Syntactic Probability Influences Duration *

Claire Moore-Cantwell

August 27, 2012

1 Introduction

There is strong evidence that frequencies of lexical items, word-to-word transitional probabilities, and even probabilities over large groups of words (Arnon and Miller, 2011; Arnon and Snider, 2010) must be available to the language production system, as they influence the phonetic duration of words. This paper demonstrates that probabilities over syntactic structures are likewise available to the production system, and can influence the duration of words situated in those structures[1]. Unlike lexical items, syntactic structures can be composed online according to a system of rules, and need not be stored individually.

I define 'frequency' here as simply a record kept by the production system of how many times a particular lexical item[2] has been encountered (either produced or perceived). One way this record has been formalized is in automatic spreading activation models (Collins and Loftus, 1975; McClelland and Elman, 1986) - lexical items become active when they are accessed, and that activation decays over time. Lexical items which are used more frequently get activated more often, and thus have a higher baseline activation level than lexical items which are used less frequently. I define the probability of an item as the frequency of that item given its occurrence in a particular context. In other words, to say that the production system tracks probabilities over lexical items is to say that the production system keeps a record of the percentage of times a particular lexical item occurs after a particular context

Because syntactic structures need not be independently stored in long-term memory (since they are compositional), if the production system tracks probabilities over syntactic structures, this implies a more complicated mechanism than automatic spreading activation over stored items. In section 7 I suggest that the production system gains 'practice' with more frequent structures through a process of implicit learning.

*Thanks to everyone who helped me out with this project, especially Lyn Frazier, Michael Wagner, Anne Pycha and Kristine Yu.

[1]There need not be an explicit representation of a structure's probability stored somewhere for the production system's use. I will make a slightly less strong claim: just that the probability of a structure influences the production system's behaviour in specific ways

[2]I will assume that a lexical item is any linguistic unit which is stored in its entirety in long-term memory as an entry in the mental lexicon. Linguistic units which must be stored in the mental lexicon have meanings and/or behavior which cannot be deduced by examination of the sub-parts of the unit. Under this definition, individual words and morphemes count as lexical items, and so do multi-word idioms like 'kick the bucket'. Crucially unlike a lexical item, a syntactic structure can be derived via some set of syntactic rules, and need not be stored in the mental lexicon.

The less probable a structure is, the more difficult it is to plan. This difficulty of planning leads to increased durations at prosodic boundaries before low-probability structures.

In order to show that probabilities over syntactic structures influence duration, it will be necessary to disambiguate between the probability of the structure and the probability of specific words or groups of words within that structure. Also, many properties of syntactic structures are highly correlated with their probability. Longer and more complex structures tend to occur less often, and shorter and less complex structures occur more often. Without separating the effects of length and complexity from any effects of probability, it will be impossible to tell whether probability itself is relevant for the production system.

Recent work (Gahl and Garnsey, 2004; Gahl et al., 2006) proposes that probabilities over syntactic structures are in fact explicitly stored as part of a speaker's knowledge of syntax. Their experiments demonstrate that words in low-probability structures have longer durations than words in high-probability structures. I believe their claim to be too strong, as the experiments conducted in these papers do not fully disambiguate between probabilites over syntactic structures, and probabilities over specific lexical items within those structures. However, their experiments do lend insight into the mechanism by which syntax-level probability might influence phonetic details. In particular, they suggest that syntactic probability influences the properties of prosodic boundaries, which in turn influences the duration of words at those boundaries.

In section 6.3, I will propose an explicit hypothesis about the relationship of syntactic probability to prosodic structure, supported by the results of an experiment. This hypothesis is the Structural Probability-Boundary Strength hypothesis (SPBS), and is based on the Left Hand Side/ Right Hand Side Boundary hypothesis (LRB) of Watson and Gibson (2004). The LRB hypothesis states that prosodic boundary strength is affected by the length (in number of phonological phrases) of both the previous syntactic constituent and the the upcoming syntactic constituent. The length of the upcoming constituent matters because it is being planned as the prosodic boundary is being produced - since longer structures take longer to plan, they will incur a greater slowdown at the prosodic boundary. The SPBS states that prosodic boundary strength is additionally affected by the probability of the upcoming structure. Low-probability structures are harder to plan than higher probability structures, and therefore incur a greater slowdown at the prosodic boundary where they are being planned. Section 7 proposes that low probability structures are harder to plan than higher probability structures because the production system gets more 'practice' with high probability structures. This process of implicit learning can occur over a very short period of time, but persist over the next several sentences.

The sections of the paper are as follows: I will first review the results of three previous studies which indicate some relationship between syntactic probabilities and phonetic duration. These studies indicate that some relationship exists, but do not provide sufficient evidence for durational effects which are completely independent of particular lexical entries. I will then report an experiment that demonstrates that speakers use explicit syntactic probabilities in production and disambiguates between probability over syntactic structures and lexical items, as well as between syntactic complexity and syntactic probability. Finally, I will describe in detail the augmented Left Hand Side/ Right Hand Side Boundary Hypothesis, and discuss its predictions.

2 Background

Much recent research has focused on predictability as a solution to the problem that the extreme phonetic variability of natural speech poses to the human comprehension system. Lindblom (1990) put forth the notion of *sufficient discriminability*, which words achieve through balancing phonetic information with contextual information. When the context is very predictive, talkers can produce very phonetically uninformative utterances and still achieve sufficient discriminability. However, when the context is uninformative, the phonetic details of the utterance must be more precise. Further work along these lines includes Jaeger (2010); Aylett and Turk (2004, 2006); Turk (2010); Aylett (2000). All of these authors provide evidence that the information content of speech, or the amount of redundancy in speech, is uniform. That is, there is a trade-off between the amount of information provided by the context, and the amount of information provided by the phonetics. Aylett (2000) calls this hypothesis the Smooth Signal Redundancy Hypothesis (SSRH). The idea is that utterances display a constant amount of total redundancy, or information, over time. This means that in a higly predictive context, since the context has high redundancy, the phonetics has correspondingly low redundancy, and vice-versa.

The works cited above focus on how smooth signal redundancy is useful to the perception system. However, they do not address the question of what mechanism the *production* system uses to produce differences in amount of phonetic detail based on probability.

2.1 The production system and probability

One possible hypothesis about the mechanism that the production system uses to differentiate phonetically between more- and less-probable words and structures is what I will call the "Theory of Mind" hypothesis. Under this hypothesis, the production system has access to detailed predictions about the state of the listener's comprehension system, and uses those to lengthen words just in case it would help the listener. Put in terms of Smooth Signal Redundancy, the production system would maintain smooth signal redundancy by explicitly calculating contextual redundancy, crucially taking into account the point of view of the listener, and adjusting phonetic redundancy accordingly.

Although an account like this fits in well with a theory of the relationship between prosodic prominence and information structure (Brown, 1983; Bell et al., 2003; Bell, 1984), studies such as Bard et al. (2000); Bard and Aylett (2000) indicate that talkers do not modulate their reductions based on the information available to the listener. In particular, they found that talkers reduce previously mentioned words even in the presence of a new listener, who has not yet heard those words. Related work has argued that a production-based account of predictability-related phonetic reduction better explains experimental data (Lam and Watson, 2010; Kahn and Arnold, 2011; Rosa et al., 2011).

In short, it is an open question to exactly what degree talkers design their productions to be easy to comprehend, but it is unlikely that the production system employs something like explicit"Theory of Mind" to calculate contextual redundancy. What then might cause the production system to vary word durations based on contextual predictability? This is an especially interesting question in the domain of probability over syntactic structures, for two reasons. First, if the probability of a structure, independent of the probability of the words in it, can really influence fine-grained phonetic details, the production system must have a mechanism through which these two quite distinct levels of representation can interact. Second, the probability of syntactic structures is highly correlated with other aspects of those structures, such as their length and their complexity, which have been independently shown to influence phonetic details of production (Ferreira, 1991; Maner et al., 2000). This raises the question: Is probability over syntactic structures really causal in the production system? That is, does the probability of a structure exert its own influence on the production process, regardless of the structure's length and complexity?

2.2 Previous experiments on syntactic probability and duration

Gahl and Garnsey (2004, 2006) argue that grammatical knowledge contains explicit information about probabilities over structures. Gahl and Garnsey (2004) attempted to vary the probability of a syntactic structure independently of (a) the identity of the words in that structure, (b) the bigram transitional probabilities of those words, or (c) the real-world plausibility of the sentences. To do this, they used verbs with varying subcategorization biases. They were able to use the same structure as the complement of a verb that frequently takes that complement type, and as the complement of a verb which rarely takes that complement type. Thus, they observed the same structure in a context where it was highly predictable, and in a context where it was unpredictable but still grammatical. Additionally, for a given verb, they compared the same string of words with two different structures. In this way, they held any properties of the particular lexical items in the structure constant across conditions, as well as any word-to-word transitional probabilities. Example stimuli are given below.

(1) Verbs like 'decide' have a strong subcategorization bias for a particular type of object:

- **Clausal complement bias:**

  Matching: The experienced judge <u>decided the appeal</u> should be started right away.

  Mismatching: The experienced judge <u>decided the appeal</u> on the merit of the case.

- **DP complement bias:**

  Matching: The journal editor <u>printed the article</u> with the footnotes at the end.

  Mismatching: The journal editor <u>printed the article</u> had been slanderous to him.

They elicited productions of these sentences by instructing participants to read through each sentence silently first until they understood it, and then to say it aloud as naturally as possible. They measured the durations separately of the matrix verb and the following NP (the underlined words). They found that words in bias-mismatching contexts (low probability structures) were longer than their counterparts in bias-matching contexts. Whether that difference showed up on the NP or on the verb varied with bias type. For clausal complement biased verbs, it was the following NP that was longer in the mismatching context (a DP-complement) than in the matching context (an SC complement). For DP-bias verbs, it was the verb that was longer in the mismatching context (an SC complement) than in the matching complement (a DP complement). They conjecture that the two bias types behave differently because duration is being manipulated just at unpredictable prosodic boundaries.

DP-complement structures and SC-complement structures have different prosodic structures in addition to different syntactic structures. DP-complement structures have a prosodic boundary after the DP, while SC-complement structures have a boundary after the verb. When a DP complement structure occurs in an SC-bias situation, the prosodic boundary after the DP is unexpected, and is therefore implemented with more phrase-final lengthening than when it occurs in a DP-bias situation, where it is expected. When an SC complement structure occurs in a DP-bias situation, the prosodic boundary after the verb is unexpected, and therefore undergoes more phrase-final lengthening than it does when it occurs in an SC-bias situation and is therefore expected.

A similar result was found by Gahl et al. (2006). In this case, verbs were biased to be transitive or intransitive, and the structures being compared were either (a) a new clause or (b) a DP-complement to the verb.

(2) Verbs are biased to be transitive or intransitive:

- **Intransitive:**

  Matching: Although the storyteller continued, the story had lost its appeal.

  Mismatching: Although the storyteller continued the story, it had lost its appeal.

- **Transitive:**

  Matching: Soon after the candidate accepted the money, it was found to be illegal.

  Mismatching: Soon after the candidate accepted, the money was found to be illegal.

The verb and a following DP, which is either its complement or the start of a new clause, remain lexically identical. For intransitive-bias verbs, the DP was longer in the bias-mismatching context than in the bias-matching context. For transitive-bias verbs, it was the verb which was longer in the bias-mismatching context than in the bias-matching context. In sentences with intransitive verbs ('Early Closure'), there is a prosodic boundary after the verb, which is unexpected in transitive-bias contexts, and therefore implemented with extra phrase-final lengthening. In sentences with transitive verbs ('Late Closure'), there is a prosodic boundary after the DP, which is unexpected in intransitive-bias contexts, and therefore implemented with extra phrase-final lengthening.

These two studies demonstrate that a structure's probability as a complement to a particular verb can influence the duration of words in that structure. Since they controlled for the identity of the words, they conclude that the relevant probabilities are really over syntactic structure, and not over particular strings of words, or n-gram transitional probabilities (though see below for a problem with this conclusion). If it was a particular word's lexical frequency which was influencing its own duration, that effect should be visible in both the probable and the improbable structures, since all lexical items appeared in both. Likewise, the same group of words in the same order appears in both probable and improbable structures, making any slowdown due to low transitional probabilities between the lexical items identical across conditions. Only the probability of the structure itself differs across conditions, and thus only that probability can be responsible for the duration differences.

Their design also rules out any kind of 'articulatory practice' explanation for their results (Bybee and Hopper, 2001), in which words are pronounced with shorter durations when the articulatory system has had more practice with them. Although the production system may get more practice with more frequent structures (see Section 7), this practice is not related to articulation. The actual words, and therefore the phones, in the structure differ across instances of it. Gahl and Garnsey (2004) also conducted a norming study, asking participants to judge the real-world plausibility of the sentences. The real-world plausibility of a sentence did not affect the duration of the critical words.

Both of these studies also found lengthening only at prosodic boundaries, and not elsewhere. This suggests that prosodic structure may play an important role. They conclude from these studies that grammatical knowledge of syntax includes explicit knowledge of probabilities over syntactic structures, but they do not propose a theory of how these probabilities affect phonetic implementation during production. They also do not propose a theory which explains the role of prosodic structure.

Tily et al. (2009) followed up on this work by examining double-object structures in a spoken corpus (the switchboard corpus, Godfrey et al., 1992). They examined structures like in (3), in which essentially the same meaning can be expressed as two different syntactic structures.

(3) Double object structures:

a. **NP NP:** The man gave the girl a book.

b. **NP PP:** The man gave a book <u>to</u> the girl.

As their predictability measure, they used a series of factors identified by Bresnan et al. (2007). They focused on NP PP structures, and for each example of such a structure in a corpus, they calculated its probability according to the factors found by Bresnan, Cueni, Nikitina, and Baayen (2007). They compared the duration of the word 'to' in very low-probability NP PP structures to its duration in high-probability NP PP structures, and found that it was significantly longer in the low-probability structures than in the high-probability structures.

One possible confound with the studies in Gahl and Garnsey (2004); Gahl et al. (2006) is their task of planned production. Although they tried to simulate normal speech production by asking the subjects to read through the sentence and understand it first, it is possible that the duration effects they found were actually a product of participants' also having to comprehend the sentences online as they were reading them. This corpus study suggests that this is not the case, since it finds similar effects in spontaneous speech, where there is no issue of interference from online comprehension. This corpus study is also interesting in that it finds a duration difference which is not at a prosodic boundary (there is no prosodic boundary at the right edge of the word 'to'), suggesting that perhaps lengthening of words in unpredictable syntactic structures is not restricted to prosodic boundaries.

2.3 Towards a theory of production mechanisms

The three studies summarized above demonstrated that phonetic duration of words can change as a function of the predictability of the syntactic structure they are situated in. Gahl and Garnsey claim that this shows that there is explicit representation of probability in the syntactic grammar, but they do not distinguish this situation from one in which syntactic probabilities are only affiliated with particular lexical items. That is, upon looking up that lexical item (say, 'claim') speakers access one of two lexical 'subentries' which subcategorize for different complement types and which vary in frequency. In this case, the duration differences that Gahl and colleagues found could be directly related to lexical frequencies, and not related to syntactic probability as such.

Syntactic structures differ in probability independently of any immediate or lexical context. An open question, then, is whether or not these differences in probability lead to differences in duration. The present experiment attempts to address this question by using two structures which differ in overall frequency in a corpus (Roland et al. (2007)). These are subject-extraction and object-extraction cleft sentences.

(4) a. **Subject-extraction (SE):** It was Edward who $_{(t)}$ scammed Melvin out of some money.

b. **Object-extraction (OE):** It was Edward who Melvin scammed $_{(t)}$ out of some money.

Subject- and Object- extraction cleft sentences are very similar syntactically, and differ only in gap position. However, subject-extraction is more frequent in a corpus than object-extraction. In addition to being less frequent, object-extraction is also more complex in that it contains a longer filler-gap dependency. Evidence for thinking of sentences with longer dependencies as more syntactically complex comes from experiments with relative clauses, showing that object extraction is generally more difficult to process than subject-extraction (Gibson, 2000; Warren and Gibson, 2002, among many others).

Because they tend to take more effort to produce and comprehend, more complex structures will necessarily be rarer in spontaneous speech than less complex ones. If words in object-extraction sentences are longer than those same words in subject-extraction sentences, this will not be enough to argue that the production system uses some explicit representation of probability. A clear alternative presents itself: Words in object-extraction sentences are longer than those same words in subject-extraction structures because object-extraction sentences are more difficult to produce, being more complex.

In order to avoid this confound, I manipulate the probability of each structure locally within the experiment. One group of subjects encountered many examples of object-extraction, and few examples of subject extraction, while a different group encountered many examples of subject extraction and only a few examples of object extraction. Thus, the first group was exposed to a context in which object extraction was very probable and subject extraction improbable. The second group was exposed to the opposite: subject extraction was very probable and object extraction was improbable. If a structure's probability is represented distinctly from its complexity in the production system, and that representation can be updated in a fairly short amount of time, we would expect the groups to behave differently according to how probable each structure was in the recent set of sentences they were exposed to.

The group for which object extraction is highly probable should produce words in object extraction structures shorter than those same words in subject extraction structures. The group for which subject extraction is highly probable should do the opposite: they should produce words in subject extraction structures shorter than those same words in object extraction structures. If the groups do not differ in their behavior, then we can conclude that probability is not explicitly represented in the production system, and the complexity of a structure is a better predictor of the duration of words in it. An alternative explanation would be that probability is explicitly represented in the production system, but that it is relatively static, and cannot be updated in a short period of time.

An augmented version of the Left Hand Side/ Right Hand Side boundary hypothesis (Watson and Gibson, 2004) (called the "Structural probability - boundary strength hypothesis", detailed in section 6.3) predicts that prosodic boundary strength is affected by the probability of upcoming syntactic structure. Since duration is a strong phonetic correlate of prosodic boundary strength (Wightman et al., 1992), this hypothesis predicts that the duration of words at prosodic boundaries preceding low-probability structures should be longer than the duration of those same words at boundaries followed by high-probability structures.

3 Methods

3.1 Participants

The experiment was conducted at the University of Colorado at Boulder. There were 21 subjects, all residents of Boulder or participants in the LSA summer institute.

3.2 Items

The two structures compared in this experiment were subject extraction and object extraction clefts. Roland et al. (2007) found that extraction out of relative clauses was much more likely to be subject-extraction than object-extraction when the extracted constituent was animate. For cleft structures themselves, they found that subject extraction was numerically more common, but that both were extremely rare, such that the difference between them was not statistically significant. I assume that the same animacy facts that constrain the predictibility of relative clause extracton are also relevant for clefts. In all items, the extracted constituent was animate.

(5) a. **Subject-extraction (SE):** It was Edward who $_{(t)}$ scammed Melvin out of some money.

b. **Object-extraction (OE):** It was Edward who Melvin scammed $_{(t)}$ out of some money.

All items followed the form 'It was DP_1 who (DP_2) Verbed (DP_2) PP'. Some items has more than one prepositional phrase after the verb or DP_2, and they varied in whether they were an argument of the verb or not. The complete list of items (16 total) is given in the appendix. All sentences were presented visually, along with a context sentence of the form 'Did DP Verb DP PP ?', where the Verb and PP were the same as those in the target sentence, and one DP was DP_2 of the target sentence, while the other was not in the target sentence.

(6) **SE:**
Did John scam Melvin out of some money?
It was Edward who scammed Melvin out of some money.

(7) **OE:**
Did Melvin scam John out of some money?
It was Edward who Melvin scammed out of some money.

The questions were formulated so as to introduce all parts of the target sentence except the extracted constituent, so that they would be old information in the target sentence. This was to encourage all participants to put narrow focus on the extracted constituent, and no focus on the rest of the sentence, reducing variability in the duration of both the extracted constituent and DP_2.

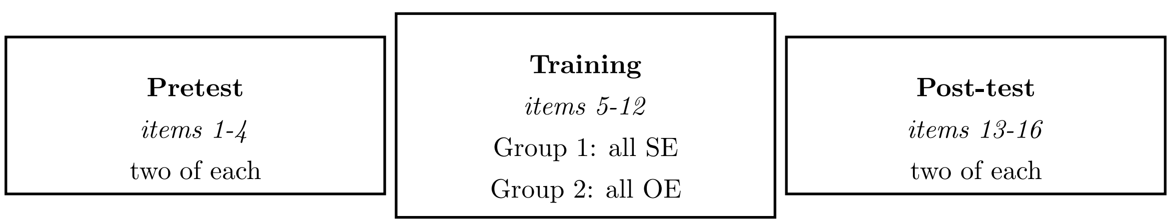
3.3 Procedure

Participants were instructed to read both sentences silently to themselves until they understood what they mean, and then to read the second aloud as naturally as possible, as if they were participating in a real conversation. They saw the two sentences on a computer screen, studied them for as long as they liked, and then pressed a button to start recording, and pronounced the target sentence.

Items were presented in three blocks: a pretest block consisting of four items, a training block consisting of eight items, and a test block consisting of four items. The task in each block was identical, and participants were not informed in any way that the experiment was divided into sections. Each participant saw the same items in each of these blocks, but each participant saw a different random order of items within each block. Items were counterbalanced, so that each participant saw each item in only one condition.

The pretest and test blocks each consisted of two items in the OE condition, and two in the SE condition. The training block consisted of eight items all in the same condition. Participants were randomly assigned to two groups, one which produced all SE sentences in the training block, and one which produced all OE sentences in the training block. A schematic of this layout is given in figure

Figure 1



3.4 Analysis

Participants' utterances were recorded with MatLab. The Prosodylab-Aligner (Gorman et al., forthcoming) was used to segment out words of interest. After running the forced aligner, the word-edge alignments were checked by hand. If the Aligner placed a boundary in the middle of a vowel, as it often did when the vowel was voiceless (in the word 'who') or creaky, the boundary was moved to the end of voicing, or the end of clear formant structure when the vowel was followed by a voiced segment. If the Aligner placed a boundary in the middle of a nasal, the boundary was moved to the edge of the nasal, identified by a sudden change in amplitude visible in the waveform, or a sudden change in formant values. The aligner occasionally mistook aspiration of a [t] for an [s] when they occurred next to each other, and in this case, the boundary was simply moved according to where the stop closure of the t occurred. A very few adjustments were made based entirely on aural judgments.

The duration of each word of interest was extracted using Praat (Boersma and Weenink, 2011). Words of interest were (1) the clefted constituent, (2) the word 'who', (3) the verb, (4) the second DP.

(8) It was $\underbrace{\text{Edward}}_{1}$ $\underbrace{\text{who}}_{2}$ $\underbrace{\text{scammed}}_{3}$ $\underbrace{\text{Melvin}}_{4}$ out of some money.

Several utterances contained disfluencies, and several utterances did not match the target sentence given to the participant. When participants produced an utterance which was fluent, but did not match the target, their utterance was identical to the target utterance, but of the opposite extraction type. All but one of these were object extraction targets pronounced as subject extraction. All disfluencies occurred on object extraction sentences.

Counts of each type of error were recorded for each condition, but those utterances were removed from the duration analysis.

4 Results

4.1 Duration differences

Figure 2 shows the durations of the four words of interest in each extraction type and in each training condition. Bar heights represent mean durations over items and subjects, and error bars signify standard errors.

4.1.1 The clefted constituent

In the pretest phase, before training altered the local probabilities of each extraction type, the clefted constituent was longer in object-extraction structures than in subject-extraction structures. A linear mixed effects model was fitted to the duration data from the pretest portion of the experiment, with extraction type as a factor and random intercepts for subjects and items. Extraction type was a significant predictor of clefted constituent duration, with $t = 2.4$, and $p = .02$. This duration difference could reflect a probability difference between the two extraction types, but could also reflect the difference in complexity between them: the object-extraction structure has a longer movement dependency than the subject-extraction structure.

After training, the clefted constituent's duration differed across training groups. Subjects trained on object extraction produced no duration difference between subject-extraction and object-extraction structures, while subjects trained on subject extraction produced the clefted constituent longer on object extraction structures than on subject extraction structures. A linear mixed effects regression was conducted on the post-training durations, with random intercepts for subjects and items. The interaction between condition and training type was significant ($\beta = .065$, $p = .04$). There was no main effect of training type ($\beta \approx 0$, $p \approx 1$), or of extraction type ($\beta = -.004$, $p = .83$).

The difference in duration of the clefted constituent between the subject-extraction and object-extraction structures after SE training was numerically greater than the difference between subject- and object-extraction structures before training, in the pretest potion of the experiment - the difference in means after SE training was 80ms, while the difference in means in the pretest was only 37ms. That the duration difference became more substantial after SE training is also reflected in the regression coefficients: 0.065 for the interaction after training vs. 0.038 for extraction type before training.

In order to statistically compare the duration difference between subject- and object-extraction structures in the pretest with the post-tests, a separate linear mixed effects model was fitted to the data from each participant group. Each model had extraction type, training condition (before or after training), and their interaction as factors, and had random intercepts for subjects and items. In the subject-extraction training group, there were no significant main effects or interactions (Training condition: $\beta = -0.02$, $p = .37$; Extraction type: $\beta = 0.02$, $p = .48$; Interaction: $\beta = 0.06$, $p = .28$). In the object-extraction training group, there was a significant main effect of extraction type ($\beta = 0.05$, $p = 0.1$), as well as a significant main effect of training condition - overall durations decreased in the later part of the experiment ($\beta = 0.05$, $p = 0.03$). The interaction between extraction type and training type approached significance ($\beta = -0.06$, $p = 0.06$).

4.1.2 The Verb

In the pretest phase, the verb's mean duration was longer in object-extraction structures than in subject-extraction structures, but this different was not significant ($\beta = 0.050$, $p = .16$). After training, the verb's duration patterns differed between the two training conditions: after SE training, the verb's duration did not differ between extraction types. After OE training, the verb's duration was numerically longer in subject-extraction structures than in object-extraction structures. A linear mixed effects model (again with random intercepts for subjects and items) yielded no significant main effect of training type ($\beta = .0094$, $p = .7$), but a marginally significant main effect of extraction type ($\beta = -.048$, $p = 0.08$). There was no significant interaction ($\beta = 0.036$, $p = 0.38$).

Although the interaction term is not significant, it is clear from the graphs that the near-significant main effect comes primarily from the OE training condition. One reason for the large difference in verb's duration after OE training and in the pretest may be a difference in prosodic structure due to the difference in syntactic structure itself rather than a difference in the probability of the structures.

In items like (9) below, the verb is at the right edge of an XP just in the object-extraction condition, and not in the subject-extraction condition. Being at the right of an XP makes the verb much more likely to be at the edge of a phonological phrase boundary according to Match theory (Selkirk, 1986, 1995, et. seq.)

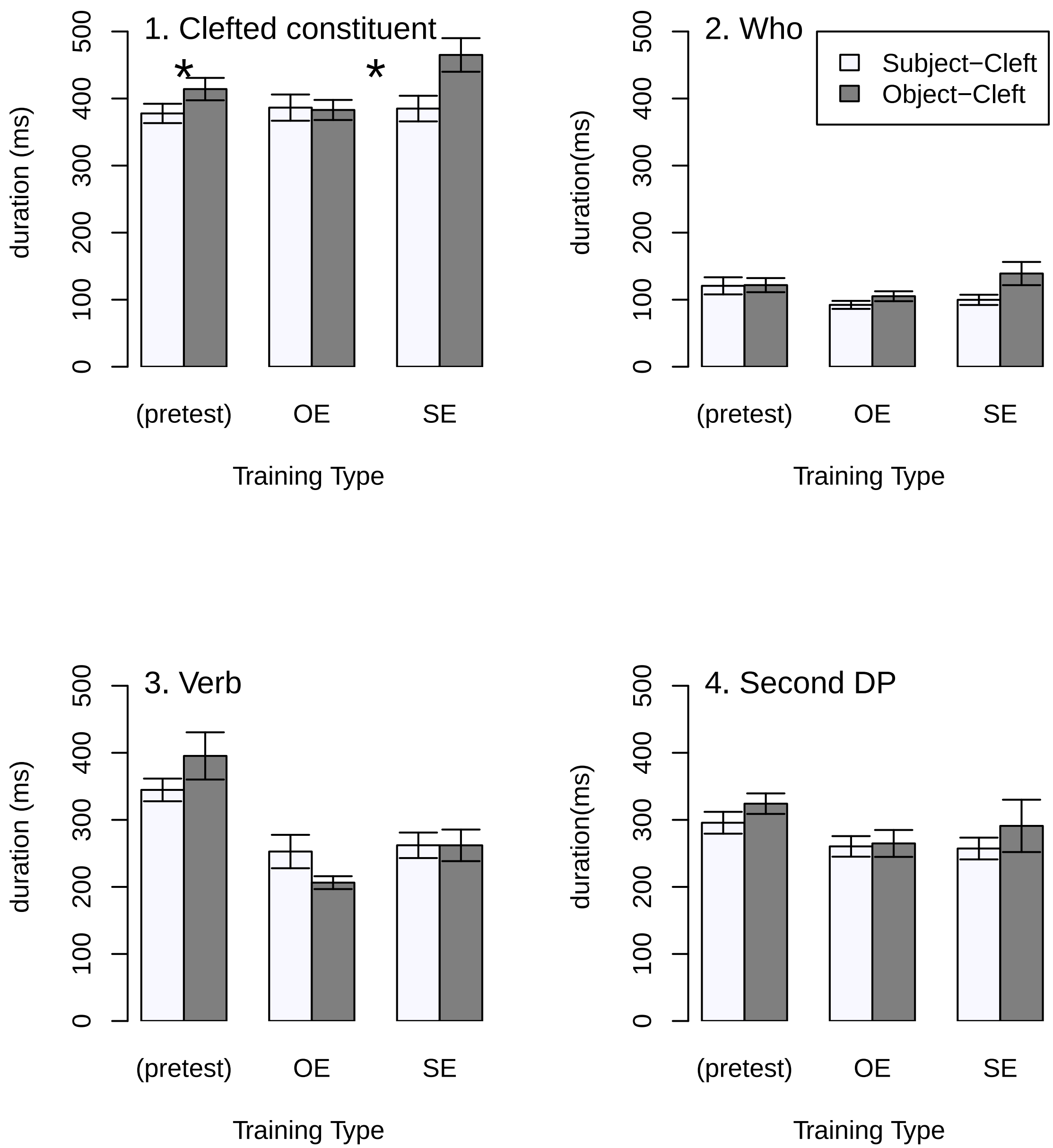
(9) a. **(SE)** It was Kim who (t) lost Jim $]_{\mathrm{VP}/\phi\mathrm{P}}$ down at the lake.
b. **(OE)** It was Kim who Jim lost (t) $]_{\mathrm{VP}/\phi\mathrm{P}}$ down at the lake.

In other items, such as (10), the material after the verb is an argument of the verb rather than an adjunct, and therefor the verb is never the rightmost item in an XP:

(10) a. **(SE)** It was Rebecca who (t) sent sam back home $]_{\mathrm{VP}}$.
b. **(OE)** It was Rebecca who Sam sent (t) back home $]_{\mathrm{VP}}$.

Of the four sentences in the pretest block, half were like (9) and half were like (10). In the post-training test block, three were like (10), and one like (9).

Figure 2: Mean durations by extraction type and training type. Error bars indicate standard errors of the mean.



These prosodic breaks explain the duration difference before training, but not the duration difference after OE training. Before training, the verb is longer on average in object-extraction structures than in subject-extraction structures. It is also in object-extraction structures that the verb is half of the time at a prosodic phrase boundary, while it is never at a prosodic phrase boundary in subject-extraction structures.

The verb may be longer in object-extraction structures in this case just because it is more often at a prosodic break.

The duration differences after OE training are not explained by the occasional prosodic breaks after the verb. The verb is actually *shorter* in object-extraction structures than in subject-extraction structures, an effect we would not expect if it were at a prosodic break in object-extraction and not at a prosodic break in subject-extraction structures. A relevant property of the verb in subject-extraction structures is that it is the first word which disambiguates the extraction type, and forces a listener to conclude that they are hearing a subject-extraction structure. It does not serve this role in an object-extraction structure, since a DP intervenes between the complementizer and the verb, and indicates that the structure must be an object-extraction structure. In subject-extraction structures after OE training, the verb becomes the disambiguating element for a low-probability structure.

The experimental design encouraged participants to fully understand each sentence before producing it, in order to make the task as much like natural production as possible. However, this attempt may not have been successful. Subjects may have actually been comprehending the sentences online while producing them, which might result in a slow-down at the disambiguating elements.

4.1.3 Other elements

The durations of the word 'who' and the second DP in each sentence were also measured, but do not show any clear effects. Linear mixed effects models were again fitted with random slopes for subjects and items. The word 'who' was not longer in object- than subject- extraction before training ($\beta = 0.008$, $p = .6$), but its duration was numerically longer in object-extraction sentences than subject-extraction sentences after both kinds of training. There were no significant main effects (Training type: $\beta = 0.007$, $p = .58$; Extraction type: $\beta = 0.011$, $p = .2$). The interaction approached significance ($\beta = 0.025$, $p = .08$). The noun's duration was numerically longer in object-extraction sentences than in subject-extraction sentences both before training ($\beta = 0.021$, $p = .29$, n.s.) and after both kinds of training. Main effects and the interaction term were not significant (Training type: $\beta = -0.003$, $p = .89$; Extraction type: $\beta = 0.018$, $p = .81$; Interaction: $\beta = 0.018$, $p = .66$).

4.2 Accuracy differences

Throughout the experiment, subjects made errors on object-extraction structures (and just once on a subject-extraction structure). These errors included disfluencies and incorrect productions. The incorrect productions were always a subject-extraction structure when an object extraction structure was presented, for example 'It was Edward who scammed Melvin out of some money.' when the target sentence was 'It was Edward who Melvin scammed out of some money.' After SE training, there were significantly more disfluencies than there were after OE training ($\chi^2 = 9.8$, $p = .001$).

4.3 Subject and Item effects

Subjects differed in their overall speaking rate, and the critical words in individual experimental items differed in length. It is possible that the length of the word, or the speaking rate of the individual speaker play a

Figure 3: Errors by training condition and error type

	Disfluency	SE instead of OE	(out of)
SE training:	7	9	24
OE training:	0	3	20

role in the degree of duration difference between probable and improbable structures. Individual speaking rates were calculated for each participant using the critical words (the clefted constituent, the word 'who', the verb, and the second DP) from all utterances produced by that subject, including the utterances from the training portion of the experiment. However, there was no clear difference in the behaviour of subjects with faster speaking rates and subjects with slower speaking rates, and no clear correlation of speaking rate with amount of duration change.

There were only eight total distinct words used as clefted constituents, all but one of which were two-syllables long. They contained various numbers of segments however. Again, no clear difference was found between particular lexical items. Graphs illustrating mean durations for each extraction type in each training condition broken down by subject, and by item can be found in Appendix B.

5 Discussion

The clefted constituent was initially longer in object-extraction structures than in subject-extraction structures. After the probability of each extraction type was manipulated through training, the clefted constituent's duration behaved differently depending on which structure was the more probable one. When object-extraction was more probable than subject-extraction, the clefted constituent's duration was no different in the two structures. However, when subject-extraction was more probable than object-extraction, the clefted constituent was much longer in the improbable object-extraction structures than in subject-extraction structures.

The clefted constituent's duration is modulated by the probability of the cleft structure it occurs in. This probability is not related to or conditioned by any particular lexical items in the structure. It is also independent of the syntactic complexity of the structure, in particular, it is not conditioned by the length of the movement dependency. In the experiment, the probability of each structure was manipulated independently of its extraction type. In the object-extraction training condition, the more common structure was the more complex one, while in the subject-extraction training condition the more common structure was the less complex one. The clefted constituent was consistently longer in the low-probability structure rather than in the complex one.

The duration of the clefted constituent depended on the local probability of the structure within the experiment, rather than on some 'global' probability calculated over all cleft sentences to which a subject has been exposed over the course of his or her life. Although participants presumably entered the experiment with some prior expectation of the relative probabilities of the two cleft types, based on a large number of sentences encountered over a long period of time, this expectation was overwritten during the training block of the experiment. The training block was only eight sentences long - a relatively tiny number. This suggests that the process of planning upcoming syntactic structure may be affected by structural probabilities calculated over recently encountered structures, or else structures particularly relevant to the speech context (in this case, the experimental setting). This result makes the alternative, that the planning process depends on structural probabilities calculated over all sentences percieved and/or produced in the individual's lifetime, unlikely.

No duration differences were found on the word 'who' or the noun. This demonstrates that the duration changes driven by structural probability are localized on particular words, and are not properties of the whole sentence. Since duration is known to correlate strongly with prosodic boundary strength (Wightman et al., 1992), I will hypothesize that there is a prosodic boundary following the clefted constituent whose strength is being modulated by the probability of the upcoming syntactic structure, while there is no such boundary after the word 'who' or the noun.

As expected, the verb's duration did change with the probability of the structure. This might have been because in some items there was a prosodic boundary next to the verb, or it may have been a result of participants' comprehending the sentences online as they were producing them.

6 The mechanism

In this experiment, the probability associated with a syntactic structure is influencing fine-grained details of the phonetic implementation. In this section I will present the "Structural probability-boundary strength" hypothesis, which describes the mechanism by which the process of syntactic structure-building in production affects the process of articulatory planning.

The Structural Probability-Boundary Strength hypothesis (SPBS) is an extension of the Left hand side/Right hand side Boundary hypothesis (LRB) (Watson and Gibson, 2004). The LRB states that the probability that an intonational boundary will occur at a word boundary is positively correlated with the size of the upcoming constitutent. They attribute this correlation to the production system, saying that the more difficult something is to plan, the more likely the speaker will be to place a prosodic boundary before it, in order to 'buy time' for planning that structure.

Additionally, the SPBS states that the probability of the upcoming syntactic structure is inversely correlated with the size of a prosodic boundary. The less probable the upcoming structure, the larger the boundary. This correlation is likewise attributed to the production system on the grounds that high-probability structures are easier to plan than low-probability structures. Higher-probability structures may become easy to plan through a process of implicit learning, which can take place over a short period of time (as demonstrated in the experiment), but whose effects can persist, at least as long as the next several sentences.

The SPBS explains the relationship of syntactic probability to the phonetic detail of duration in the following way: The strength of a prosodic boundary at a word's right edge directly affects the duration of that word - the larger the boundary, the longer the word's duration. The boundary's strength is itself affected by the probability of structure currently being planned. Therefore, any lengthening of a word's duration because of the low probability of an upcoming structure occurs only when that low probability has a chance to affect the strength of a prosodic boundary. This explanation accounts for the fact that in the experiment, only the duration of part of the sentence - the clefted constituent - was affected by the structural probability. If we suppose that the clefted constituent is usually followed by some sort of prosodic boundary, and furthermore that it is at this boundary that planning of the rest of the sentence happens, then we would expect the duration of the clefted constituent, and only that duration, to be affected by the probability of the rest of the sentence's structure.

6.1 The LRB hypothesis

Watson and Gibson's (2004) hypothesis states simply that the probability of an intonational phrase boundary between two words is a function of (a) the size of the left-hand, or preceding constituent, and (b), the size of the right hand, or upcoming constituent. As the size of each constituent increases, the probability of a boundary increases.

They measure the size of both preceding and upcoming constituents in terms of the number of phonological phrases. They suggest a definition for phonological phrase as 'all the words within the maximal projection of a lexical head on the lexical head's left side,' but point out that the exact definition is not important to their general claim. They also take into account the semantic relatedness of immediately preceding and upcoming material.

(11) (Watson and Gibson, 2004, pg. 732) LHS/RHS boundary (LRB) weight: The LRB weight at a word boundary between w_1 and w_2 is defined to be the sum of

  a. the size of the LHS constituent terminating in w_1, in terms of phonological phrases
  b. the projected size of the RHS constituent in phonological phrases starting at w_2, if this is not an argument of w_1;
  c. 1, if w_1 marks the end of a phonological phrase.

This definition of weight is predictive of the probability of occurrence of an intonational phrase boundary:

(12) The LRB weight hypothesis: The LHS/RHS boundary weight is proposed to be correlated with the probability of producing a boundary at a given location.

They point out that if planning is incremental, the size of the RHS constituent might not be fully known at the point of producing w_1. Therefore, what counts as the RHS constituent is simply the amount of sentence which has been or is being planned at the time that w_1 and w_2 are being produced. They propose, based on Ferreira (1991), that the reason that the size of the right hand constituent matters is that longer constituents take longer to plan. Placing an intonational boundary results in lengthening of the word whose right edge coincides with the boundary, and also a pause at the boundary, both of which might allow the speaker a little extra time for planning the upcoming structure.

Ferreira (1991) measured initiation times in the production of memorized sentences, comparing long utterances to short ones (where length was measured in number of prosodic words), and more syntactically complex utterances to less complex ones. She found that initiation times were longer when the sentence was long, and they were longer when the sentence was more complex. More recent work, for example Allum and Wheeldon (2007) has continued to find slower initiation times for longer and more complex structures than for shorter and less complex ones.

6.2 Probability vs. size of an intonational boundary

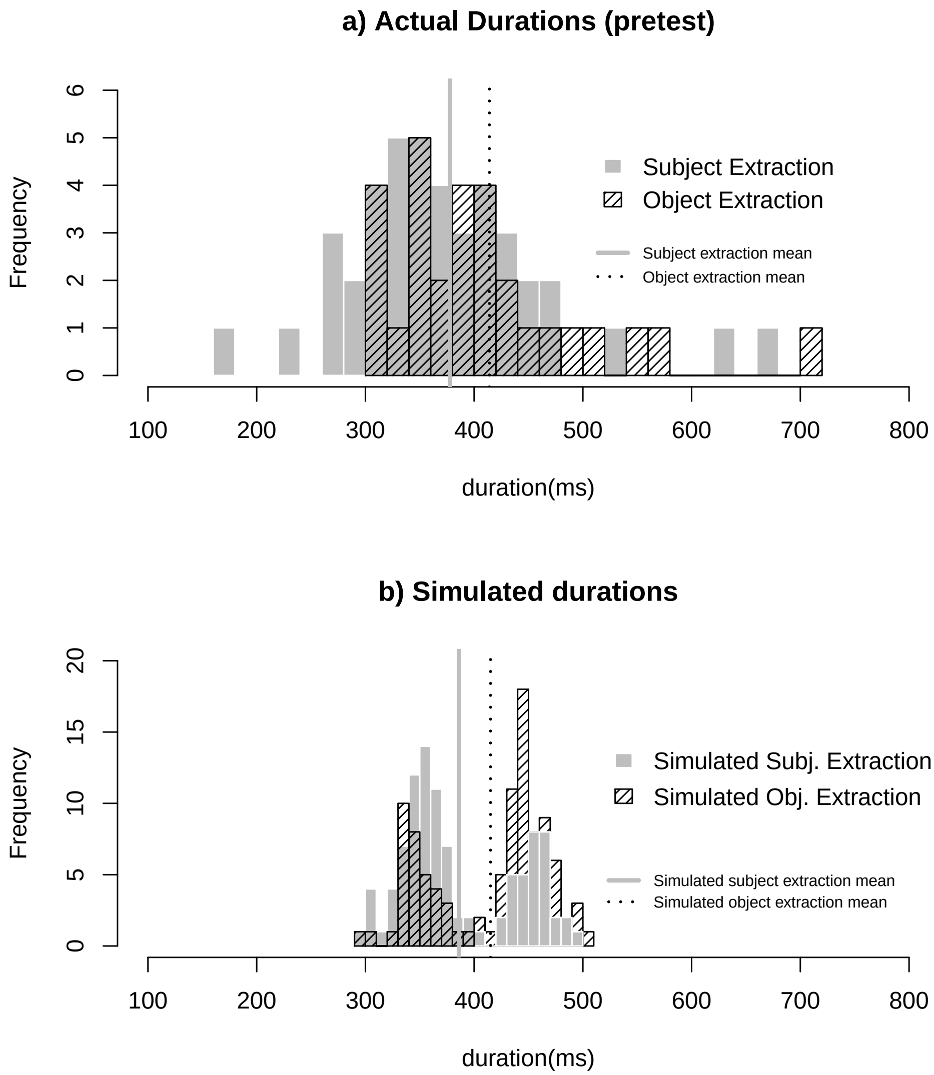
The LRB weight hypothesis predicts the probability of occurrence of an intonational phrase boundary. The boundary is either present or not in a particular utterance, but a boundary is more likely depending on the lengths of the constituents on either side of the boundary. Wagner and Watson (2010) (pg. 911) point out that whether prosodic boundary strength is really gradient or whether it falls into discrete categories is still an open question. They note that although different prosodic boundary strengths are often thought of as categorically different, the acoustic correlates of those boundary strengths tend to vary gradiently. Furthermore, they note that in studies such as de Pijper and Sanderman (1994), subjects are much more reliable at annotating relative boundary strengths than at categorizing boundary types.

In the utterances collected in the present experiment, the duration of the clefted constituent seems to vary gradiently rather than categorically, and therefore it seems that the 'size' or strength of implementation of the prosodic boundary at its right edge is being affected in a gradient way. Consider the histogram in Figure 4 a). This is a plot of all durations of the clefted constituent gathered in the pretest portion of the experiment (about 70 utterances). Grey bars indicate durations gathered from subject-extraction utterances, and dashed bars indicate durations from object-extraction utterances. The distribution of durations from each type of sentence are close to normal, except that the distribution of durations from object extraction structures is shifted to the right relative to the distribution of durations from subject extraction structures.

If the difference between object and subject extraction sentences were the result of differing probability of occurrence of a boundary, we should see a bimodal distribution of durations - one distribution corresponding to utterances with an IP boundary and one corresponding to utterances without an IP boundary. The difference between the two extraction types should be only a different number of utterances in each category. This situation is depicted in Figure 4 b). To produce this figure, two groups of durations were produced by simulation. Clefted constituents in subject extraction structures were assumed to have a 30% chance of being followed by an IP boundary, while clefted constituents in object extraction structures were assumed to have a 70% chance of being followed by an IP boundary. 100 durations were generated for each type of extraction by first generating boundary types according to these probabilities. Next, duration values for IP boundaries were sampled from a normal distribution with a mean of 450ms and a standard deviation of 20ms. Duration values for PP boundaries were sampled from a normal distribution with a mean od 350ms and a standard deviation of 20ms. These means and standard deviations were chosen so that the mean duration associated with each extraction type would be close to its mean duration in the experiment.

This simulation does not prove that the durations observed in the experiment in no way reflect a categorical difference in boundary type, but it does illustrate a clear case of durations differing categorically, and conditions differing only in the number of cases assigned to each category. The results from the experiment do not resemble this clear case. Based on this I propose that what is varying across conditions in the experiment is not the probability of occurrence of an IP boundary, but rather the strength of implementation of that boundary.

Figure 4: The top histogram comes from observed durations of the clefted constituent. The bottom histogram is the result of sampling durations from two normal distributions, and represents an idealized case of what we would expect the histogram of clefted constituent durations to look like if the durations fell into two categories, namely (1) an IP boundary present and (2) an IP boundary not present.



6.3 The Structural Probability - Boundary Strength Hypothesis

Watson and Gibson's LRB hypothesis relates the 'weight' of a boundary between two words to the planning difficulty of the the upcoming structure. In particular, the longer the upcoming structure currently being planned, the greater the LRB weight. The Structural Probability - Boundary Strength Hypothesis (SPBS) extends this proposal to include the probability of the upcoming structure as a factor influencing the LRB weight on the boundary between two words. The lower the upcoming structure's probability, the higher the LRB weight.

(13) Watson and Gibson, 2004, pg. 732, augmented: LHS/RHS boundary (LRB) weight: The LRB weight at a word boundary between w_1 and w_2 is defined to be the sum of

a. the size of the LHS constituent terminating in w_1, in terms of phonological phrases

b. the projected size of the RHS constituent in phonological phrases starting at w_2, if this is not an argument of w_1;

c. 1, if w_1 marks the end of a phonological phrase.

d. **1 minus the probability of the upcoming structure**

To flesh out this proposal a bit: Upcoming syntactic structures take longer to plan the longer they are, and the more complex they are. Additionally, they take longer to plan the less probable they are, where probability is calculated over some recent, relatively short period of time. The need for extra planning time on low-probability structures is fulfilled via prosodic structure building. Specifically, placing a prosodic boundary or increasing the strength of its phonetic realization increases the amount of time available to the production system in which to plan an upcoming chunk of syntactic structure. Therefore, the production system aids itself by placing more boundaries with stronger phonetic implementations while planning long, complex, and low-probability structures.

This proposal leads us to expect that not just duration ought to be affected by the probability of the upcoming structure, but also other aspects of the phonetic realization of a prosodic boundary. These include F0 excursion, pause probability and duration of pauses, intensity, and voice source effects (Wagner and Watson, 2010).

7 Why are low-probability structures hard to plan?

The proposed amendment to the LRB weight hypothesis given above depends on the notion that low-probability structures are in some sense harder to produce than high-probability structures, independently of their length or complexity. I argue here that low-probability structures are difficult to produce relative to high-probability structures because the production system gets more 'practice' with high-probability structures.

Before fleshing out this proposal, I will return to the 'Theory of Mind' hypothesis I discussed in section 2.1. According to this hypothesis, lengthening of low-probability material occurs for the benefit of the listener. Lengthening of a word that is unexpected given the context, for example, gives the listener extra time to process that word and to correctly perceive it. One possible interpretation of the results of this experiment is that the speakers are lengthening the clefted constituent in order to give a potential listener time to process the low-probability upcoming structure.

Hale (2003, 2006) discusses the relation between the processing difficulty of a particular word and the amount of information conveyed by that word about the structure of the sentence. A word that conveys more information to the listener has a higher 'surprisal' value, and processing it incurs a higher cognitive load (as measured by e.g. reading times). It seems unlikely that the clefted constituent is such a high-information word in this experiment. While the clefted constituent does consistently convey new information (the identity of the person acting or being acted upon in the rest of the sentence, which is old information), that information is not syntactic in nature. A subject-cleft sentence and an object-cleft sentence in which the same word has been extracted are identical until after the complementizer.

(14) It was Edward who | $_{(t)}$ scammed Melvin out of some money.
It was Edward who | Melvin scammed $_{(t)}$ out of some money.

If speakers were lengthening high-surprisal words, we might expect duration differences on the material after the word 'who', the verb and the noun. In particular, in the subject-extraction training condition, when subject extraction is more likely than object extraction, the DP encountered immediately after the complementizer ('Melvin' above) in an object-extraction structure would have a high surprisal value compared to the verb encountered immediately after the complementizer ('scammed' above) in a subject-extraction structure. Likewise in the object-extraction training condition, the verb immediately following the complementizer in a subject-extraction sentence would have a high surprisal value compared to the DP encountered immediately after the complementizer in an object-extraction structure.

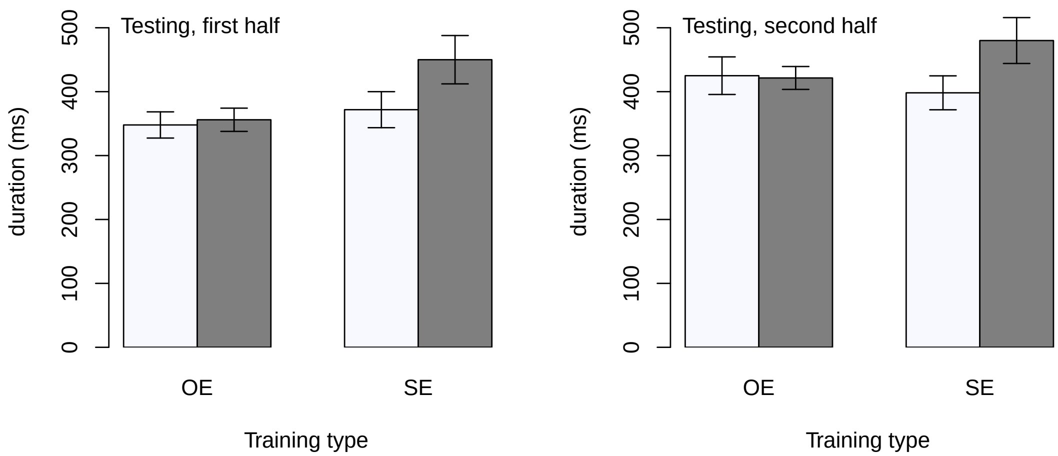
The results of the present experiment provide no evidence that speakers are lengthening high-surprisal words. Rather, they are lengthening (or shortening) the clefted constituent, which is potentially at a prosodic boundary, and therefore at the left edge of some planning domain. The locus of the duration difference in the experimental item suggests that speakers are lengthening (or shortening) for their own benefit rather than for the benefit of the listener.

The error data from the experiment provides additional empirical support for this position. Object-extraction sentences were more likey to have production errors when they were low-probability (in the subject-extraction training condition), than when they were high-probability (in the object-extraction training condition). This indicates that production of a structure is in fact more difficult when that structure is improbable than when it is probable.

Wheeldon and Smith (2003) argue that a process of implicit learning is responsible for some structural priming effects, namely those effects which persist longer than a single experimental item. **?**, among others, report persistent structural priming effects which last several minutes, and over ten experimental items or more. Wheeldon and Smith contrast these persistent effects with short-lived structural priming effects, which last only as long as the very next production and are gone after just one intervening sentence. In the experiment reported here, the effects of the training session persist throughout the entire testing session, over at least two intervening sentences of the less-probable type (given the training).

Figure 5 compares the durations of sentences of each extraction type after each training type from the first half of the testing session with those from the second half. The order of experimental items and what condition they appeared in was randomized, but all subjects saw two sentences whose extraction type matched the one they were trained on, and two sentences with the other extraction type. For most participants, the first two sentences of the testing block consisted of one of each extraction type, as did the second two sentences. The pattern of durations does not differ between the two halves of the testing session, indicating that the effects of the training session persist throughout the entire testing session. If the effects in the present experiment were due to short-lived priming, then the duration differences apparent after training would be observable only on the first two sentences (one object-extraction and one subject-extraction) after training. In the second two sentences after training, the duration differences would be the same as those

Figure 5



found before training. Thus the post-training duration differences appear to be indicative of something like Wheeldon and Smith's implicit learning.

Implicit learning of syntactic structures is modeled by Chang et al. (2000) in a simple recurrent network model. Units in this model represent lexical items and syntactic category and thematic role labels. As the model learns to associate lexical items with syntactic/thematic labels, it adjusts the weights between all the units. Once it has learned how to correctly parse a training set of sentences, it continues to make minute adjustments to the weights between syntactic/thematic units each time it encounters a sentence. The effects of adjusting these weights are long-lasting, over many intervening sentences, thus mimicing long-lived priming effects.

In Chang et. al.'s model, implicit learning of syntactic structures happens through the same mechanism as syntactic acquisition itself. This learning is entirely independent of any social or other context, and is quickly reversed. As soon as the model is exposed to structures different from the one it was primed on, the weights begin to change again. Warker takes issue with this model as a model of long-lasting priming. She conducts a series of experiments on implicit learning of a phonological rule, demonstrating that subjects maintain knowledge of the rule a full week after training. Since the pattern she taught her subjects is not true of English as a whole, it cannot be the case that subjects are simply adjusting weights in their phonological grammar. If this were the case, their experience with normal English outside of the experimental context would completely overwrite the rule they learned in the experiment. Warker proposes a slightly different mechanism, in which learned rules or tendencies can be indexed to particular social contexts.

The results of the experiment presented here could be captured with Chang et. al.'s model, but the speed with which subjects adjusted their durations (and production accuracy) suggests a mechanism which is separate from the mechanism responsible for first-language acquisition - a much slower process. Future work should test the durability of the learning exhibited here (as in Warker's experiments), and also to what degree subjects learn contextually-dependent probabilities over syntactic structures (for example, can subjects learn that an object-cleft is very likely when Talker A is speaking, but very unlikely when Talker B is speaking? Can they learn that an object-cleft is very likely when the computer screen is red, but very unlikely when it is green?).

8 Conclusion

In the experiment reported above, the duration of one part of the sentence (the clefted constituent) was modulated by the probability of the structure it occurred in. This probability was not related to or conditioned by any particular lexical items in the structure - rather the identity of the particular lexical items was held constant across high- and low- probability versions of the structure. It is also independent of the syntactic complexity of the structure. In the experiment, the probability of each structure was manipulated independently of its complexity, through training.

The duration of the clefted constituent in the object- and subject-extraction cleft structures used in the experiment depended on the local probability of the structure within the experiment, rather than on some 'global' probability calculated over all cleft sentences to which a subject has been exposed over the course of his or her life. Although participants presumably entered the experiment with some prior expectation of the relative probabilities of the two types of sentences, this expectation was overwritten in a relatively short period of time during the training block of the experiment.

The Structural Probability-Boundary Strength hypothesis (SPBS) was proposed in section 6.3. This hypothesis claims that the probability of upcoming syntactic structure affects the strength of implementation of a prosodic boundary. Low-probability structures take more time or effort to plan than higher probability structures, and therefore the production system 'buys time' for them by increasing the strength of the prosodic boundary at which they are being planned. Section 7 discusses a process of implicit learning by which the production system gains practice at structures to which it is exposed more often. Implicit learning can occur over a short amount of time (just eight sentences of training in the experiment) but persists longer than short-lived structural priming effects like those discussed in Wheeldon and Smith (2003).

References

Paul H. Allum and Linda R. Wheeldon. Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4):791–810, 2007.

Inbal Arnon and Sam Miller. More than words: phonetic duration is reduced in more frequent chunks. Poster presented at CUNY, 2011.

Inbal Arnon and Neal Snider. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62:67–82, 2010.

Matthew Aylett. Stochastic suprasegmentals: Relationships between redundancy, prosodic structure, and care of articulation in spontaneous speech. In *Proceedings of ICSLP-2000*, Beijing, 2000.

Matthew Aylett and Alice Turk. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47:31–56, 2004.

Matthew Aylett and Alice Turk. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119(5):3048–3058, May 2006.

E. G. Bard and M. P. Aylett. Accessibility, duration, and modeling the listener in spoken dialogue. Gtelog 2000, Fourth workshop on the semantics and pragmatics of dialogue. Gtalog, Sweden: Gothenburg University, 2000.

Ellen Gurman Bard, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon, and Alison Newlands. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42:1–22, 2000.

Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle L. Gregory, and Dan Gildea. Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *Journal of the Acoustical Society of America*, 113:1001–24, 2003.

Allan Bell. Language style as audience design. *Language in Society*, 13:145–204, 1984.

Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. In *http://www.praat.org/*, 2011.

Joan Bresnan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. Predicting the dative alternation. In G. Bourne, I. Kraemer, and J. Zwarts, editors, *Cognitive foundations of interpretation*, pages 69–94. Royal Netherlands Academy of Science, Amsterdam, 2007.

G. Brown. Prosodic structure and the given/new distinction. In A. Cutler and D. R. Ladd, editors, *Prosody: Models and Measurements*, pages 67–77. Springer, Berlin, 1983.

Joan Bybee and Paul Hopper. Introduction to frequency and the emergence of linguistic structure. In Joan Bybee and Paul Hopper, editors, *Frequency and the emergence of linguistic structure*, volume 45 of *Typological studies in language*. John Benjamins, 2001.

Franklin Chang, Gary S. Dell, Kathryn Bock, and Zenzi M. Griffin. Structural priming as implicit learning: a comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29(2):217–229, 2000.

Allan M. Collins and Elizabeth F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975.

J. R. de Pijper and A. A. Sanderman. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, 96:2037–2047, 1994.

Fernanda Ferreira. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30:210–233, 1991.

Susanne Gahl and Susan M. Garnsey. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronounciation variation. *Language*, 80(4):748–775, 2004.

Susanne Gahl and Susan M. Garnsey. Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, 82(2):405–410, 2006.

Susanne Gahl, Susan M. Garnsey, Cynthia Fisher, and Laura Matzen. "That sounds unlikely": Syntactic probabilities affect pronounciation. In *28th Annual Conference of the Cognitive Science Society*, 2006.

Edward Gibson. The dependency locality theory: a distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, editors, *Image, Language, Brain.* MIT Press, 2000.

J. Godfrey, E. Holliman, and J. McDaniel. Switchboard: Telephone speech corpus for research and development. International conference on acoustics, speech and signal processing, 1992.

Kyle Gorman, Jonathan Howell, and Michael Wagner. Prosodylab-aligner: a tool for forced alignment of laboratory speech. The annual conference of the Canadian Acoustical Association, forthcoming.

John Hale. The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2): 101–123, March 2003.

John Hale. uncertainty about the rest of the sentence. *Cognitive Science*, 30:643–672, 2006.

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

Jason Kahn and Jennifer E. Arnold. Reduction comes from facilitation of levels of language production. Paper presented at the second conference on experimental and theoretical advances in prosody, September 2011.

T.Q. Lam and D.G. Watson. Repetition is easy: Why repeated referents have reduced prosodic prominence. *Memory and Cognition*, 38(3):1137–1146, 2010.

B. Lindblom. Explaining phonetic variation: a sketch of h and h theory. In William J. Hardcastle and Alain Marchal, editors, *Speech production and speech modeling*, pages 403–439. Kluwer Academic Publishers, The Netherlands, 1990.

Kimberly Jones Maner, Anne Smith, and Liane Grayson. Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of speech, language, and hearing research*, 43: 560–573, April 2000.

James L. McClelland and Jeffrey L. Elman. The trace model of speech perception. *Cognitive Psychology*, 18:1–86, 1986.

Douglas Roland, Frederic Dick, and Jeffrey L. Elman. Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57:348–379, 2007.

Elise C. Rosa, Kayla Finch, Molly Bergeson, and Jennifer E. Arnold. The effects of adressee attention on prosodic prominence. Paper presented at the second conference on experimental and theoretical advances in prosody, Montreal., September 2011.

Elisabeth Selkirk. On derived domains in sentence phonology. *Phonology Yearbook*, 3:371–405, 1986.

Elisabeth Selkirk. The prosodic structure of function words. In Suzanne Urbanczyk Jill Beckman, Laura Walsh Dickey, editor, *Papers in Optimality Theory*, pages 439–70. GLSA Publications, Amherst, MA, 1995.

H. Tily, S. Gahl, I. Arnon, N. Snider, A. Kothari, and J. Bresnan. Syntactic probabilities affect pronounciation variation in spontaneous speech. *Language and Cognition*, pages 147–165, 2009.

Alice Turk. Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis. *Laboratory Phonology*, 1(2):227–262, 2010.

Michael Wagner and Duane Watson. Experimental and theoretical advances in prosody: a review. *Language and Cognitive Processes*, 25:905–945, 2010.

Jill A. Warker. Investigating the retention and time-course of phonotactic constraint learning from production experience. Received from Joe in email 10 May '12, he got it from the author.

Tessa Warren and Edward Gibson. The influence of referential processing on sentence complexity. *Cognition*, 85(1):79–112, August 2002.

Duane Watson and Edward Gibson. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6):713–755, 2004.

Linda R. Wheeldon and Mark C. Smith. Phrase structure priming: A short-lived effect. *Language and Cognitive Processes*, 18(4):431–442, 2003.

Colin W. Wightman, Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91(3):1707–1717, March 1992.

A Experimental Items

Pretest

1.	**SE:**	Did John scam Melvin out of some money?
		It was Edward who (t) scammed Melvin out of some money.
	OE:	Did Melvin scam John out of some money?
		It was Edward who (t) Melvin scammed out of some money.
2.	**SE:**	Did Mary blame Bob in order to ruin her reputation?
		It was Alice who blamed Bob in order to ruin her reputation.
	OE:	Did Bob blame Mary in order to ruin her reputation?
		It was Alice who Bob blamed in order to ruin her reputation.
3.	**SE:**	Did Meg beam Mark up to the spaceship?
		It was Ellen who beamed Mark up to the spaceship.
	OE:	Did Mark beam Meg up to the spaceship?
		It was Ellen who Mark beamed up to the spaceship.
4.	**SE:**	Did Ellie lose Jim down at the lake?
		It was Kim who lost Jim down at the lake.
	OE:	Did Jim lose Ellie down at the lake?
		It was Kim who Jim lost down at the lake.

	Post-Training	
5.	**SE:**	Did Kit meet Joe for tea? It was Erik who met Joe for tea.
	OE:	Did Joe meet Kit for tea? It was Erik who Joe met for tea.
6.	**SE:**	Did Max set Beth up with Manny? It was Larry who set Beth up with Manny.
	OE:	Did Beth set Max up with Manny? It was Larry who Beth set up with Manny.
7.	**SE:**	Did Kim leave Jeff behind after the party? It was Brandon who left Jeff behind after the game.
	OE:	Did Jeff leave Kim behind after the party? It was Brandon who Jeff left behind after the game.
8.	**SE:**	Did Brandon let Liz win? It was Kevin who let Liz win.
	OE:	Did Liz let Brandon win? It was Kevin who Liz let win.

Training Items		
1.	**SE:**	Did Ann bang Ed on the head with a rolling pin? It was Sally who banged Ed on the head with a rolling pin.
	OE:	Did Ed bang Ann on the head with a rolling pin? It was Sally who Ed banged on the head with a rolling pin.
2.	**SE:**	Did Julia bind Samuel to a chair before making her escape? It was Abby who bound Samuel to a chair before making her escape.
	OE:	Did Samuel bind Julia to a chair before making his escape? It was Abby who Samuel bound to a chair before making his escape.
3.	**SE:**	Did Jake cram all the children into a phone booth? It was Alan who crammed all the children into a phone booth.
	OE:	Did the children cram Jake into the phone booth? It was Alan who the children crammed into a phone booth.
4.	**SE:**	Did Leonard cast Jacob to play Mephistopheles? It was Rodney who cast Jacob to play Mephistopheles.
	OE:	Did Jacob cast Leonard to play Mephistopheles? It was Rodney who Jacob cast to play Mephistopheles.
5.	**SE:**	Did Maria typecast Bert before even thinking? It was Susan who typecast Bert before even thinking.
	OE:	It was Susan who Bert typecast before even thinking. Did Bert typecast Maria before even thinking?
6.	**SE:**	Did Max send Sam back home? It was Rebecca who sent Sam back home.
	OE:	Did Sam send Max back home? It was Rebecca who Sam sent back home.
7.	**SE:**	Did Joe teach Bill to play? It was Mary who taught Bill to play.
	OE:	Did Bill teach Joe to play? It was Mary who Bill taught to play.
8.	**SE:**	Did Amy follow Julia home? It was Joanne who followed Julia home.
	OE:	Did Julia follow Amy home? It was Joanne who Julia followed home..

B Extra graphs

Figure 6: Durations by subjects. While all subjects participated in the pretest portion of the experiment, each subject participated in only one of the OE- or SE-training conditions. Subjects also differed in the number of disfluent object-clefts they produced. Subjects with a duration of zero for the object-cleft condition produced no fluent object clefts. The y-axis represents durations in milliseconds, and the number above each graph is that subject's average speaking rate.

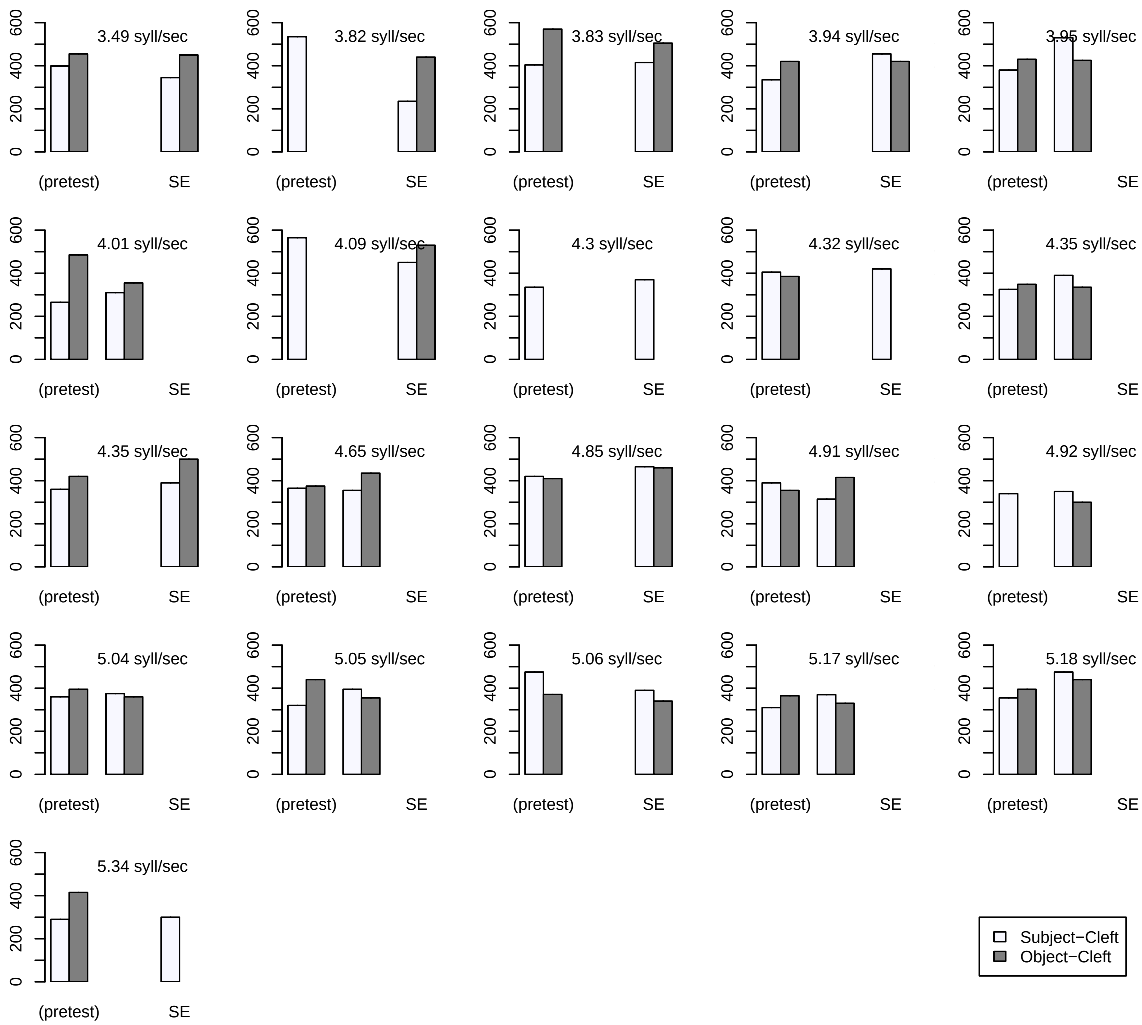


Figure 7: Durations by words.