero, based on the observed distribution of possible parameters, and is functionally equivalent to a standard *z*-test.

All results reported here reflect the entire dataset. A potential problem with this design is that each object was mentioned 3 times, raising potential practice effects. Nevertheless, analyses on the first three blocks (i.e., the first mention of each object) produced the same pattern of effects, except where noted. See http://www.unc.edu/~jmkahn/models.html for reports of these additional models.

Two analyses were performed. The primary analysis examined the critical third instruction across the three conditions (linguistic, non-linguistic, control). This analysis afforded two comparisons: (1) between the linguistic and nonlinguistic conditions, to quantify any contribution of linguistic givenness over and above non-linguistic givenness and referential predictability, and (2) between the non-linguistic and control conditions, to test whether reduction occurred for non-linguistically given/predictable referents, even in the absence of linguistic givenness. The primary analysis was restricted to only the experimental target items, which were fully determined as of the third mention, with predictability held constant across the two given conditions. Nevertheless, analysis of all three instructions generally revealed the same pattern as the 3rd-instruction analysis.

A secondary analysis attempted to isolate the effect of predictability by comparing reduction between the three instructions, with givenness held constant. Condition was included as a predictor in each model, in order to test the simultaneous effect of givenness and predictability. Because the effects of predictability should only show in the linguistic and non-linguistic conditions, the analysis was restricted to this data. Conversely, this analysis necessarily included data from the first and second instructions, in order to determine the cumulative effect of predictability, but the objects in these instructions were not experimentally controlled to any degree. These results should thus be interpreted with caution.

For both segments reported here, I first constructed a baseline model to identify the significant control variables. Each model included random intercepts for subject and item, as well as all of the following control variables: log frequency (based on the original Snodgrass & Vanderwart (1980) English names), imageability, visual complexity, and familiarity of the experiment target (taken from Rossion & Purtois (2001)), the number of syllables of the experimental target and of the action word, and the trial number (to control for practice effects). Each of the control variables was centered on its mean. Only variables that approached significance in this baseline analysis (a *t value* approximately greater than 1.5) were included in the final models. Final models consisted of the significant control variables as well as the predictors of interest.

In order to fully assess the effects of the experimental conditions, two sets of final models were fit in the primary analysis. Both included a pair of Helmert contrasts that tested for significant differences between conditions. The first pair compared the linguistic condition to the non-linguistic condition, and the two experimental conditions to the control condition. The second pair compared the non-linguistic condition to the control condition and the linguistic condition to the other two conditions. The final models in the secondary analysis also included a similar set of Helmert contrasts that compared the first and second mentions, as well as the third versus a combination of the first and second, and then another set that compared second and third mentions, as well as the first to a combination of the second and third. Only the final models are reported here. Reported effects are based on the *p* values that emerged from the MCMC simulations of the final models. Only *p* values are reported, because the parameter estimates correspond with log-unit changes in the dependent variables. For reports of the baseline models, and full information about the parameter estimates of the final models, please visit http://www.unc.edu/~jmkahn/models/.

Predictable Targets		Onset to Speak	*the*	Object Word	Action Word	Total
Linguistic	1st	1118	94	419	778	2409
	2nd	1066	96	416	795	2374
	3rd	940	96	381	764	2181
Non-linguistic	1st	1160	100	448	809	2517
	2nd	1090	105	448	803	2446
	3rd	932	102	406	795	2235
Control	1st	1441	105	451	795	2793
	2nd	1314	107	440	805	2666
	3rd	1269	102	416	789	2576

Table 1. Experiment 1 - Mean durations of each segment from all blocks, in milliseconds.

Results: Primary Analysis. The primary analysis on the object word is reported first. As shown in Table 1, and graphically in Figures 5 and 6, the linguistic and non-linguistic conditions both exhibited numerical reduction relative to the control condition, with more reduction in the linguistic than non-linguistic conditions.

Two models tested the significance of these patterns, as described above. The first model included Helmert contrasts between the non-linguistic condition and the control condition, and also between the linguistic condition and the other two conditions. This model revealed that both contrasts were significant. The parameter estimates indicated that the non-linguistic condition had a significantly smaller duration than the control condition (t = 3.07, p < .002, pMCMC < .002), and the linguistic condition had a significantly smaller duration than the other conditions (t = -5.2, p <. 0.0001, pMCMC < 0.0001). A second model included the opposite set of contrasts, namely between the linguistic and non-linguistic condition, and between the control condition and the other two conditions. Again, both contrasts were again significant. This model's results, coupled with the previous model's, indicated a three-way contrast between the conditions. The estimate for the difference between the linguistic and non-linguistic condition was significant (t = 2.91, p < .002, pMCMC < .003), as was the estimate for the difference between the control condition and the other conditions (t = 5.2, p < 0.0001, pMCMC < 0.0001).

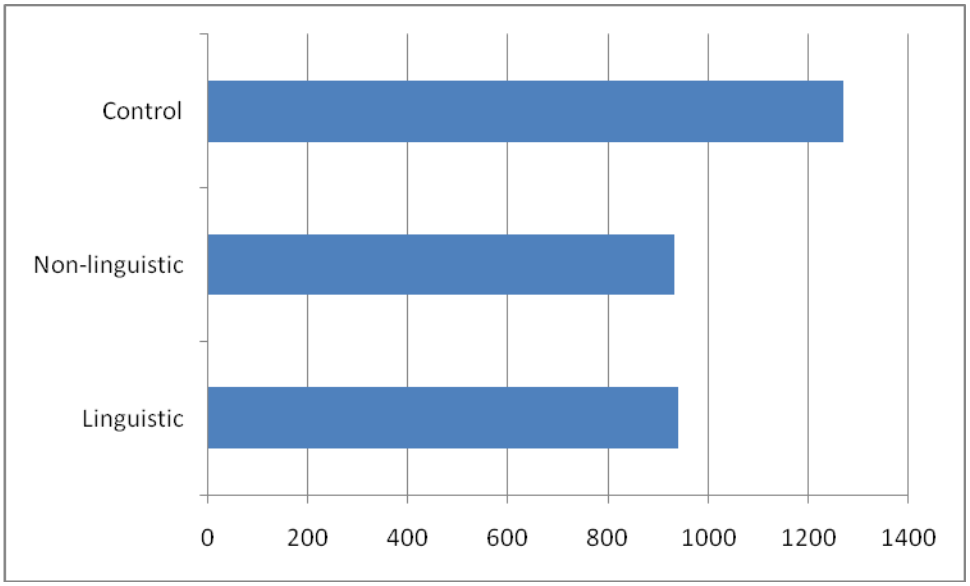

Figure 5- Experiment 1 - Graph of the onset to speak latency of the third instructions, in milliseconds

The onset to speak latency displayed a different pattern. Table 1 shows that the linguistic and non-linguistic condition were both shorter than the control condition, but did not differ from each other. This bore itself out in the statistical analysis, where the difference between the control condition and the other two conditions was significant (t = 17.39, $p < 0.0001$, pMCMC < 0.0001), while the difference between the linguistic and non-linguistic conditions was not (t = -.24, $p > .80$, pMCMC > .79). A notable difference emerged between the first-blocks-only analysis and the full complement of third-mentions. Despite the smaller number of observations, the first-blocks-only analysis revealed a significant difference between the linguistic and non-linguistic condition (t = 2.35, $p < .01$, pMCMC < .01), and a significant difference between the non-linguistic condition and the control condition in the model with the other set of contrasts (t = 2.26, $p < .01$, pMCMC < .01).

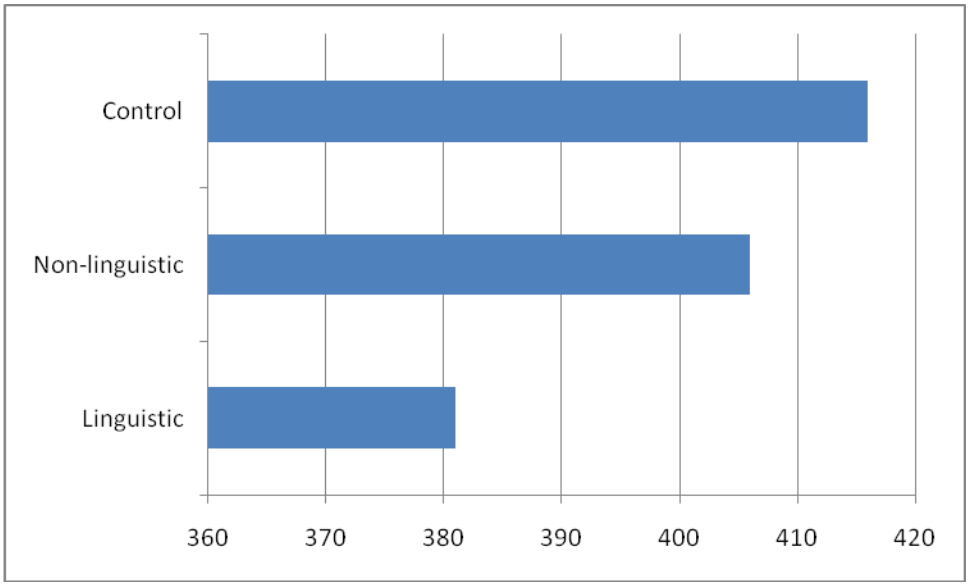

Figure 6 - Experiment 1 - Graph of the object word durations of the third instructions, in milliseconds

Results: Secondary Analysis. As a secondary analysis, I sought evidence for the role of predictability by measuring the duration of the latencies and object words across the three instructions, within each condition. The means for each segment are shown in Table 2, and graphically in Figures 6 and 7. In the two experimental conditions, all three objects were predictable at the beginning of the item, but their order of mention was not known. This means that the target of the first instruction was 33% predictable, the target of the second instruction was 50% predictable, and the target of the third was 100% predictable. If predictability effects map onto acoustic reduction in a gradient fashion, reduction should increase over the three trials. One caveat attends these results, however: the non-experimental targets (i.e. the first and second instruction objects) were not experimentally controlled.

.

Mention	Onset to Speak	*the*	Object Word	Action Word	Total
1st	1259	101	443	789	2592
2nd	1142	102	434	804	2482
3rd	1023	99	396	781	2300

Table 2: Experiment 1 - Means from all blocks for all three mentions

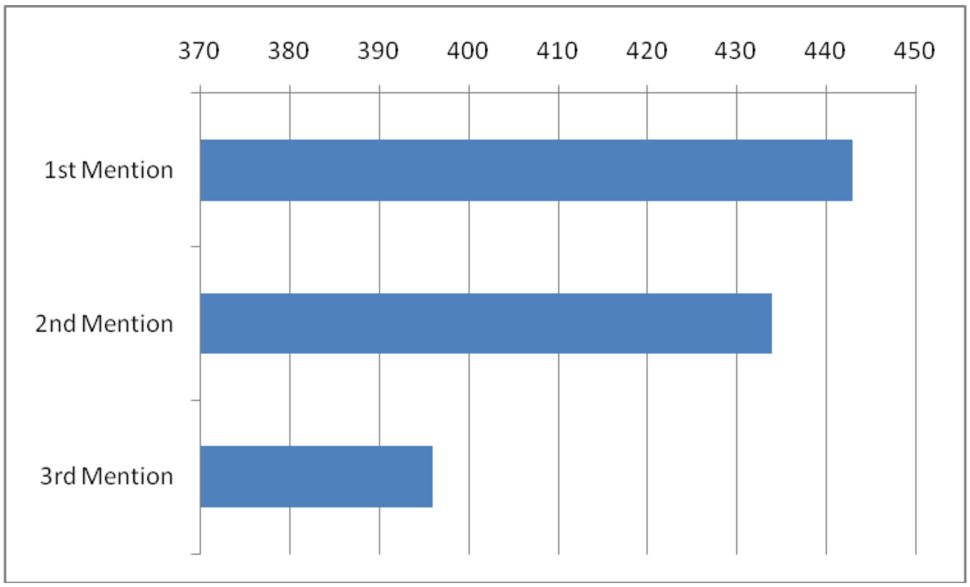

Figure 7– Experiment 1 - Means of the object word duration for all three mentions in milliseconds

The object duration data leads again. The parameter estimate for the difference between the first and second mention was not significant ($t = -.13$, $p > .89$, pMCMC > .9), while the estimate for the difference between the third mention and the other two was marginally significant by t value, and significant by MCMC value ($t = -2.38$, $p < .07$, *p*MCMC < .02). An examination of the means for the object duration reveals that the first and second mentions elicited nearly equivalent durations, while the third mention was shorter. The difference between the third mention and the other two mentions most likely reflects a difference in planning that allowed for greater activation of representations. That is, with a fully predictable object, speakers could begin planning their whole utterance, or portions of it, sooner than with less-predictable objects.

This view is supported by the onset duration data, where second mentions had shorter onsets than first mentions, and third mentions had shorter onsets than second mentions. All of the parameter estimates for the differences between conditions were significant at or beyond the .0001 level. The three-way contrast between the mentions suggests that speakers were using the information in the experimental conditions to begin planning their utterances sooner when that information made the entities increasingly predictable.

A somewhat puzzling result emerges from an analysis of the control condition. Because it included no explicit priming information, no reduction should emerge across mention number, barring a slight decrease due to the fewer number of items remaining after each mention. That is, after the first mention, only 7 of the objects on screen could serve as targets. However, an analysis of mention number revealed a significant difference between the first mention and the other two (t = -3.01, p < 0.0001, pMCMC < 0). No significant differences were found in the object durations, however. The reduced onset time in the second and third mention onset durations probably reflects a greater familiarity with the array, which allowed speakers to disperse their attention among more objects, and thus detect an action and beginning planning for it sooner.

Analysis of Speaking Rate. An additional analysis was performed as a follow-up to the primary analysis to determine the extent to which the effects were driven by overall speaking rate. For these models, the durations of the other segments in the instruction were included. The model for object duration included the onset, article, and action word duration, for example. The test of interest then became whether the original condition contrasts remained significant even after controlling for the effect of speaking rate. Indeed, after including the additional parameters in the model, all of the previously-reported effects remained significant. The manipulation appears to have had an effect over and above speaking rate.

Discussion. The results demonstrated that referential events can serve as a locus for givenness and predictability information, as evidenced by the sheer presence of reduction in this experiment. This allowed for a manipulation of both linguistic and non-linguistic givenness with predictability held constant, and, to a lesser extent, predictability with givenness held constant. Comparisons among givenness conditions revealed that non-linguistic information elicits reduction, and that linguistic information elicits more reduction. This effect appeared in both the onset to speak latency and the object duration, which most likely reflects an effect of both planning and facilitation. Comparisons among the mention numbers (i.e. the first, second, and third instructions) revealed reduction on the onset to speak latency but not the object duration. Keeping in mind that the first and second instruction objects were not experimentally controlled, this tentatively suggests that the predictability manipulation had only the effect of reducing planning time, and not on durational reduction *per se*.

The most general result to emerge from the first experiment, the efficacy of referential events as determiners of reduction, experimentally verifies both theoretical claims (Arnold, 1998, 2001) and findings from corpus studies (Arnold 1998; Tily & Piantodosi, 2009). The design here directly manipulated whether participants construed reference to particular objects as given and/or predictable, as opposed to construing a particular referential *form* as given and/or predictable. The instructions themselves contained very little structure that would have allowed word-level probability measurements to take hold. That is, although they were complete sentences, utterances like "The airplane rotates right" do not vary systematically enough from utterances like "The accordion rotates left" to observe effects of word-level predictability. My point is not to claim that word -level probability is unimportant. Rather, this result establishes that referential events can stand among the various predictability measures used to predict reduction.

The second major result of these manipulations is the effect of non-linguistic givenness on reduction. Both the onset to speak latency and the object duration displayed reduction in the non-linguistic condition, suggesting that this type of information, even though it comes from outside of the discourse itself, counts along many, if not all, of the same dimensions as linguistic information. Non-linguistic information produces reduction most likely because it activates representations and/or facilitates processing at the conceptual level of language production. This facilitation appears to allow both for quicker planning, as evidenced by the reduced onset to speak latencies, and for overall reduced word duration, as evidenced by the effect on the object word. The results persist even after controlling for speech rate.

The result of greatest theoretical interest, however, is the increased reduction in the linguistic condition relative to the non-linguistic condition. This effect appeared on the object duration, but not the onset duration, suggesting that it was not an effect of planning. The inclusion of speech rate as a predictor also did not eliminate the effect. Reduction most likely came about due to the activation of representations and/or facilitation of processing throughout the production system. The difference between the linguistic and non-linguistic condition emerged because linguistic givenness affected a greater proportion of the language production process, which resulted in a greater degree of reduction on the object word.

The predictability manipulation, although the weaker of the two, also produced interesting results. Within the two experimental conditions, predictability had a gradient effect on the onset duration, but not the object duration. This suggests that referential predictability in this task had its primary effect on planning the utterance, while givenness had its primary effect on the object word duration. The second experiment will attempt to further delineate the effects of givenness and predictability.

Experiment 2: The Independence of Givenness and Predictability

Motivation. Experiment 1 used givenness and predictability simultaneously to investigate whether linguistic and non-linguistic givenness exerted different effects on reduction. The design paired these two factors in part to allow the direct comparison of different types of givenness in the presence of strong predictability, but also in order to elicit the strongest reduction effect possible. The factors worked in conjunction, which invites asking whether they can also work separately. This question has two parts: does predictability elicit reduction even in the absence of givenness, and does givenness elicit reduction even in the absence of predictability? Corpus analyses give an affirmative answer to the first part of the question; highly predictable word forms (e.g. Bell et al., 2009) and structural forms (e.g. Levy & Jaeger, 2007) both lead to reduction, even though none of the influential linguistic material appears in the text.

The second question remains unanswered, and is important for understanding the dynamics of reduction for two reasons. Givenness and predictability very frequently co-occur, which makes it possible that one factor simply cashes out in terms of the other. Because predictability has been shown to have an independent effect, the most likely direction for this relationship is that givenness merely makes entities predictable in context, as opposed to having an effect of its own. An independent effect of givenness would thus invite an alternative explanation, perhaps one that related givenness and predictability to some third factor like activation. The second reason comes out of general interest about reduction, but more specifically from the results of Experiment 1. Discourse accounts of givenness do not typically make a distinction between non-linguistic and linguistic givenness, as was found in the first experiment. This result is more readily explained by processing accounts, such as the ARH. If a processing account has something to offer, it leaves open the possibility that non-linguistic and linguistic givenness differ also in their dynamics with respect to predictability. That is, linguistic information may more readily produce activation in the absence of predictability than non-linguistic information, because of the greater proportion of the language production system it primes.

Up to this point, “givenness” has referred to both linguistic and non-linguistic information, and has glossed over the theoretical history of the term. Accounts of givenness often allow non-linguistic information to make entities given in a discourse, but none provides a principled account of why it should, or whether it counts in the same way as linguistic information. In order to distinguish the typical use of “givenness” from the type of information processing the subsequent experiments induce, I will use the term “conceptual priming” to refer to what I have previously called non-linguistic givenness. Similarly, I will use “priming” to refer to situations where either linguistic or non-linguistic “givenness” would have previously been appropriate. The new terminology captures the fact that linguistic and non-linguistic information seem to differ, both intuitively and empirically, based on the first experiment. It also highlights the psychological, as opposed to linguistic, focus of this paper. The manipulations in these experiments prime the representations associated with language production, and those primed representations count as given and/or predictable. Here, the psychological phenomenon takes precedence over the linguistic phenomena.

Experiment 2 investigates two questions in particular: 1) Does priming operate independently of predictability for the purpose of producing reduction? 2) Do non-linguistic and linguistic information differ, as they did in Experiment 1, in their propensity to produce reduction, even in the absence of predictability? Because priming inevitably creates some amount of predictability, especially in a laboratory setting, completely eliminating predictability is not possible. The task is designed to mitigate its effects as much as possible by creating a situation where using the priming information to predict the target on a particular trial will “pay off” infrequently.

Predictions. This experiment tests the prediction that priming should matter even in the absence of strong predictability. According to the Activation-Based Reduction Hypothesis, priming activates the representations that are relevant to word production, which results in facilitation of the production process. Even without strong predictability, some activation should take place. The interpretation offered for the results from Experiment 1 placed the activation due to conceptual priming lower than that due to linguistic priming. Reducing or eliminating the effect of predictability in Experiment Two may thus reduce the detectability of an effect due to conceptual priming. In principle, however, both types of priming should result in reduction if the ARH is correct.

The manipulation also permits asking whether priming affects both word reduction and latency to begin speaking, as it did in Experiment 1. In particular, the results from Experiment 1 suggested that predictability operated primarily on reducing the onset to speak latency, and to the same extent in both the linguistic and non-linguistic conditions. Priming, by contrast, operated primarily on the object duration, and produced a difference between the linguistic and non-linguistic conditions. This suggested that predictability might affect planning while priming affects word duration more directly. Experiment Two will put that conjecture to the test.

Participants. 24 undergraduates participated in an instruction-giving cooperation task very similar to the task in Experiment 1, and were offered course credit or $10 an hour for their participation. 4 were excluded due to technical difficulties, for a total of 20 speakers. A lab assistant served as the listener in this task.

Materials. The same materials, including the colorized line drawings, recordings, and hardware were used as in Experiment 1.The experimental targets were chosen as the

16 two-syllable target words from Experiment 1 that elicited the greatest amount of reduction in the linguistic condition, relative to the control condition. This method was used partly to control for the number of syllables in the object words and partly to give speakers the greatest opportunity to reduce, all while controlling the reduction environment.

Design and Procedure. The design was the same as in Experiment 1, with the following exceptions. Each trial consisted of only one instruction, and either a linguistic or conceptual prime. The prime was either the name of one of the objects on the screen or the flashing of one of the objects on the screen. Only the speaker saw the flashing object, while both the speaker and listener heard the name of the primed object played over the speakers. I will return to this difference in the discussion. The prime itself thus had the possibility of being valid or invalid. A valid prime served as both the prime object and the target object in the trial, such that the speaker either saw an object flash, or heard an object named, and then saw that same object complete an action. An invalid prime did not serve as the target object on a trial, such that the speaker saw an object flash or heard an object named, and then some other object performed an action. After the prime and the object's action, the speaker issued an instruction, which the listener executed.

The entire experiment consisted of 128 trials, with a break at the halfway mark. With 16 experimental items, four lists were created that crossed whether each participant saw items according to the 2x2 combination of linguistic vs. conceptual prime and valid vs. invalid prime. 8 experimental items appeared in the first half of the experiment, with half of these appearing as linguistically primed, and the other half as conceptually primed. The other 8 experimental items appeared in the second half of the experiment, again as half linguistically primed and half conceptually primed. Half of the experimental items that appeared in the first half of the experiment were validly primed, while the other half were invalidly primed. Across all lists, each experimental item appeared an equal number of times in each of the cells formed by the 2x2 condition pairings. Each participant saw items twice, once in each half of the experiment. If an item appeared in one condition pairing in the first half (e.g. linguistically and validly primed), it appeared in the exact opposite pairing in the second half (i.e. non-linguistically and invalidly primed).

The only valid primes participants saw were for experimental items. The large number of invalid trials ($7/8^{th}$ of each block) made pre-activation of any particular object a poor strategy, one that would pay off on only $1/8^{th}$ of trials. In other words, valid primes had little predictive value on any particular trial, such that any reduction on the object word should come about primarily because of conceptual or linguistic priming.

Participants completed 12 practice trials prior to the experimental trials. These were designed to familiarize them to a greater extent with the action words, which several subjects in Experiment 1 struggled with. No experimental items appeared as moving objects in the practice trials. Participants were offered an opportunity to take a break halfway through the experiment. The entire process took approximately 45 minutes.

Analysis. The analysis took the same general form as in Experiment 1, with two exceptions. No secondary analysis across instruction numbers within a trial was performed, because trials included only one instruction. Then, because the experiment was designed to test for the presence of simple effects of priming, the linguistic and non-linguistic conditions were first examined separately, and then pooled to test the overall effect of the prime.

Results. Means of the onset to speak latency and the object word can be found in Table 3, and graphically in Figures 8 and 9. Examination of the table reveals that the prime effect (i.e., the difference between the valid and invalid prime conditions) was numerically greater for both linguistic and conceptual primes, for both latency and object word duration. However, the model revealed that only the object duration differed significantly, and only for the linguistic prime.

The analysis bore out this trend, and revealed no significant difference due to validity of the prime (t = -.48, p > .63, pMCMC > .65). None of the other parameters of interest approached significance, either.

An examination of the means for the object durations shows a numerical trend in both the linguistic and non-linguistic condition toward reduction due to valid priming, but a smaller trend in the non-linguistic condition. An analysis broken down by condition (i.e. simple effects) revealed a significant difference due to prime validity in the linguistic condition (t = 1.838, p < .04, pMCMC < .05), but not in the non-linguistic condition (p > .57). A model of the full dataset revealed a significant difference between valid and invalid prime (t = -2.24, p < .03, pMCMC < .03). Neither condition nor the interaction between condition and prime validity was significant.

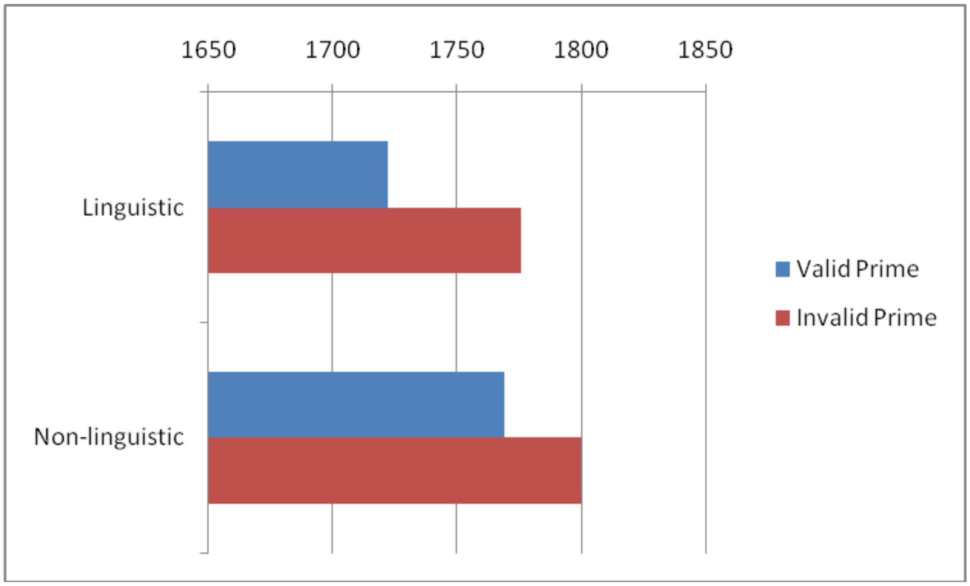

Figure 8– Experiment 2 - Means of the Onset Duration in milliseconds

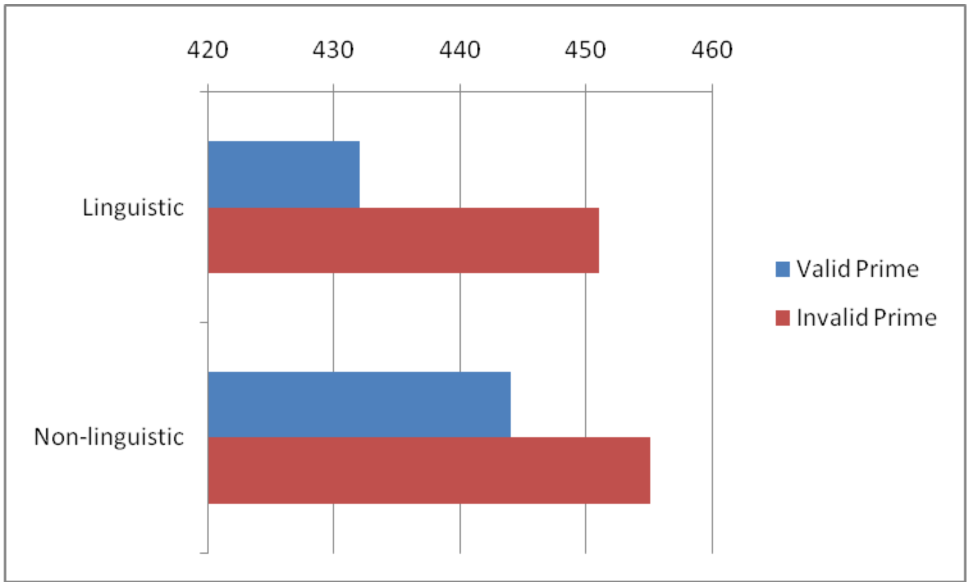

Figure 9– Experiment 2 – Means of the object word in milliseconds

	Onset to Speak			Object Word		
	Valid Prime	Invalid Prime	Priming Difference	Valid Prime	Invalid Prime	Priming Difference
Linguistic	1722	1776	54	432	451	19
Non-linguistic	1769	1800	31	444	455	11

Table 3: Experiment 2 – Means of the onset to speak and object word in milliseconds

Discussion. The results provided two positive answers to the questions: priming and predictability do appear to have separable effects, and linguistic and conceptual priming once again differ in their capacity to produce reduction. The valid primes elicited reduction even though they did not make the referential event particularly predictable. This effect only appeared on the object word durations, however, and not the onset to speak durations, suggesting that this facilitation effect resulted from activation, rather than planning. Mitigating the usefulness of the prime as a cue to the upcoming referential event appears to have eliminated speakers' ability to plan the utterance quickly.

The effect of valid priming also appeared only in the linguistic condition, not the non-linguistic condition. This may reflect a true null result, in that it is possible that conceptual priming does not elicit reduction without strong predictability. However, the manipulation may simply have lacked the power necessary to detect a subtle effect. The numerical trend for the non-linguistic condition was small, but in the right direction. Conceptual priming may simply activate fewer of the necessary representations than linguistic priming, watering down its effect.

One difference between the experimental design in these two experiments may have had an influence on the results. In Experiment 1, both the speaker and the listener had access to the priming information, which was either spoken aloud for both to hear, or flashed on both participants' screens. In Experiment 2, the priming information was explicitly directed at the speaker, who had to react to it by clicking on the primed object. This was especially true for the conceptual priming information, which only the speaker saw, and which is thus confounded with conceptual priming. The exclusivity of the priming information may have led them to treat it as irrelevant to the task, because they did not share it with their listener. At the very least, this works against the manipulation, and the presence of an effect due to the linguistic priming shows its robustness. On the other hand, a separate (unreported) experiment that was designed to explicitly test whether speakers in this task paid attention to whether their listener also received the same priming information showed that only the speakers' primes mattered. That is, speakers reduced when and only when they had the priming information, and paid no attention the listener's state.

Experiment 3: A Further Test of the Independent of Conceptual Priming and Predictability

Motivation. Experiment Two demonstrates the persistent effect of linguistic priming, even in the absence of strong predictability. However, the manipulation does not elicit an effect of conceptual priming, unlike in Experiment 1. The disparity between the two experiments raises the possibility that linguistic and non-linguistic priming differ significantly in their ability to produce reduction. Linguistic priming appears to operate at least partly independently of predictability, while the results of Experiment Two suggest that conceptual priming does not. More specifically, both priming and predictability appear to serve as a trigger for linguistically-centered reduction, while non-linguistically-centered reduction appears to require predictability. The ARH would explain by this difference by noting that linguistic priming would activate a greater portion of the representations in the language production system, relative to conceptual priming.

Predictability may matter for conceptual priming more simply because it would allow speakers to capitalize on the (presumably) minimal activation that arises due to the prime.

Experiment 3 is designed as a more direct test of the effect of conceptual on reduction, independent of the effect of predictability. More specifically, it asks two questions: 1) Is conceptual sufficient to produce reduction by itself, in the absence of strong predictability? 2) Do eye movements reveal anything about whether participants actually use the priming information? Another possible difference between the linguistic and conceptual primes in the first two experiments is that the linguistic primes have a much greater capacity to permit prior naming. That is, speakers could simply have repeated the names of the primed objects silently to themselves, and this rehearsal could have produced the effect. Presumably, such rehearsal would have had less of an effect on conceptually primed objects, simply because of the diminished initial activation. Eye-tracking will permit Experiment 3 to investigate whether participants actively seek out information throughout the course of a trial, or whether they simply absorb the priming information wholesale and use it implicitly.

Participants. 20 undergraduates participated in an instruction-giving task similar to the tasks in the prior experiments, and were offered course credit for their participation. 2 were excluded due to technical difficulties, for a total of 15 speakers.

Materials. The same materials, including the colorized line drawings, recordings, and hardware were used as in Experiment 1. The experimental targets were once again chosen from the list of objects in Experiment 1 that elicited the most reduction, with consideration given to objects that had poor naming agreement from the first two experiments.

Design. A trial consisted of two parts: a judgment task and an instruction-giving task. For the judgment task, 3 pictures of objects appeared on a computer screen, centered vertically and spaced horizontally to discourage grouping. Under each picture were a Yes and a No button, which the speaker used to answer one of three questions: 1) “Would this object fit in a drawer?” 2) “Is this an object you would find around the house?” 3) “Is this object a living thing?” The question appeared at the top of the screen. Questions were randomized, such that no question appeared twice in a row. All three of the objects from the judgment portion of a trial appeared as objects in the instruction-giving portion.

The instruction-giving task took the same form as in Experiment 1, with the exceptions that no priming information was provided, and no listener was present. 8 objects appeared on the screen, and after 2.5 seconds, one object moved. The speaker described this movement, and then clicked on the object to advance the trial. A second object moved, then the speaker issued a second instruction then clicked to advance the trial to the third object.

The experiment consisted of three conditions: predictable-and-primed, only-primed, and neither. In the predictable-and-primed condition, the three objects that featured in the judgment task were the three objects that moved for the instruction-giving task. The order of movement was constrained such that the leftmost object in the judgment task moved first, followed by the center object, and finally the rightmost object. The judgment task thus made the instruction-giving objects predictable. In the only-primed condition, the first two objects to move were randomly selected from nonjudgment task objects, while the third was always the rightmost judged object, which was always an experimental item. In the neither trials, none of the judged objects served as instruction-giving task objects, although the third object to move was always an experimental item. The entire experiment consisted of 80 trials, 32 of which were fillers. The filler trials resembled the neither condition trials, in that none of the judgment objects moved in the instruction-giving task.

Procedure. Speakers were seated in front of a computer and had their eyes tracked with an Eyelink II (X). After calibration of the eye-tracker, they were given instructions about the judgment and instruction-giving tasks. They were told to answer questions about the judged objects from the leftmost object to the rightmost object, in order to facilitate the predictability manipulation. The instructions emphasized that they should make their best guess about each object. Speakers then completed 5 practice trials, where they were familiarized with the instruction-giving task, as well as the details of each question. For the drawer question, for example, they were told to answer whether a typical example of the object would fit in one of the filing cabinet drawers in the lab.

Experimental items appeared in a constrained random order, where each item had to appear once and only once in each condition, and could not appear twice in a row.

Judgment and instruction-giving portions of the trial alternated, with the judgment task giving way to the instruction-giving task upon answering all three questions. During the instruction-giving task, objects appeared on the screen for 2.5s before the first movement occurred, and the second and third object did not begin movement until 1s after the previous object had been clicked. Trials ended when speakers issued the third instruction and clicked the third object.

Acoustic Analysis. The analysis took the same form as Experiment 1.

Results. Means of the onset to speak latency and the object word can be found in Table 4, and graphically in Figures 10 and 11. The table shows that the predictable-and-primed condition had a shorter object duration on average than the only-primed condition, which suggests that the predictability manipulation worked. The shortest average duration overall, however, is in the not-primed condition, suggesting that the priming manipulation did not take hold.

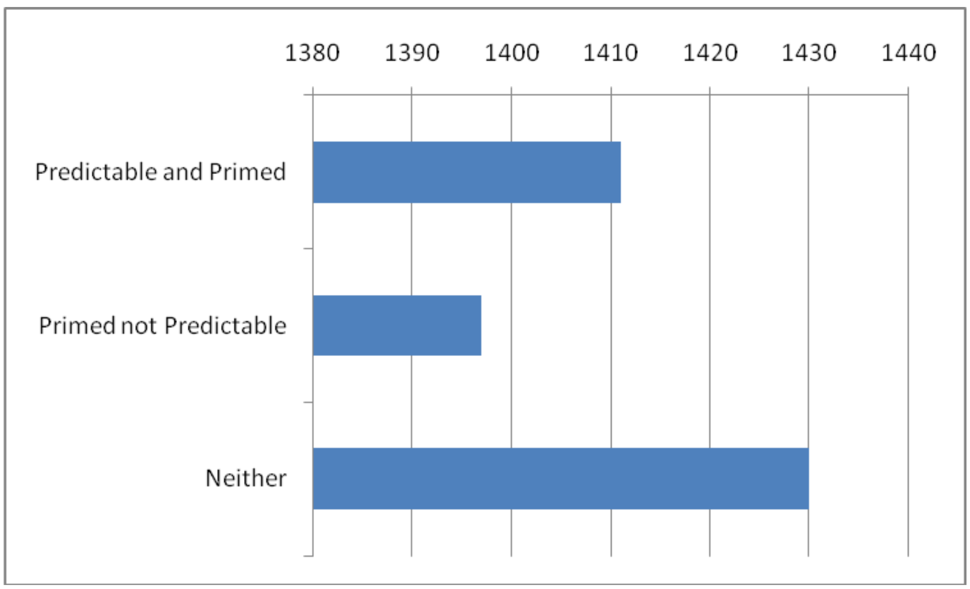

Figure 10 – Experiment 3 – Means of the onset duration in milliseconds

	Onset	Object word
Predictable and Primed	1367	419
Predictable not Primed	1356	427
Neither	1419	418

Table 4: Experiment 3 – Means of the onset to speak and object word durations in milliseconds

The statistical analysis of the duration of the object word shows that the predictable-and-primed condition differs significantly from the only-primed condition ($t = 1.97$, $p < .04$, pMCMC < .04), but does not differ significantly from the not-primed condition ($t = -11$, $p > .9$, pMCMC > .89). The analysis of the onset duration data revealed no significant differences among the predictors of interest (all t's < 1).

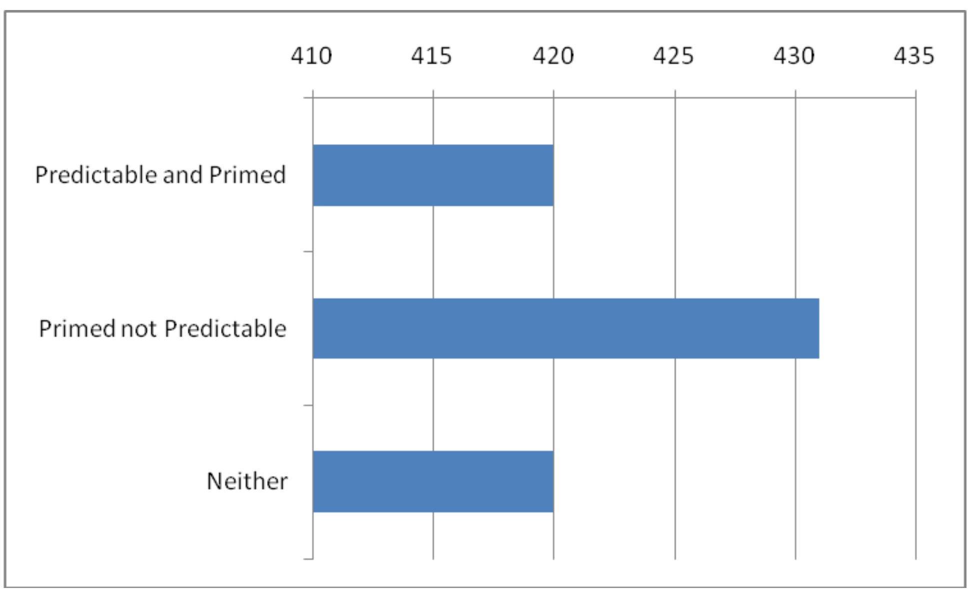

Figure 11– Experiment 3 – Means of the object word in milliseconds

Eyetracking analysis. The purpose of the eyetracking analysis was to determine the extent to which participants used the predictability manipulation provided by the judgment task. If their eye movements anticipated the third object in the predictable-and-primed condition relative to the other two conditions, that would serve as evidence that, even in the absence of acoustic data reflecting activation, the participants were still attending to the manipulation.

The region of interest was thus between the end of the second object's portion of the trial (i.e. after the participant had clicked on it), and before the average onset of utterances about the third object. Figure 12 shows a time window from 600-1500ms after the second object was clicked. An inspection of the movements over time shows that none of the three conditions produced any noticeable anticipatory behavior. The scale of proportion of looks to the target in this region shows minimal looks overall, with the baseline hovering below 10% of time spent looking at the target. Toward the end of the region, participants begin to look at the target, but only after it has started to move, which also suggests no anticipation.

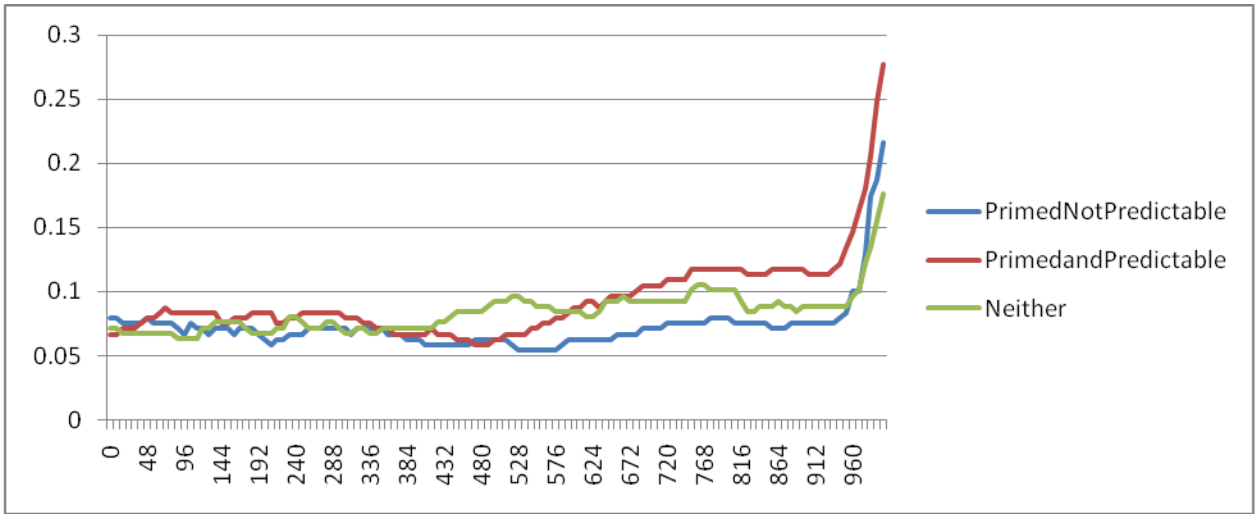

Figure 12– Eye movements from Experiment 3, ranging from 600-1500ms after the participant clicked on the 2nd object.

A statistical analysis revealed the same result. Looks to the target in the time region were aggregated first by subjects, then by items. These aggregated data were placed into 300ms bins (i.e. 600-900, 900-1200, 1200-1500) in order to permit calculation of an empirical logit. The empirical logit in this case is a rough measure of the number of looks to the target over the total number of potential looks to anywhere on the screen. It thus reflects a proportion, but has mathematical properties that make it a better candidate for serving as a dependent regression variable (see Barr, 2008 for a more detailed explanation). The empirical logits, calculated by subjects and then items, were submitted to a multilevel logistic regression. No significant effect of condition emerged in either analysis (all t's < .05).

Discussion. The null result shown here both behaviorally and with eye movements could have come about for several reasons. The most obvious is that the effect of sheer conceptual priming simply does not exist. The original formulation of the

ARH may have been too bold, and should be scaled back to include only certain types of activation. The difference between the non-linguistic and control condition in Experiment 1 suggests otherwise, however. That is, conceptual priming appears to have an effect, even though it does not achieve discourse status in any proper sense. Of course, in Experiment 1, conceptually primed information also thus became predictable. Further research, discussed below, will address this issue in more detail. Another possibility is that the effect was too small for this sample size to detect. An effort was made to create sample size parity among the experiments reported in order to facilitate a comparison among the effects, albeit an imperfect one. Because conceptual primes should activate fewer representations, conceptual priming may simply have a smaller effect on the production system, even when paired with predictability. Doubling the number of subjects might reveal an effect of this size.

A more theoretically (as opposed to methodologically) motivated possibility is that priming, especially conceptual priming, must feature in the same task as the relevant utterance. The judgment task and instruction-giving task always shared objects, but at no point did any of the information in the judgment task bear explicitly on the information in the instruction-giving task. Participants could simply perform the tasks completely separately, ignoring the implicit priming of the objects in the judgment task. Similarly, the predictability manipulation might not have worked as well as it did in Experiment 1, or for the linguistic primes in Experiment Two, simply because the domain of predictability shifted with the change in tasks. Participants had no reason to expect that the presence of an object in the judgment task was related in any way to the movement of the objects in the instruction task. Indeed, almost no subjects reported being aware that the judgment objects occasionally served as the moving objects in the other task in the order in which they were judged. Together, these two alternatives suggest that a manipulation that made the connection between the two tasks more explicit, or use conceptual primes that were either predictable or not within the same task, would stand a better chance of detecting any existing effect.

***General Discussion*.**

Summary of findings. The reported experiments tested the Activation Reduction Hypothesis, repeated here: any stimulus or processing that activates representations associated with a to-be-produced word should lead to some measurable amount of reduction. In other words, facilitation of any part of the language production process should result in reduction, *ceteris paribus*. A separate claim, which the experiments also tested, were that the amount of reduction a stimulus creates should be a function of the proportion of the language production processes it facilitates, and the means by which the representations become activated. Experiment 1 addressed three more specific questions: 1) Does referential-event predictability and/or priming 'count' for the purposes of reduction? 2) Does non-linguistic information 'count' for the purposes of reduction? 3) Do linguistic and non-linguistic information elicit different degrees of reduction? Experiment Two addressed a further pair of questions pertaining to priming, predictability, and linguistic versus conceptual priming: 4) Does priming operate independently of predictability for the purpose of producing reduction? 5) Does the same relationship between priming and predictability hold for linguistic and non-linguistic information? Experiment 3 examined conceptual priming in more detail, by asking the following two questions: 6) Is conceptual priming sufficient to produce reduction by itself, in the absence of strong predictability? 7) Do eye movements reveal anything about whether participants actually use the priming information?

All three experiments affirm that reduction can arise because of information pertaining to referential events. The focus on referential events stems from the idea that much of the patterning of natural language production arises from the expectations speakers have about what they will talk about next. Both priming and predictability, the two major factors in these experiments, can operate over referential events, in addition to the variables and entities they operate over in most other studies on reduction. Corpus work, the primary source of evidence for the effect of predictability on reduction, has not included referential events. This set of experiments thus represents an important extension of the range of variables that affect reduction, especially inasmuch as most corpora do not contain the situational information necessary for drawing inferences about given and/or predictable references to entities. Another way to view this result is that the design controlled for the *n*-gram and frequency information internal to the utterances, but still showed an effect of both priming and predictability. To produce cleaner results, future reduction studies should similarly attempt to control for referential event priming and/or predictability. In natural conversations (such as those found in corpora) and experimental settings, the next word in an utterance often refers to an entity that the context has made predictable independent of the particular form chosen as a reference. Referential predictability may be confounded with standard measures of predictability in some situations, making unambiguous attribution of an effect on reduction to standard predictability impossible.

The results from Experiment 1 also provide an important demonstration that referential events, and priming and predictability more generally, can affect reduction in both a linguistic and a non-linguistic format. Like referential events, nearly all corpora suffer from the limitation that very little non-linguistic information, if any, gets coded alongside the text. Although this is not the first experiment to demonstrate an effect of conceptual priming on reduction (Bard & Anderson, 1994), it is the first to demonstrate that a direct *manipulation* of conceptual priming produces reduction relative to a new control, and does so by making reference to a particular object given *per se*. Theories of givenness tend to allow non-linguistic information to ”count” for the purposes of givenness, but do so by merely changing the information status of the entity, and without an explicit claim about the relationship between linguistic and conceptual priming’s tendency or capacity to produce reduction. The ARH naturally accounts for the inclusion of non-linguistic information in the cluster of factors that affect reduction, because it activates the conceptual representations associated with reference to a particular object.

Perhaps the most important result to emerge from Experiment 1 is the difference between linguistic and non-linguistic priming. After holding predictability constant, linguistic givenness elicited shorter object word durations than conceptual priming. This follows naturally from the ARH, which predicts that a greater amount of facilitation of the language production process should lead to a greater amount of reduction. Indeed, linguistic givenness should have activated at least the lexical and phonological representations of the object words, in addition to the conceptual representations. The additional activation post-conceptually would explain the difference between the two priming conditions. Just as importantly, this difference emerges in the presence of 100% predictable entities, suggesting that predictability and priming can, in principle if not normally in practice, have separable and at least semi-independent effects on reduction.

One concern is that the results reflect an overall increase of speaking rate in the experimental conditions, as opposed to a "genuine," word-level reduction. The models in the follow-up analysis support the opposite conclusion. These models included the durations of the other segments of each utterance, and still retained a significant effect of condition. Even after controlling statistically for speaking rate, reduction appeared in both experimental conditions.

Experiment 2 also tests the novel predictions of the ARH by asking whether priming and predictability operate independently, and whether linguistic and non-linguistic information respond to priming and predictability manipulations in the same way. The results from this experiment suggest that priming can operate independently from predictability for the purposes of producing reduction, and that linguistic and conceptual priming differ in that capacity. After mitigating the effects of predictability, there is an effect of priming on the object word, but not the onset to speak duration. This pattern only holds for the linguistic condition, however. Conceptual priming does not have a significant impact on either onset or word duration. Both the lack of a significant difference in the onset duration and the null effects for the non-linguistic condition may simply reflect a lack of power. In both cases, a small numerical trend in the right direction suggests a not-quite detected effect. As mentioned in the discussion for Experiment 3, sample size across these experiments is approximately the same, in order to facilitate cross-experiment comparisons. A potentially more viable alternative would be to pick sample sizes based on the expected effect sizes.

Experiment 3 narrowed the focus of the first two experiments to only conceptual priming. The judgment task appears not to have primed participants sufficiently well for them to count the judged objects as given for the purposes of reduction. The priming information did not result in reduced duration on the object word nor the onset duration. Adding predictability information, where the judged objects served as instruction-giving objects, in the same order, also did not produce a significant difference relative to the control. As previously mentioned, the small sample size may prohibit drawing any firm conclusions about the absence of an effect here. Although Experiment 1 showed an effect of conceptual priming, the task made the analogy between linguistic and non-linguistic priming clear, whereas in this task, the relatively weaker manipulation may not have activated representations sufficiently to produce reduction. That is, participants had to group the judgment objects together as a set, remember them, and interpret their grouping as relevant to the instruction-giving task, which may not have been readily apparent. The lack of an effect in the eyetracking analysis supports the idea that the predictability manipulation did not work.

Assessing the ARH. Broadly speaking, the ARH posits that any stimulus that activates relevant representations in the language production system should lead to reduction. Priming and predictability both fall under this umbrella, although potentially for different reasons. Non-linguistic and linguistic priming do as well. Referential events should lead to reduction to the extent that they produce priming and/or predictability.

Priming should directly activate representations associated with particular levels of production, and predictability should facilitate the processing associated with production.

The results of Experiment 1 accord quite well with the ARH, which readily explains all of the theoretically interesting outcomes. Referential events lead to reduction. Both of the experimental conditions, linguistic and non-linguistic priming, elicit reduction on the object word and onset duration. The linguistic condition elicited more reduction than the non-linguistic condition on the object word, which is consistent with the separation of priming and predictability effects. Linguistic priming activated more representations in the language production system than conceptual priming. As the predictability of the objects increased across mentions, the duration of the onset to speak decreased for both the linguistic and non-linguistic condition. On the third mention, both experimental conditions elicited the same amount of reduction, which is consistent with the idea that predictability specifically affects planning.

The ARH also plausibly explains the results of Experiment 2. The linguistic condition again activated a greater proportion of the language production system, and thus created reduction on the object word, even in the absence of strong predictability, whereas the non-linguistic condition activated only the conceptual representations, at best, producing either an undetectable or non-existent amount of reduction. The minimal predictability associated with each prime decreased the degree to which speakers could plan ahead and streamline the entire production process, which explains the null effect on the onset durations. A less charitable interpretation would recommend scaling back the scope of the ARH to include any linguistic stimuli but only non-linguistic stimuli that become predictable as well as given.

The results from Experiment 3 do not require much in the line of interpretation – a null result offers little to the conversation. As noted earlier, one possible explanation for the ineffectiveness of the judgment task at eliciting givenness status is its complete disjunction from the instruction-giving task. Participants did not have to pay attention to the relationship between the tasks in order to successfully complete them, and in fact reported that they did not do so. A modified version of the ARH might pull back to include only stimuli that become relevant for some linguistic task – a conversation, an experimental setting, a lecture, etc. Looking around the room while giving a lecture on phonemes and seeing windows, for example, should not necessarily lead to reduction. Looking around the room, finding a window, and then talking about it, might do.

The ARH and other proposals. The ARH is not the first proposal to claim that activation of representations in the production system leads to reduction. Most recently, Bell et al. (2009) claim that activation of the lexical level of representation accounts for the results they report, and further that coordination between the lexical and phonological levels explains word lengthening. The ARH differs from the hypothesis proposed by Bell et al. in at least two important ways. The ARH subsumes activation at any level of representation in the language production system, not only the lexical level. In predicts that any stimuli that activate both the lexical *and* phonological levels, for example, should produce measurably more reduction than a stimulus that activates only the lexical level, irrespective of the coordination between the two. It also predicts that conceptual priming should lead to reduction, despite its at best indirect activation of lexical representations. The scope of the ARH thus extends beyond simple lexical activation. It remains to be seen whether this will be to its benefit or detriment.

One possible implementation of predictability-based accounts of reduction is neural coding of frequency effects. Frequency factors into many of the equations used to calculate probability in corpus studies. Bybee (2001), for example, argues based on homophone frequency attraction data that the shorter durations associated with highly frequent words come about because of routinized articulation. In other words, because speakers often produce highly frequent words, the motor programs associated with pronouncing those words become overlearned. Levelt et al. (1999) situate frequency effects at the phonological level, based on frequency inheritance, where a low frequency word with a high frequency homophone expresses the same behavior as the high frequency homophone. Gahl (2008), *contra* both Bybee and Levelt et al., presents a corpus study that suggests the lemma level affects word duration after controlling for phonological effects. All three of these proposals posits that the language production system is sensitive to activation, in the sense that forms coded based on frequency should exhibit differential activation profiles during production. The ARH is compatible with these claims, in that any type of facilitation and/or activation in the production system should lead to reduction. Frequency information is no exception.

Proponents of the Uniform Information Density Hypothesis (hereafter UID; Jaeger 2006; Levy & Jaeger 2007) take the frequency argument one step further, and argue that speakers manage probability at every level of linguistic representation in order to modulate word durations (among other speech phenomena). In order to communicate efficiently, from an information-theoretic perspective (Shannon, 1949), speakers should modulate the amount of information they put into a unit of time. To accomplish this, they use probabilistic information about the language, in the sense that they should pronounce a predictable word with shorter duration and an unpredictable word with longer duration in order to make their speech as uniformly informative as possible. Like the frequency-based hypotheses, UID predicts that the probability of linguistic information should affect reduction, but it extends the number of levels of linguistic representation involved in this calculation. Predictability information at *all* levels of linguistic representation should affect reduction.

There are two primary differences between the ARH and UID: 1) the ARH explicitly links predictability and givenness in the common medium of activation, while UID leaves them as separate contributers to the choices involved in reduction; 2) UID makes explicit predictions and offers evidence for categorical forms of reduction. Like many other proposals, UID recognizes the contributions of both givenness and predictability in eliciting reduction. Its mathematical formulation, however, emphasizes only predictability, and relegates givenness to a completely separate predictor in a regression equation. It also posits no direct relationship between the two. In principle, the theory could be extended to include a common substrate, or a more formal recognition of the informational contribution of givenness, but based on its current formulation, the ARH makes broader, simpler, and more cognitively plausible predictions. In its favor, however, UID shows that categorical reduction, where speakers leave out entire words (e.g. *that's* in *the towel on the floor*), occurs in exactly the environments the theory predicts. The ARH as it stands makes no predictions about reduction that pertains to inclusion or exclusions of words, but could be extended to do so.

Although some linguistic theories do make a distinction between linguistic and non-linguistic givenness, the ARH is the first proposal to make predictions about behavioral differences between the two, and explicitly propose an underlying reason. Most accounts of givenness claim that once an entity enters the discourse, through whatever means, it counts as given and retains that status for a set period of time. Givenness is thus somewhat monolithic, and has little to do with the processes that lead to actual language production. By contrast, the ARH draws a principled distinction between the contributions of linguistic and conceptual priming to reduction, and allows priming to take on graded values. As with the other proposals, this is not to say that givenness accounts could not extend to make the same predictions as the ARH, but currently none do.

A final, somewhat orthogonal issue is the debate about speaker- versus listener-centric reduction. One extreme says that speakers constantly take their listener's state of mind into account during production, while the other says that speakers only rarely track the listener. Neither is likely to be the case. The more interesting line of investigation thus becomes determining when, why, and how speakers track their listener's internal state, and what they do with this information. The studies reported here have little to say directly about this debate, because they all either provided speakers and listeners with the same information or eliminated the listener altogether. At best, the ARH can merely say that the internal state of the speaker appears to matter considerably for calculations of givenness, but alternative interpretations of the experimental findings are possible. More strongly, the evidence here suggests that audience design is not the only factor in operation during reduction.

Future directions. The ARH makes a host of novel predictions, but still has limitations. These studies also only ask a small subset of the interesting questions that arise from its formulation. In this closing section, I would like to outline some ongoing work, as well as the next steps for testing the ARH.

The question that most demands an answer pertains to the null result from Experiment 3. Does conceptual priming count for the purposes of reduction, even in the absence of strong predictability? The task appears to have fallen too far on the side of distal priming, with the judgment task doing too little to elicit activation. More powerful manipulations could create a closer relationship between the conceptual prime and the task that elicits the utterances, such as in Experiment 1. Activation that had a greater, more direct bearing on the utterance's task might experience less filtering due to task-switching. A second type of manipulation could emphasize the *degree* of activation, which for the simple and easily-identifiable objects here is presumably minimal. More complex objects that demanded more attention and thoughtful consideration, or questions with less straightforward answers, might produce greater or more lasting activation, and lead to more reduction.

A second, more theoretically motivated question, is whether activation at different levels of the production system have measurably different effects on reduction. The ARH does not make a principled distinction between activation at one level versus another, but may be too coarse-grained, in that sense. Activation at the lexical level, for example, may play a greater role in determining reduction than activation at the phonological level. A line of experiments that attempts to prime only one level of the production system at a time, while holding activation in the others constant, is already underway.

Another unanswered question is whether planning and sheer activation differ. The results of Experiments 1 and 2 suggest that predictability and priming have separable effects on reduction. Predictability appears to affect planning, while priming appears to lead to activation *per se*. Because neither of these experiments featured eye-tracking, the results do not fully support the conclusion that the two factors of interest operate complete in separate domains. A replication of Experiment 1 that used anticipatory eye movements to factor out predictability would go some way toward delimiting its role in reduction.

Finally, the role of the listener remains unknown. Experiment 1 intentionally sidestepped the question by giving priming information to both speaker and listener. After a follow-up experiment that showed no effect of the presence of the listener, Experiments 2 and 3 essentially treated that factor as irrelevant. The low interactivity of these experiments may have been the primary factor in the speakers' inattention to the listener. None of the tasks required the listener to supply any information to the speaker, reducing the comparability to earlier work on audience design. A more interactive task, where speaker and listener alternated, while still adhering to the linguistic versus non-linguistic priming structure, or the priming-but-not-predictable constraint, might serve as a better test.

Final summary. Generally speaking, these results support the ARH, although some fine-tuning may be necessary, depending on the susceptibility of the language production system to conceptual priming. Referential events cause reduction, which extends previous findings to a new, more naturalistic variable. Conceptual priming also causes reduction under certain circumstances, namely in the presence of predictability.

Linguistic priming causes *more* reduction, arguably because it activates more of the representations associated with language production. It also causes reduction even in the absence of strong predictability, which provides further evidence for a sheer-activation account like the ARH.

Appendix: List of object names

Accordion, airplane, alligator, anchor, ant, arm, arrow, artichoke, ashtray, asparagus, axe, ball, balloon, banana, barn, barrel, basket, bat, bear, bed, bee, beetle, bell, belt, bike, bird, blouse, book, boot, bottle, bowl, box, bow, bread, broom, brush, bus, butterfly, button, cake, camel, candle, cannon, cap, car, carrot, cat, caterpillar, celery, chain, chair, cherry, chicken, chisel, church, cigar, cigarette, clock, clothespin, cloud, clown, coat, comb, corn, couch, cow, crown, cup, deer, desk, dog, doll, donkey, door, doorknob, dress, dresser, drum, duck, eagle, ear, elephant, envelope, eye, fence, finger, fish, flag, flower, flute, fly, foot, football, fork, fox, frog, giraffe, glass, glasses, glove, goat, gorilla, grapes, grasshopper, guitar, gun, hair, hammer, hand, hangar, harp, hat, heart, helicopter, horse, house, iron, jacket, kangaroo, kettle, key, kite, knife, ladder, lamp, leaf, leg, lemon, leopard, lettuce, lightbulb, lightswitch, lion, lips, lobster, lock, mitten, monkey, moon, motorcycle, mountain, mouse, mushroom, nail, nailfile, necklace, needle, nose, nut, onion, orange, ostrich, owl, paintbrush, pan, pants, peach, peacock, peanut, pear, pen, pencil, penguin, pepper, piano, pig, pineapple, pipe, pitcher, pliers, plug, potato, pumpkin, purse, rabbit, raccoon, rhinoceros, ring, rollerskate, rooster, ruler, sailboat, salt, sandwich, saw, scissors, screw, screwdriver, seal, sheep, shirt, shoe, skirt, skunk, sled, snail, snake, snowman, sock, spider, spool, spoon, squirrel, star, stool, stove, strawberry, suitcase, sun, swan, sweater, swing, table, telephone, television, thimble, thumb, thumbnail, tie, tiger, toaster, tomato, toothbrush, top, train, trashcan, tree, truck, trumpet, turtle, umbrella, urn, vest, violin, wagon, watch, watermelon, well, wheel, whistle, windmill, window, wrench, zebra

References

Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J. Miller, J., Sotillo, C., Thompson, H. S. & Weinert. R., (1991). The HCRC Map Task Corpus. *Language and Speech, 34,* 351-366.

Arnold, J. E. (1998). Reference form and discourse patterns (Doctoral dissertation, Stanford University, 1998). *Dissertation Abstracts International*, *59*, 2950.

Arnold, J. E. (2008). THE BACON not *the bacon*: How children and adults understand accented and unaccented noun phrases. *Cognition*, *108* (1), 69-99.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*, 521–536.

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.

Aylett, M. & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119* (5), 3048-3058.

Balota, D. A., Boland, J. E. & Shields, L. W. (1989). Priming in Pronunciation: Beyond Pattern Recognition and Onset Latency. *Journal of Memory and Language*, *28*, 14-36.

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

Bard, E. G., & Aylett, M. P. (2004). Referential Form, Word Duration, and Modeling the Listener in Spoken Dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173-191). Cambridge, MA: MIT Press.

Bard, E. G., Lowe, A., & Altmann, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), *Proceedings of EUROSPEECH ’89 vol. 2* (pp. 573-576). Paris: European Speech Communication Assocation.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38,* 1636-1657.

Baumann, S., & Hadelich, K. (2003). Accent Type and Givenness: An Experiment with Auditory and Visual Priming. *Proceedings of the 15th ICPhS, Barcelona, Spain*. (pp. 1811-1814).

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behavior in conversation*. Milton Keyes: Open University Press.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*, 1001–1024.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92-111.

Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Philadelphia: John Benjamins.

Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.15) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.

Brown, G. (1983). Prosodic Structure and the Given/New Distinction. A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 67-77).

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25-56). New York: Academic Press.

Chafe, W. L. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 20-51). Amsterdam: John Benjamins.

Chafe, W. L. (1994). *Discourse, Consciousness, and Time*. Chicago: Chicago University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73-111.

Clark, H.H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L. & Sag, I.S., (Eds.). *Elements of Discourse Understanding* (pp. 10-63). New York: Cambridge University Press.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453-459.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292-314.

Deusch, W., & Pechmann, T. (1982). Social interaction in the development of definite descriptions. *Cognition, 11,* 159-184.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409-436.

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 197-827). Cambridge, Mass.: MIT Press.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31*, 307-319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Gahl, S. & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language, 80* (4), 748-775.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1-42). Amsterdam: John Benjamins Publishing.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability on lexical production. In *CLS-99* (pp. 151-166). University of Chicago.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review*, *10* (3), 603-609.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*, *38*, 313-338.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. Mouton: The Hague.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67–104.

Jaeger, T. F. (2006a). Redundancy and syntactic reduction in spontaneous speech. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: Benjamin.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 34* (4), 391-405.

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

Lambrecht, K. 1994.*Information structure and sentence form: Topic, focus, and the mental representation of discourse referents.* In Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.

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

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

Lieberman, P. (1963). Some effects of the semantic and grammatical context on the production and preparation of speech. *Language and Speech*, *6*, 172-175.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer Academic.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Coehn et al. (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge, MA: MIT Press.

Prince, E. (1981). Toward a taxonomy of given-new information. In Cole, P. (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thomspon & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295-325). Philadelphia: John Benjamins B.V.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition [On-line]. *Journal of Vision, 1*, 413a.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing "given" information: Effects of persistence of grammatical functions and surface position. *Language and Speech, 37*, 125-145.

Terken, J. (1985). *Use and Function of Accentuation: Some Experiment.* Ph.D. thesis, University of Leyden, Leyden, The Netherlands.

Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on Production of Referring Expressions*, Cogsci 2009.

Venditti, Jennifer J. and Julia Hirschberg. (2003) Intonation and discourse processing. Invited paper in *Proceedings of the International Congress of Phonetic Sciences* (ICPhS). Barcelona, Spain.

Watson, D., & Arnold, J. E. (2005). Not just given and new: The effects of discourse and task-based constraints on acoustic prominence. In Poster presented at the 18th annual CUNY conference on human sentence processing, Tuscon, AZ.

Watson, D. & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*, 713-755.

Watson, D.G., Arnold, J.E., & Tanenhaus, M.K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*, 156-1557.

Zipf, G. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology, 40*, 1-95.