Relative Scaling of Prosodic Boundaries Size in Sentence Production*

Yelena Fainleib

First Generals Paper, 2/19/2013

1. Introduction

This paper presents an experimental study, which explored the factors that influence the realization of prosodic boundaries in sentence production. Previous perception studies (Carlson, Clifton and Frazier 2001; Clifton, Carlson and Frazier 2002, 2006; Carlson, Frazier and Clifton 2009) have shown that the size of a prosodic boundary is processed relative to the size of earlier boundaries. In ambiguous sentences such as *Sally learned$_A$ that John telephoned$_B$ last night,* more interpretations of temporal adverbial modifying the matrix verb were perceived in cases when boundary A was smaller than boundary B, rather than A being bigger than or equal to B.

The present study used stimuli sentences intended to elicit production of two prosodic boundaries. They consisted of matrix clause, subordinate clause and temporal adverbial. The content of the adverbial indicated whether it modified the matrix or the subordinate verb. This manipulation made it possible to test whether the relative size of produced prosodic boundaries would be different in these two types of syntactic structures.

In order to make predictions about the manner in which the prosodic scaling of boundaries would be manifested, models of prosodic planning had to be considered. The expected results were predicted to be different depending on availability of large look-ahead in the planning of prosodic structure. If the speakers have access to the structure of the whole sentence, both boundaries A and B were predicted to be adjusted to reflect the type of

* I am deeply grateful to Lyn Frazier and Kristine Yu for their invaluable help and guidance throughout all the stages of this project, as also to Alice Harris for the comments on the later version of this paper, and to Michael Becker for the input in the earlier stages of this work and for the help with the experimental software. I also wish to thank John Kingston the insightful comments on data analysis, the members of the 2nd year seminar at UMass Spring 2012, and the audience of the UMass 2nd year mini-conference. Many thanks to the participants in this study and a special thank you to Bailey Hillst for all the help with finding the participants. All errors are mine alone.

attachment. In the case of the look ahead spanning only over the next available word or two, only the size of boundary B was predicted to change between conditions.

The results confirmed that the relation between the sizes of boundaries A and B was different in structures where the adverbial modified the higher verb and the structures where the lower verb was modified. The scaling was realized by adjusting the size of the later boundary, consistent with no large look-ahead model.

This paper is organized as follows: section 2 presents the relevant theoretical and experimental background to the current study; section 3 describes the experiment and its results and section 4 presents the results. Section 5 presents the summary of the results and their implications in the context of existing literature. Section 6 concludes this paper.

2. Background

This section begins by presenting the relevant background on prosodic structure in section 2.1, and on its matching with syntactic structure in section 2.2. Section 2.3 describes the studies that examined the scaling of prosodic boundaries in perception and production. Section 2.4 describes the present study.

2.1 Prosodic structure

Prosodic structure, like syntactic structure, is composed of discrete units which are arranged in a hierarchical relation with each other. It organizes phonetic and phonological material both above and below the level of the word. The lowest units below the word level are the morae, which are parsed into syllables. The syllables are parsed into feet; these are further grouped into prosodic words, which embody the lexical items.[1]. The number of constituents above the prosodic word level varies in different theories. For example, according to Selkirk (1978), prosodic words are grouped into a larger unit called phonological phrase; (one or

[1] Functional words, such as *to* are assumed to be appended to lexical items to which they relate, and parsed in the same prosodic word (eg Selkirk and Chen 1990, Selkirk 1996).

more) phonological phrases are parsed into intonational phrase, which are dominated by the largest constituent called utterance. Another theory by Beckman and Pierrehumbert (1986), assumes only two levels above prosodic word: the intermediate phrase (ip) which dominates the prosodic words, and the intonational phrase (IP), which consists of one or more intermediate phrases. The Beckman and Pierrehumbert model was adopted for this study. An example of a possible prosodic parsing of a phrase "visit Mississippi" is shown in Figure 1 below:

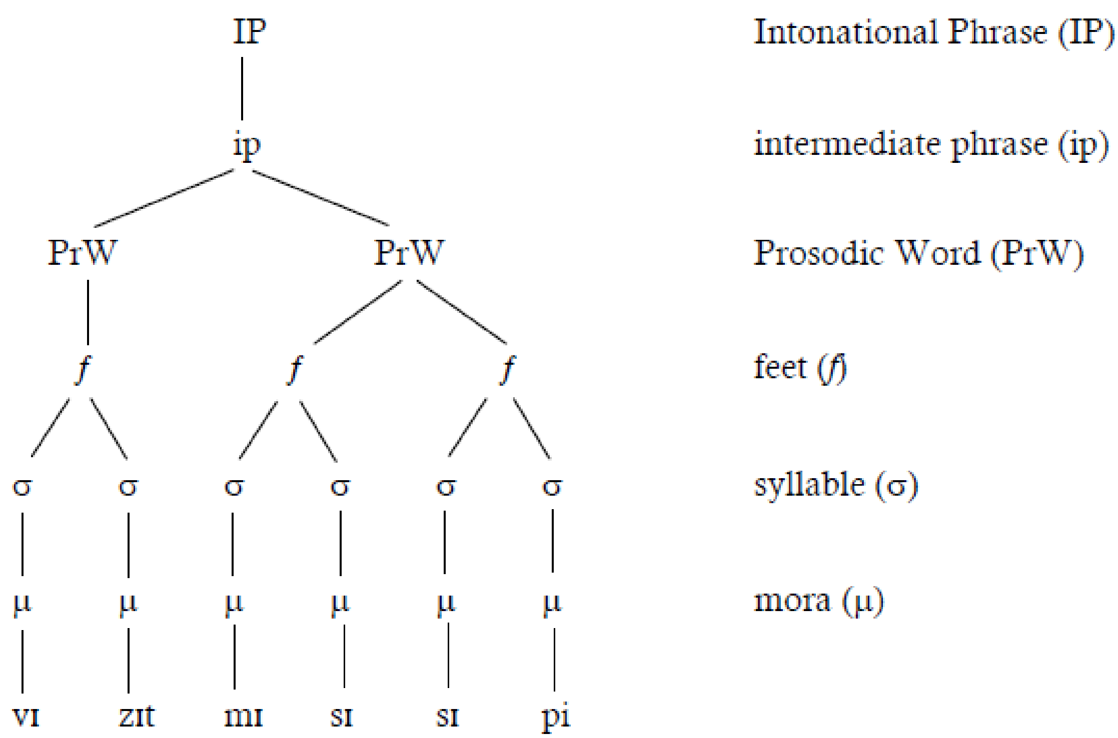

Figure 1: prosodic structure of the phrase "visit missisipi" according to model compatible with Beckman and Pierrehumbert's approach (adapted from Selkirk 2012).

The boundaries between prosodic constituents are marked by acoustic properties which include lengthening of the pre-boundary segments (Klatt 1975, Shattuck-Hufnagel and Turk 1996, Wightman, Shattuck-Hufnagel, Ostendorf and Price 1992) and restarting of the level of fundamental frequency (F0) (which declines over the course of the sentence production) at a higher level (dePijper and Sanderman 1991, Beckman and Pierrehumbert 1986). The magnitude of these acoustic properties tends to be bigger at the boundaries between two constituents which occupy a higher position in the prosodic hierarchy, than at the boundaries between two constituents which occupy a lower position. Specifically, segments preceding a boundary between two intonational phrases (IPs) were found to be longer compared to the segments preceding two intermediate phrases; these in turn were found to be longer than segments preceding the boundary between two prosodic words dominated by the same intermediate phrase (e.g., Byrd and Saltzman 1998, Byrd 2000, Edwards, Beckman and Fletcher 1991). With respect to the restarting of F0 level, dePijper and Sanderman (1991) have found that more significant interruptions in F0 were perceived as cues for boundaries between larger constituents in Dutch. Selkirk (2005) cites Truckenbrodt (2002) as observing that the "upward pitch reset found at the left end of Intonational Phrase goes higher than the pitch reset at the end of the Major Phrase (which they call Phonological Phrase)" in German and Ladd (1986, 1988) in English.

Based on these findings, the present study assumes that the magnitude of pre-boundary segment lengthening and of the F0 reset is correlated with the position of constituent in the prosodic hierarchy (as in Figure 1). This approach implies that the boundaries between two prosodic constituents which occupy a higher position are assumed to always be bigger than the boundaries between the hierarchically lower constituents.

2.2 Syntax-prosody mapping

There is much evidence that shows there is correspondence between syntactic and prosodic structures. Studies by Lehiste (1973), Nespor and Vogel (1986), Price et al (1991), Schafer at al (2000) have shown that the listeners are quite accurate in perceiving the intended meaning of a syntactically ambiguous utterance. An example of such sentence is *They rose early in May* (Price et al 1991), which has two possible meanings: (1) they rose early every day during May and (2) the onetime event of their rising occurred short after May has begun (as opposed to when May was close to its end).[2] In the first case, the adverb *early* modifies

[2] The exact contexts that were given to the speakers that produces the sentences in Price at al (1991) were as follows: (1) In spring there was always more work to do on the farm. May was the hardest month. *They rose early in May*. (2) Bears sleep all winter long, usually coming out of hibernation in late April, but thisy eart heyw erea little slow. *They rose early in May.*

the verb *rose*. In the second case *early* modifies the adverbial phrase *in May*. The participants in Price et al's study, who heard the sentences could identify the intended meaning in 95%. Acoustic and phonological analyses revealed a large prosodic boundary (ip or IP) after *early* in the first reading and before *early* in the second reading. This findings implies, that the speakers (who initially produced the sentence) chose to parse the sentence into two prosodic constituents, with the second one containing the final prepositional phrase *in May*, shown in (1) below.

(1) *Two possible prosodic parsings of the sentence "they rose early in May".*

a. $([\text{they}]_{NP}\ [\text{rose early}]_{VP})_{IP1}\ ([\text{in may}]_{PP})_{IP2}$

b. $([\text{they}]_{NP}\ [\text{rose}]_{VP})_{IP1}\ ([\text{early in may}]_{PP})_{IP2}$

As the example shows, in the first reading, *early* is not part of the prepositional phrase , but of the preceding VP, the right edge of which corresponds with the right edge of IP1. This parsing results in a prosodic break between *early* and *May.* In the second reading, *early* is part of the adverbial phrase, the left edge of which corresponds with the left edge of IP2 resulting in the break between *rose* and *early*.

The above example demonstrates the **alignment** between the syntactic and the prosodic constituents. There are several proposals on the specific rules on how this alignment is manifested, such as Selkirk (1986), Trockenbrodt (1999), with the latest development being the Match theory (eg Selkirk 2009). This framework proposes the following rules:

(2) *Match theory of correspondence between syntactic and prosodic constituents:*

a. <u>Match Clause</u>: A clause in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ι, in phonological representation.

b. <u>Match Phrase</u>: A phrase in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it φ, in phonological representation.

c. <u>Match Word</u>: A word in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ω, in phonological representation.

Selkirk (2011) states that Clauses which are the complements of the illocutionary head $Force^0$ correspond more often to intonational phrases (eg Dehé 2009 for English). However, there are also cases in which clauses which are the complements of the functional head C also corresponding to intonational phrases (eg. Kisseberth 1994 for Xitsonga, Truckenbrodt 1999 for German).[3] Moving on to the Match Phrase rule, a Phrase can roughly be defined as the maximal XP projection (Jackendoff 1977). This rule predicts therefore, that a simple SVO sentence which has the syntactic structure [NP] [V NP] could be parsed as $(NP)_{ip}(V (NP)_{ip})_{ip}$.

These matching rules are violable (as in the Optimality Theory framework, Prince and Smolensky 1993, McCarthy and Prince 1993), and can be outranked by other rules, such as these imposing phonological wellformedness. The result might be a deviation from the 1:1 mapping between syntactic and prosodic constituents. For example, if the Match (ip, XP) rule is outranked by a rule that requires that an ip would contain maximum two prosodic words (Selkirk 2000), it will create a mismatch between the syntactic and the prosodic structures, such as in the following example from Japanese (cf. Selkirk and Tateishi 1988, Kubozono 1993; Shinya, Selkirk and Kawahara 2004). In this example, a four word noun phrase with a left branching recursive syntactic structure is being parsed into two prosodic constituents (φ):

(3) *Matching rule being outranked by a phonological well formedness rule*

$[[[\text{N-no N-no}]_{NP}\ \text{N-no}]_{NP}\ \text{N-ga}]_{NP} \rightarrow {}_{\varphi}((\text{N-no N-no})_{\varphi 1}\ (\text{N-no N-ga})_{\varphi 2})_{\varphi}$

The prosodic constituent φ_2 does not have a syntactic match. This case is analyzed as the rule which matches between NPs and φs is overridden by a rule which requires that the constituent φ would contain maximum two prosodic words. Prosodic parsing that would be faithful to syntactic structure $(((\text{N-no N-no})_{\varphi 1}\ \text{N-no})_{\varphi 2}\ \text{N-ga})_{\varphi 3}$ would violate this requirement for prosodic phrases φ2 and φ3.

[3] Selkirk 2009 proposed two types of Match clause, for illocutionary clauses and for standard clauses, with the former dominating the latter.

Another factor that can influence the syntax-phonology matching is the length of the constituents, measured in prosodic units, such as the number of syllables. Jun (2003) reports that, in Korean, a syntactic phrase is parsed into one prosodic constituent (called Accentual Phrase) if it is five or less syllables long; otherwise it is often parsed into two Accentual Phrases. D'Imperio et al (2005) have found that in Standard European Portuguese the length of the subject in a (SVO) sequence determined whether it was parsed in the same prosodic constituent as the verb and the object (if 5 and less syllables long), or in a separate one. Watson and Gibson (2004a) show, that an algorithm that predicts the occurrence of intonational breaks, and takes into account the length recently parsed and the length of incoming constituents outperforms those based only on syntactic and semantic factors (such as (Cooper and Paccia-Cooper, 1980; Gee and Grosjean, 1983; Ferreira 1988).

Speech rate also influences the syntax-prosody mapping: prosodic phrases tend to be larger when the rate of speech is fast, than when it is slow. The example from Italian brought in (4) below demonstrates that by showing the domain of spirantization in Italian: inside but not across a phonological phrase (PhP) (Nespor 1993). The examples below show, that when the speech rate is faster, some of the IP boundaries are erased, and therefore, spirantization k → h (underlined in the below example) can apply across wider domain:

(4) *Example from Italian of prosodic boundaries being influenced by speech rate*

a. [[La[h]amomilla]$_{PhP}$]$_{IP}$ [[[k]ome è noto]$_{PhP}$]$_{IP}$ [[k]alma] $_{PhP\ PhP}$ [anche i bambini] $_{PhP}$]$_{IP}$ [[k]on problemi] $_{PhP\ PhP}$ [diinsonnia] $_{PhP}$]$_{IP}$

'Chamomile as known calms even children with problems of insomnia ' hamomile, as it is known, calms even children having insomnia problems'

b. [[La[h]amomilla] $_{PhP}$]$_{IP}$ [[[k]ome è noto] $_{PhP}$]$_{IP}$ [[k]alma] $_{PhP\ PhP}$[anche i bambini] $_{PhP\ PhP}$ [[h]on roblemi] $_{PhP\ PhP}$ [dinsonnia] $_{PhP}$]$_{IP}$

To summarize this section: there is ample evidence supporting the correspondence between syntactic and prosodic constituents. This matching, however, is not isomorphic, and is dependent on additional factors, such as prosodic wellformedness, length and/or rate of speech. Therefore, many of prosodic breaks which are realized, are not obligatory, but are dependent the speaker's spellout decisions.

2.3 Comprehension and production of prosodic boundaries

Comprehension studies have revealed another factor which can influence the size (the position of the constituent in the prosodic hierarchy in section 2.1) of the optionally realized prosodic boundaries. Specifically, it can be influenced by the size of other prosodic boundaries (at and above the prosodic word level) which were realized in the sentence. Carlson, Clifton and Frazier 2001, Clifton, Carlson and Frazier 2002, Carlson, Clifton and Frazier 2006 and Carlson, Frazier and Clifton 2009 used ambiguous stimuli sentences such as in (5) below, (already documented in the introduction) in which the size of prosodic boundaries at locations marked by A and B was varied:

(5) Sarah learned$_A$ that Bill telephoned$_B$ last night.

The temporal adverbial *last night* can modify either the main verb thus producing the meaning that it was last night that Sarah learned that Bill had telephoned (some other time), or it can modify the subordinate verb which would imply that Bill telephoned last night (and Sarah learned that next morning for example). The first syntactic structure henceforth is referred to as high attachment, and the second one as low attachment.

It was discovered that when boundary B was bigger than boundary A (A = ip, B = IP or A = 0[4], B = ip), significantly more high attachment readings were facilitated than when boundary B was smaller or same size as boundary A. The authors concluded that boundary at B can be interpreted only in the context of the preceding boundary. The approach called the "Rational Speaker Hypothesis" states that the listeners are sensitive to the global prosodic context of the sentence, since they assume that the speakers consistently make their prosodic choices based on the relevant syntactic structure and intended meaning of the utterance that they will produce. The Rational Speaker Hypothesis predicts that the speakers cannot intend a low attachment structure (temporal adverbial modifying the subordinate

[4] Boundary between two prosodic words, using notation of Snedeker and Casserly (2010)

verb) and produce a large prosodic break at B for no reason. The Rational Speaker Hypothesis was tested not only in perception of sentences like (5), but also for other types of structures such as conjunction (*old men and women with large houses*), possessive phrases (*Johnny and Sharon's in-laws*), ambiguous relative clause attachments (*the daughter of the colonel who was on the balcony*) and was consistently supported.

The question whether in production prosodic boundaries are scaled with respect to each other was addressed in just one study by Snedeker and Casserly (2010). They have re-examined the data from an earlier study by Snedeker and Trueswell (2003), in which the speakers had to give instructions to the listeners using a toy set. The instructions were a finite set of ambiguous sentences, such as *Tap*$_A$ *the frog*$_B$ *with the flower* and the specific meaning of the utterance (and thus the attachment of the PP *with the* flower) was manipulated within the speakers. The authors examined whether in their realizations of high versus low attachment structures, the speakers scaled both boundaries A and B, or just boundary B, which, according to Price et al (1991) and should be IP in the case of high attachment and ip in the low attachment case (as opposed to a possible (0, ip) but not (IP, IP) case in the global scaling option). Only moderate support for the Rational Speaker Hypothesis was found: the data indicated that the speakers seemed to modify only boundary B to realize specific attachment type, and not to scale both boundaries. However, the authors themselves admit that the number of tokens in each cell was not balanced, and sometimes was quite small (less than 20 productions), therefore, making it impossible to draw solid conclusions about the nature of prosodic boundaries scaling.

Going back to comprehension studies, another factor which was found to be significant in the perception of prosodic boundaries is the size of the post-boundary constituent (defined in number of syllables). In an experiment of Carlson, Clifton and Frazier (2006), there were sentences such as in (5), in which the temporal adverbial was long (averaged 2.25 vs average 11.5 syllables):

(6) Sarah learned$_A$ that Bill telephoned$_B$ last night after the general meeting.

The experiment revealed no difference in high attachment interpretations for sentences in which there was an IP boundary before the temporal adverbial (at B) and the sentences in which the adverbial was parsed in the same IP as the rest of the sentence. The authors concluded that when the temporal adverb was long, the participants perceived the size of the boundary as being indicative of the long constituent that is following it, and not of the type of the syntactic juncture. These findings are in concord with these of Watson and Gibson (2004b), who argue that speakers are more likely to place intonational phrase boundaries before long constituents than before short ones, since they need time to plan the next portion of the utterance (Right Hand Side constituent size) . Their conclusions correspond to findings from Ferreira (1991) that found that the initiation times to the production of the sentence positively correlated with the complexity of the sentence's subject, as also with similar results from Wheeldon and Lahiri (1997).

2.4 Overview of the current study

As it is possible to see from the background presented so far, there was only one study that has attempted to check the factors that influence the principles of realizations of prosodic boundaries during production. However, its stimuli were not specifically designed for that purpose. The rest of the mentioned studies, while showing important mechanisms in the perception of sentence prosody, can at most imply what the speakers could be doing in production, since all the stimuli were produced with a pre-defined prosodic structure.

Therefore, this study was designed to access which factors influence the production of prosodic boundaries. Specifically, if speakers choose to realize more than one boundary during the production of a sentence, will the size of a certain boundary be influenced by the size of other boundaries in the sentence?

In order to examine the scaling of the boundaries, the study used sentences which included a main clause, a subordinate clause and a temporal adverbial. The temporal adverbial could modify either the main verb, thus resulting in a high attachment structure, or the subordinate verb, thus resulting in a low attachment structure. The sections of the sentences before the temporal adverb were relatively long in order facilitate a production of prosodic boundaries after the main verb and the subordinate verb. For the sake of convenience, the first boundary was referred to as boundary A, and the second one as boundary B. Due to the difference in perception of the attachment site between the short and the long temporal adverbials in Carlson, Clifton and Frazier (2006), as also due to the influence of constituents length on the prosodic parsing of the utterance the temporal adverbials in the present study had to be manipulated for their number of syllables. In this manner, two additional categorical conditions were established: a short adverbial condition and a long adverbial condition (see section 3.1 for detailed description of stimuli sentences).

Our predictions are as follows: the evidence for scaling of prosodic boundaries at A and B should be most evident in the condition where the adverbial is short, since the size of boundary B would not be confounded by the possible preference to signal a large constituent (Clifton, Carlson and Frazier 2006, Warren and Gibson 2004b). The exact manner in which the scaling would be realized is dependent on whether the speakers have a large look ahead while planning the production of their utterances. The general assumption is that speakers do not plan the pronunciation details (such as the specific phonetic realization of segments) for the whole utterance, but proceed in chunks, with the later portion of the sentence planned in parallel with the utterance of an earlier chunk (eg Ferreira 1993, Keating and Shattuck – Hufnagel 2002, Meyer et al 2007). There is, however, variation among different the models in their assumption concerning how much control the speaker has over this incremental production.

A model by Keating and Shattuck-Hufnagel (2002) suggests that prior to encoding the complete phonetic specification of phonological output, a rough prosodic skeleton of the entire sentence is built based on the syntactic structure, thus ensuring a large look ahead and the possible informing of the nature of the later constituents during the encoding of the earlier ones. An opposite approach is held in Levelt (1989), which assumes the look ahead of maximum one or two words. According to this theory, the prosodic skeleton is built incrementally, by querying the available syntactic constituents (which constitute the base for the prosodic constituents) one by one. The type of prosodic boundaries that is realized is dependent on the nature of the currently processed word. Under this assumption, the influence of a later prosodic boundary on an earlier one would be impossible.

In sum, the nature of prosodic boundary scaling in this study would have to depend on the existence of a large look-ahead. If there is such, the sizes of both boundaries can be scaled reflecting the high attachment or the low attachment structures, in other words reflecting that the size of boundary A can be adjusted according to the size of the later boundary B. We named this option the "Large look-ahead scaling hypothesis". If there is no large look-ahead, then the only option to scale the boundaries would be to adjust the size of boundary B according to the already established size of boundary A. We named this option "No look-ahead scaling hypothesis".

The other factor that could influence the size of boundary B was predicted to be the length of the temporal adverbial (short of long), which will be mostly evident in the low attachment structures. According to Warren and Gibson (2004b), the speakers were expected to produce a larger boundary before a long constituent than before a short one. Based on Carlson, Clifton and Frazier (2006), it could be expected that in high attachment structures, the participants might choose to realize the same size boundary both when the adverbial constituent is short and also when it is long: only in the first case, the size of the boundary might indicate the syntactic structure, while in the second case it can indicate a length of incoming constituent.

The boundary sizes were parameterized using acoustic durations of pre-boundary segments and the difference in the level of F0 between the pre-boundary and the post-boundary constituents. As already noted in section 2.1, they are both predicted to be bigger as the boundary increases thus providing a good estimate of the relative size of prosodic boundaries.

3. Materials and Methods

This section describes the production experiment, beginning with the description of experimental stimuli in section 3.1, continuing to participants in 3.2, and procedures in 3.3, and the data analyses in section 3.4.

3.1 Stimuli

Each experimental item consisted of a minimal quadruplet of sentences, which were identical up until boundary B. The temporal adverbial part provided the means for creating the high attachment/low attachment conditions and for short adverbial/long adverbial conditions. The distinction between the high attachment and the low attachment structures was achieved by matching the semantic content of the temporal adverbial with either the higher or the lower verb. The difference between the short adverbials and the long adverbials were expressed in their number of syllables: the short adverbials averaged 4.5 syllables and the long ones averaged 12.5 syllables. An example of a stimulus item is given in Table 1 below; for the full list of items please see Appendix A.

	High attachment structure	**Low attachment structure**
Short adverbial	Sarah will definitely try to evaluate$_A$ some data she obtained from her experiment$_B$ **<u>later this week</u>.**	Sarah will definitely try to evaluate$_A$ some data she obtained from her experiment$_B$ **<u>earlier this week</u>.**
Long adverbial	Sarah will definitely try to evaluate$_A$ some data she obtained from her experiment$_B$ **<u>later this week after the start of spring semester</u>**.	Sarah will definitely try to evaluate$_A$ some data she obtained from her experiment$_B$ **<u>earlier this week at the start of spring semester</u>**.

Table 1: an example of an experimental item. The temporal adverbial portion different in each condition is bolded and underlined. The high attachment condition is created by the matching the tense of adverbial with the meaning of the higher verb, and being incompatible with the meaning of the lower verb. In the low attachment condition the situation is reversed.

The sentences within the quadruplet were made with the intention of their being maximally similar to each other so that any possible difference in their production could be attributed only to the differences between the high attachment and the low attachment structures, and to the length of the temporal adverbial. It was impossible to maintain a fixed number of syllables in the adverbial for each short and for each long condition sentence due to the restrictions on the possible content of the adverbial, the preceding sentence content, and in order to avoid the option of making the temporal adverbial attachments identical across all the stimuli items.

Words flanking the potential boundary sites were controlled for their phonological and phonetic properties, to establish as uniform as possible environment for the phonetic measurements. Pre-boundaries words were controlled for prosodic structure; they were 3 or 4 syllables long and had antepenultimate stress[5]. The purpose of this manipulation was to increase the length of prosodic constituents before boundary A, thus facilitating the production of a boundary where intended by the experiment design. Post boundaries words were controlled for their initial segments: they never began with a stop, in order to not mix the pre-release silence with the potential prosodic break pause. Additionally, all sentences were produced in two different manners of speech: casual and emphatic. This manipulation was included in order to check for consistency in production of boundaries size and locations in the sentence, and to maximize the chances of eliciting pronunciations with boundaries at both relevant locations. The two speech manners were elicited by asking the participants to speak in either: (1) their regular manner as they do in their everyday lives in non-official situations to elicit a casual speech style (2) in a more deliberate and exaggerated manner to elicit an emphatic speech style.

In total, there were 8 stimuli items. The stimuli variants from each item were distributed between 4 lists using the Latin Square design. They were intermixed with 8 fillers, 4 early/late closure and 4 object/clause sentences taken from Anderson and Carlson (2010). These sentences were shorter than the experimental stimuli, and also involved potential gardenpath condition (eg. *As Louisa undressed the baby started to cry*), thus still forcing the participants to attend to the prosodic planning of the utterance pronunciation. For the full list of stimuli please see Appendix A.

[5] Except for one word before boundary B which was overlooked as having penultimate stress.

3.2 Participants

The participants were 16 undergraduate students from UMass Amherst, 12 females and 4 males, recruited from the departments of Linguistics and Psychology. All participants identified themselves as English monolinguals (grew up in a monolingual environment, spoke English with friends and have not lived abroad for more than a month) were naïve to the purposes of the experiment. They were compensated with $10 for their participation.

3.3 Procedures

The experiment was executed using the free online experimental software Experigen (Becker and Levine 2010, https://github.com/tlozoot/experigen), with all the instructions appearing on the screen as the experiment progressed. The experiment was self-paced, each new screen appearing after the participant hit a key. One sentence at a time was presented to the participants.

At the beginning the participants were presented with three sentences which were not included in the stimuli or the fillers, and were asked to read each of them out loud in order to practice their reading before having to read the actual stimuli. After completing this section, they were presented with a demo procedure for reading a sentence, which also included audio samples for the casual and the emphatic speech manners, in order to give an example of each. Since the emphatic speech manner was defined as a sort of exaggeration of the speaker's normal speech, the casual speech manner was elicited first, to provide a normal baseline for the speakers themselves, from which they can then deviate.

The utterance given in the demo procedure (and also recorded for the audio example) consisted of a main clause and a subordinate clause, but without the temporal adverbial part. Another purpose of the audio samples was to try and elicit a prosodic break at location A, which was present in both the casual and the emphatic speech audio examples. The samples were recorded by a female native English speaker, a graduate student in the Linguistics Department. For the sentence used in the audio example, as also sentences used in the training session please also see Appendix A.

For each stimulus sentence which the participants saw, the procedure was as follows: first they were asked to read the sentence silently to themselves. Next they were presented with three obligatory forced choice questions about the content of the sentence to force them to actually read the sentences and parse their underlying syntactic structure. Then they were asked to produce it in their casual way, and afterwards to produce it again with more expression, as if they were giving a speech. For the screenshots of the experimental procedure, including the instructions, please see Appendix B.

The experiment was conducted in a sound-attenuated booth in the phonetic laboratory at the Department of Linguistics at UMass Amherst. The participants' productions were recorded using a AKG C420 head mounted microphone, which was connected to a Macintosh desktop computer through a preamplifier. The sound file for each participant was recorded using Audacity (http://audacity.sourceforge.net/), at a mono channel, sample rate 44100 Hz, 32 bit float.

3.4 Data Analysis

Acoustic measurements were pre-boundary segments durations and the difference in post- and pre- boundary F0, using the free software PRAAT (Boersma and Weenik 2013). Pre-boundary segments duration – henceforth referred to as word durations – were collected by measuring the durations of three syllables immediately before boundaries A and B – starting from the antepenultimate stressed syllable (section 3.1). This procedure accounted for the possible segment lengthening not only of the very last pre- boundary syllable, but also the possible lengthening starting from the stressed syllable and until the end of the word (Turk and Shattuck-Hufnagel 2007).

The F0 difference – henceforth referred to as the pitch reset – was calculated from the average F0 on the sonorant segments of the last syllable before boundaries A and B (the nucleus and the coda), and the average F0 on the sonorant segments of the first syllable after the respective boundaries (the onset and the nucleus). The measurements were log transformed in order to account for the difference in pitch ranges between the male and the female participants. Then, the pitch reset was calculated as log(post- boundary pitch) – log(pre- boundary pitch).

To test the scaling of prosodic boundaries assuming the Large look-ahead scaling hypothesis, and also the influence of temporal adverbial length, a series of linear regression models were constructed. The dependent variables were the word durations and the pitch reset. Separate models were constructed for the following subsets of data for **each** of the boundaries A and B (shown in Table 2 below): short adverbial condition, long adverbial condition, low attachment condition, high attachment condition. This manner of examination was necessary, since according to the predictions laid out in section 2.4, the effects high/low attachment and of adverbial length would be easier to find in different subsets of data. Specifically, the effect of attachment should produce a significant difference in the short adverbial condition at the data from both boundaries, while the length of constituent effect should produce significant difference in the low attachment condition only at the subset of data from boundary B. The remaining subsets of data - high attachment condition and long adverbial condition - were not predicted to be influenced by the experiment design, but they were nonetheless examined for potential main effects.

	Data subset	Linear Model Factors
Boundary A	Short adverbial condition	Attachment height + speech manner
	Long adverbial condition	Attachment height + speech manner
	High attachment condition	Adverbial length + speech manner
	Low attachment condition	Adverbial length * speech manner
Boundary B	Short adverbial condition	Attachment height + speech manner
	Long adverbial condition	Attachment height + speech manner
	High attachment condition	Adverbial length + speech manner
	Low attachment condition	Adverbial length + speech manner

Table 2: factors which were examined for each of the data subsets at boundary A and at boundary B. The plus sign indicates that the final model did not include interaction between the factors, while the asterisk indicated that the final model included interaction. For coefficients of the factors in each model, please see Appendix F.

Since each linear model was specific only for its subset of data, the specific factors which were entered into each model varied, except for the speech mode, which was examined in all the models, and was tested for its interaction with word durations and pitch reset. The factors that were entered into the model of each data subset were identical both when the dependent variable was the word durations and when it was the pitch reset. The random effects were the item and the participant for all the models. [6]

The execution of the analysis and the model fitting was executed in R software (www.cran-project.org), using the lmer command from lme4 package (Bates and Maechter 2012). Initial models included the interactions of all the possible factors for a specific subset of data, and backward comparison was performed to find the best fitting model. The model comparison was performed in a likelihood ratio test, using the anova command. The

[6] Additional scaling of the word durations was performed later to account for the possible influence of speech rate on the measurements. No differences in their analyses from the raw word durations was found. See the end of Appendix E for details.

reported approximations to the p values of the factors in each model were made with the Monte Carlo (MCMC) distribution method (Robert and Cassella 2004), using the pvals.fnc() command from the package languageR.

In the course of examination, it was discovered that the results from the pitch reset produced different results from the word durations, and were often unpredictable. Therefore, only word durations were considered as evidence and as a measurement for the evidence for the influence of scaling.[7] For results from the pitch reset and a short discussion of their implications please see appendix C. Table 2 below presents the specific factors which were entered into the model in each subset of data.

To test the scaling of prosodic boundaries assuming No look ahead hypothesis, Pearson's R correlations were performed between the word durations at boundary A and at boundary B. The scaling would manifest itself in a strong positive correlation in the high attachment short adverbial condition (since the size of boundary B would have to be adjusted according to the size of boundary A), while in the case of the low attachment short adverbial condition, the correlation would be significantly smaller. Pearson's R correlation analysis was executed in R, using the corr() command from the package stats (R Development Core Team 2011) The significance of the difference between the correlations was determined using Fisher's r to z transformation (from http://www.vassarstats.net/rdiff.html). The correlations were computed for each of the conditions: high attachment short adverbial condition, low attachment short adverbial condition, high attachment long adverbial condition and low attachment long adverbial condition. [8]

4. Results

This section describes the results that were obtained from the data analysis described in section 3.4. Section 4.1 presents the results of the examination of data subsets from

[7] Lehiste (1973) also used pre-boundary words as a main evidence for prosodic disambiguation.

[8] Another possible evidence for prosodic boundaries scaling without the lookahead could look as follows: WORD DURATIONS A/WORD DURATIONS B < 1 for high short attachment condition; WORD DURATIONS A/WORD DURATIONS B > 1 for the low attachment conditions. See appendix D for that.

boundary B, to test for the effect of constituent length. Section 4.2 presents the data subsets from boundary A, which should provide the main evidence for the Large look-ahead scaling hypothesis. Finally, section 4.3 presents results from the correlations of word durations between the two boundaries, which should constitute evidence for No look-ahead scaling hypothesis. The main effects of the speech manner are also reported (where appropriate), however, they are not discussed in length. In all the cases, word durations in emphatic speech manner were longer than in the casual speech manner.

4.1 Examination of boundary B for influence of adverbial length

Figure 2 below shows that word durations were longer in the long adverbial condition (mean = 539.37, standard deviation = 119.9), than in the short adverbial condition (mean = 519.67, standard deviation = 131.1). This difference, although not very big, proved to be significant, ($t = 2.125$, $p = 0.03$), implying that the speakers chose to realize a bigger prosodic boundary before longer constituent. The high attachment condition did not reveal any influence of adverbial length. The effect of speech mode was found to be significant in both conditions ($t = 2.249$, $p = 0.02$) for low attachment condition and ($t = 2.504$, $p = 0.02$) for high attachment condition.

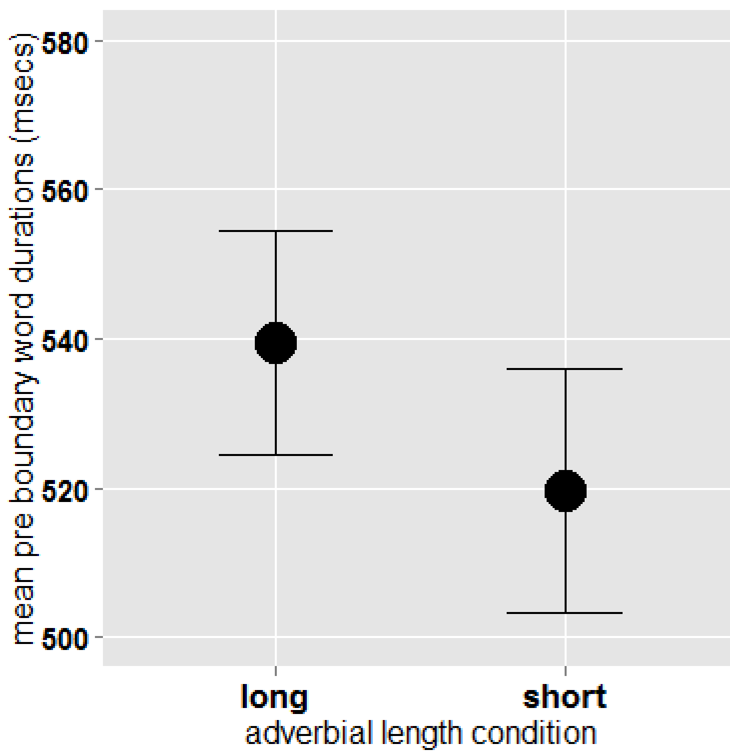

Figure 2: mean pre- boundary B word durations, for sentences with low attachment structure, showing the effect of adverbial length on the estimated size of boundary B. The dots indicate the mean word durations for the long adverbial condition (mean = 539.37, standard deviation = 119.9) and the short adverbial condition (mean = 519.67, standard deviation = 131.1); the error bars indicate +/- 1 standard error. The figure shows that word durations are longer when the adverbial was long, a difference which was found significant.

The examination of the short adverbial condition data subset, presented in Figure 3 below, revealed that word durations were longer in the high attachment condition (mean = 554.41, standard deviation = 158.53) than in the low attachment condition (mean = 519.67, standard deviation = 131.1). This difference was highly significant ($t = 3.592$, $p < 0.001$), There was an additional main effect of speech manner ($t = 2.234$, $p = 0.02$). The examination of the long attachment condition, not surprisingly, did not reveal any effects of height and the only main effect in this condition was of speech manner ($p < 0.01$).

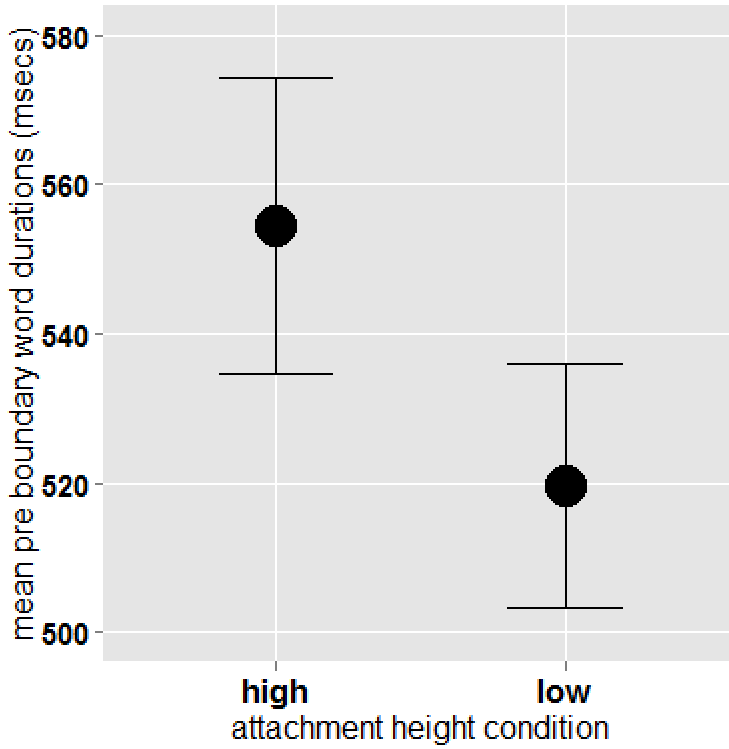

Figure 3: mean pre- boundary B word durations, for sentences with short adverbial, showing the effect of attachment on the estimated size of boundary B. The dots indicate the mean word durations for the high attachment condition (mean = 554.41, standard deviation = 158.53) and the low attachment condition (mean = 519.67, standard deviation = 131.1); the error bars indicate +/- 1 standard error. The figure shows that word durations were significantly longer when in the sentences with high attachment structure than in the sentences with low attachment structure.

4.2 Examination of boundary A for evidence of scaling as a function of syntactic structure type

As Figure 4 below shows, the short adverbial condition did not show a significant difference between the high attachment (mean = 470.57, standard deviation = 107.81) and the low attachment conditions (mean = 466.35, standard deviation = 91.56) at boundary A. No main effect of height was found ($t = 0.401$, $p = 0.63$).

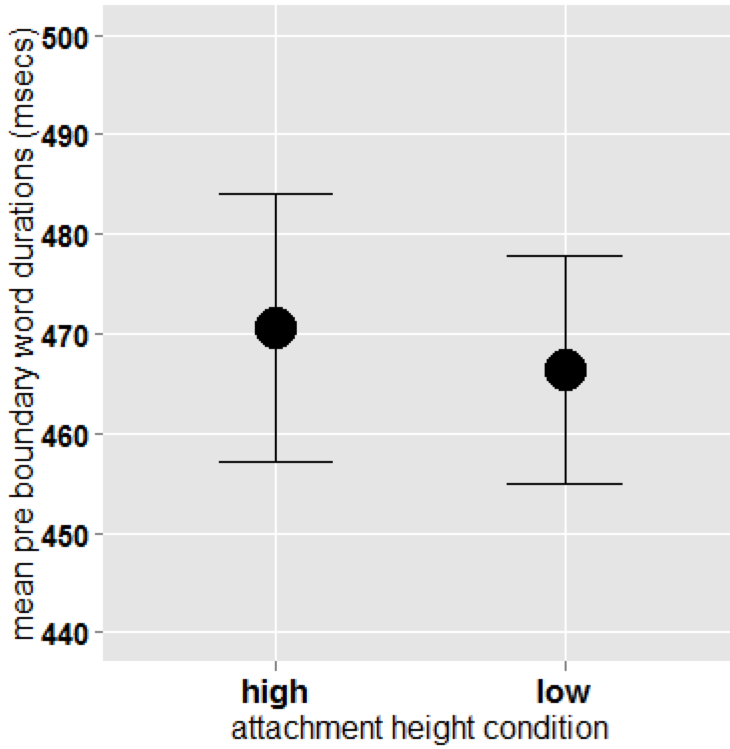

Figure 4: mean pre- boundary A word durations, for sentences with short adverbial, showing the effect of attachment on the estimated size of boundary A. The dots indicate the mean word durations for the high attachment condition (mean = 470.57, standard deviation = 107.81) and the low attachment condition (mean = 466.35, standard deviation = 91.56); the error bars indicate +/- 1 standard error. The figure shows that word durations are virtually identical in sentences with high attachment structures and in sentences with low attachment structures.

The examination of the long adverbial condition revealed that the pre boundary A word durations were longer in the high attachment condition (mean = 470.57, standard deviation = 107.81) than in the low attachment condition (mean = 466.35, standard deviation = 91.56), as shown in Figure 5 below. Surprisingly a near main effect of height ($t = 1.816$, $p = 0.07$), was found. There was also a main effect of speech manner ($t = 3.299$, $p = 0.001$).

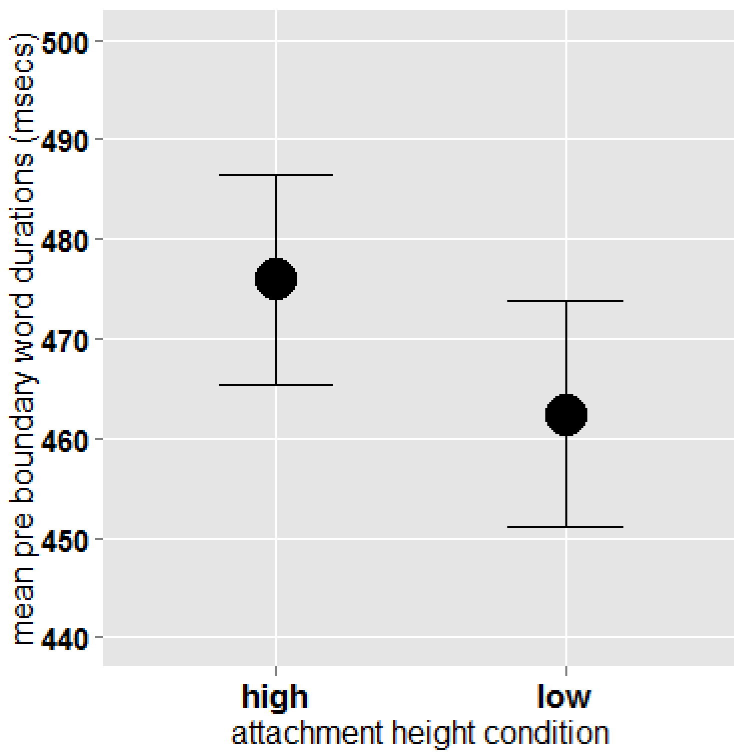

Figure 5: mean pre- boundary A word durations, for sentences with long adverbial, showing the effect of attachment on the estimated size of boundary A. The dots indicate the mean word durations for the high attachment condition (mean = 470.57, standard deviation = 107.81) and the low attachment condition (mean = 466.35, standard deviation = 91.56); the error bars indicate +/- 1 standard error. The figure shows that word durations are longer when the adverbial was long, a difference which was found to be approaching significance.

No predictions were made whether the influence of the length of adverbial would influence the size of boundary A; however both high and low attachment conditions were examined for main effects. The examination of low attachment condition (which was supposed to be indicative of the influence of adverbial length for boundary B) did not reveal any effect of either height of attachment or speech manner.[9] The examination of the high attachment condition also revealed only near main effect of speech style ($t = 1.76$, $p = 0.08$).

[9] There was a near-significant interaction between the two factors; however, its effects do not provide any insights into the nature of the boundaries production.

To summarize the results: the examination of the main effects at boundaries B and A show an influence of constituent length on the local realization of prosodic boundaries. The data does not show that speakers have insight into the size of a later prosodic boundary: the data from boundary A, which was supposed to bear the main evidence for prosodic scaling, did not reveal any influence of the height of attachment on its size (the duration of pre-boundary words).

4.3 Examination of boundaries A and B for change of correlation values as a function of syntactic structure type

The correlations were r = 0.78 for the high attachment short adverbial condition; r = 0.49 for the low attachment short adverbial condition; r = 0.45 for the high attachment long adverbial condition and r = 0.57 for the low attachment long adverbial condition.[10] The difference between the high attachment short adverbial condition and the low attachment short adverbial condition, which were predicted to be the main indicator for the scaling between the two boundaries proved to be significant, p = 0.005 (two tailed). The statistical difference between the high attachment long adverbial and the low attachment long adverbial conditions did not reach significance.

Two charts below in Figure 6 below present the scatter plots and the regression lines for the correlations for the conditions for which the correlations were calculated.

[10] For the rest of correlations that were performed to verify the validity of the ones presented in this section, see appendix E.

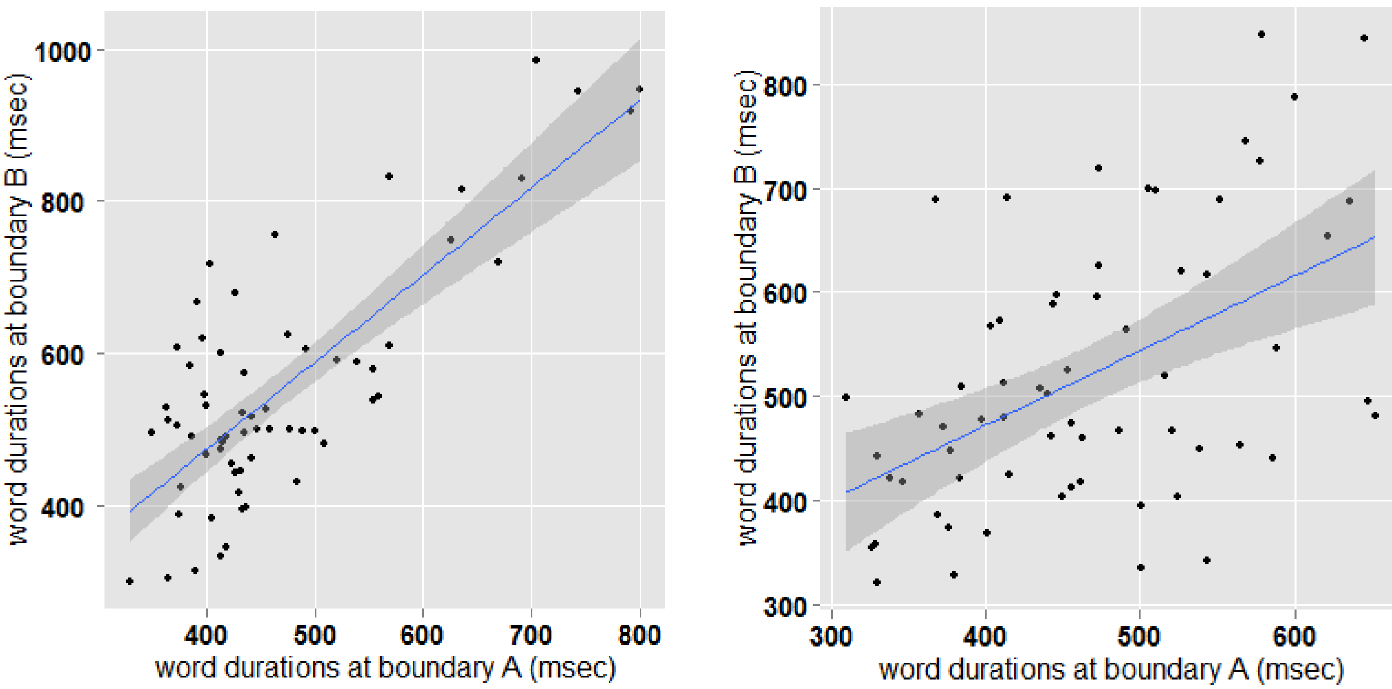

Figure 6: scatter plots and regression lines for the Pearson's R correlations. Figure 6a on the left presents the data for the high attachment short adverbial condition[11], while Figure 6b on the right presents the data for the low attachment short adverbial condition. It is possible to see that the scattering of the data for the correlation between the word durations in the high attachment short adverbial condition is narrower and has a higher slope than for the low attachment short adverbial condition. The grey area represents smoother using lm function.

The results presented in this section show evidence for the scaling of two prosodic boundaries; word durations at boundary B were adjusted in the high attachment long adverbial condition to signal larger boundary, than the earlier produced A.

[11] Figure 6a contains many outlier points. A more thorough examination of their influence of the result of correlation (which is currently beyond the scope of this paper) is required here.

4. Discussion

In this section we first present the summary of the results in 4.1, and then continue to their examination in the context of previously established data in 4.2. Section 4.3 presents possible directions for future research.

4.1 Results summary

This study has explored the scaling of prosodic boundaries in sentence production. It examined the two possible ways the speakers could scale the sizes of prosodic boundaries in production of sentences, should they choose to realize two boundaries, outlined in section 2.4. The Large look-ahead hypothesis (based on Keating and Shattuck – Hufnagel 2002 among others) suggested that if the speakers have an option to build a prosodic structure of the entire sentence before its spellout, they would adjust the sizes of both boundaries according to reflect a high attachment or a low attachment structure (references). The No look-ahead hypothesis, based on Levelt's (1989) approach, suggested that the speakers might lookahead of maximum two words, thus being able to adjust only the size of a later boundary B.

The examination of word durations before boundaries, which served as a measure for the size of the boundaries, revealed a significantly bigger correlation in the high attachment short adverbial condition than their correlation between word durations at the low attachment short adverbial condition (section 4.3). This finding suggests that the longer the word durations at the earlier boundary, the longer the word durations at the later boundary and that the speakers strove to have the later boundary larger than the earlier one. Thus, the Rational Speaker Hypothesis was confirmed for production.

The correlations between the word durations also constitute evidence in favor of the No look-ahead hypothesis and of Levelt's (1989) approach, that the speakers do not have a large look ahead at the time of spell out. Additionally, no main effect was found in the values of pre boundary A word durations between the high attachment short adverbial and low attachment short adverbial condition. This indicates that the size of boundary was not adjusted to reflect the syntactic structure, presumably since the size of boundary B was yet unknown at the moment of planning the acoustic realization of boundary A. Additional examination of the long adverbial condition at boundary A revealed an unexpected near main effect between the high attachment and the low attachment structures. At this point, it is impossible to say whether this effect represents rule in a different parsing of the sentence in the high long and the low long attachment condition, or is just confound of the specific experimental stimuli.

The examination of word durations at boundary B produced several findings, the first of which was the main effect between the high attachment and the low attachment conditions in the short adverbial data subset. This finding might also reflect the scaling of boundaries size, which is produced by increasing the size of boundary in the high attachment but not in the low attachment structures (to make it bigger than the earlier boundary). Another finding was the replication of already established phenomenon of bigger boundaries being realized before longer constituents (eg. Warren and Gibson 2004a,b). The examination of the low attachment condition revealed the word durations in the long adverbial condition were significantly longer than word durations in the short adverbial condition, confirming that the participants produced a larger boundary before the longer constituent. Additionally, the examination of the high attachment condition did not reveal any significant difference between word duration in the short adverbial and the long adverbial conditions, confirming the possibility that boundaries of the same type are used to signal both a high attachment and a long constituent in both perception (Carlson, Clifton and Frazier 2006) and production.

4.2 Examination of the findings in the broader context

The data obtained in this study brings further support to several issues which are already discussed in the literature. First of all, it supports the approach that the realization of prosodic boundaries is dependent on the preceding context. As already mentioned in section 2.3 this issue was extensively explored in perception studies, but only one production study attempted to explore this issue. Therefore, the experiment presented in this paper brings further evidence that the global (meaning of the whole sentence) prosodic context is considered when realizing a prosodic boundary.

Another dimension that this experiment has touched is this of prosodic planning. The manner in which the prosodic scaling was realized supports speech planning models which assume the lack of (large) look ahead in planning the production of an utterance. No evidence was found in this study that the earlier prosodic boundary is affected by the type of the later boundary. Interestingly, these results seem to contradict the results of the experiments in Krivokapić (2012), which claim to show evidence for the models which propose a large look-ahead. Specifically, the findings showed that the duration of a pause at an earlier prosodic break was affected by both the length of the phrase following it, and by the length of the phrase following a later prosodic boundary: when the latter phrase was longer, the pause following the earlier prosodic boundary was longer too. However, Krivokapić does not discuss the following: (1) the effects of manipulations on the later pause and (2) the option that what affected the duration of the first pause was not (just) the size of the later constituent, but the **total** size of the sentence portion that followed the first pause (which includes the portion between the first and the second pause and the portion following the second pause). Therefore, increase in the length of the attachment after the second pause leads to increase in the total post boundary portion of material. Perhaps what happened was that the **total** sentence portion following the first pause was parsed as one prosodic constituent, hence the longer pause to allow the planning of a longer string (Watson and Gibson 2004b).

The support for the No look-ahead hypothesis might also be surprising regarding the well established fact that the speakers are able to adjust the initial height of F0 and the slope of its fall during the cross of the utterance as a function of the utterance length (Cooper and Sorensen 1981, Swerts et al 1996). For shorter utterances, the initial level of F0 is lower, and the declination slope is steeper, than for longer utterances. These data clearly indicate that the speakers do have some access to the whole utterance, thus large look ahead of at least some aspect of the utterance is available to them.

4.3 Possible directions for future research

Another issue that emerged in this study was the difficulty in eliciting the exact number of prosodic boundaries in particular predefined locations. Both the perception experiments conducted in support of the Rational Speaker Hypothesis, and the examinations of production data and the further perception experiment by Snedeker and Casserly (2010) assumed the presence of two and only two specific prosodic boundaries. The analysis in this study also concentrated on two specific predefined locations. However, sometimes participants produced (a) more than just two boundaries and/or (b) boundaries in locations other than the predefined A and B. The below figures show the additional locations of the boundaries in the sentences, produced in the high attachment short adverbial condition:

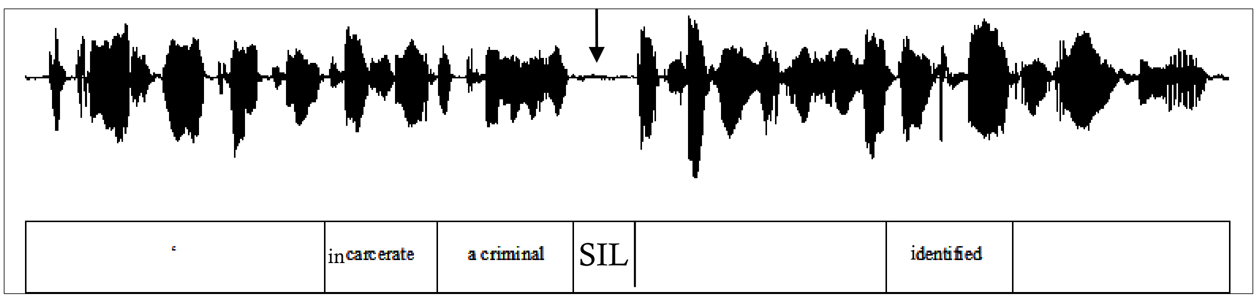

Figure 7: illustrating the additional boundary A1. The boundary is sigpresence of a relatively big prosodic boundary. The sentence is: ***The police will attempt to incarcerate a criminal that several witnesses identified an hour from now.***

The above waveform snapshot shows a distinctive pause after the word *criminal*, which indicates a presence of a larger than prosodic word boundary (Beckman and Pierrehumbert 1986, Beckman and Elam 1997). While the presence of pauses is not obligatory for the indication of a prosodic boundary, it nonetheless seems that the boundary after the segment *carcerate* – the intended location of boundary A – is much smaller than the one indicated on the snapshot. In this case, the participant chose to realize a break not after the main verb, but before the relative clause constituent. This boundary was termed A1, since it is located after the intended boundary A but before intended boundary B. Figure 8 below demonstrates a production from another speaker, which shows an additional boundary, termed boundary C, which was produced in the middle of the long adverbial. Additionally, boundary B was also clearly produced after the word *identified*.

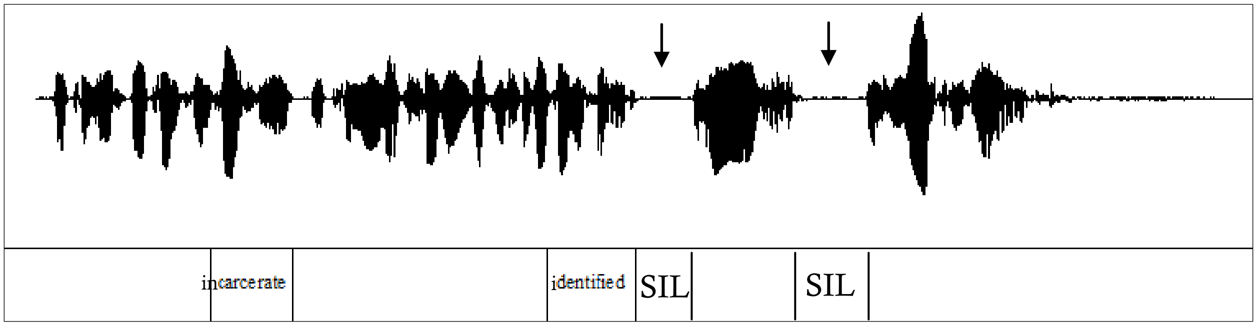

Figure 8: illustrating the additional boundary C. The large pauses, which are taken as a evidence for the presence of a relatively big prosodic boundary, are marked by arrows. The sentence is: *The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.*

The two logical options to explore here would be (1) how (or whether) the scaling was realized between boundaries A1 and B and (2) how the scaling was realized in cases where the long adverbial was pronounced as more than one constituent. One possibility for further examination would be to examine each production separately (as opposed to the quantitative measurements performed in this study) and to estimate the magnitudes of boundaries in the more detailed manner. An appropriate method could be to use the ToBI transcription system (Beckman and Elam 1997), which takes into consideration factors such as juncture strength, presence of pauses and the tonal patterns. The option of performing quantitative measures on these additional boundaries is also possible, however, it would require a larger scale experiment.

6. Conclusion

This study has examined the scaling of prosodic boundaries in speech production. For this purpose, stimuli sentences intended to elicit two prosodic boundaries were constructed. The data on word durations prior to the two boundaries collected from 16 participants confirmed that the sizes of prosodic boundaries were in different proportions as a function of the underlying syntactic structure of the sentence. Thus, influence of the global prosodic structure on the realization of the boundaries was confirmed for production.

Confirming the evidence from previous perception experiments, the study brings additional support to production models which postulate a lack of large look-ahead in the planning of prosodic boundaries in sentence production. Possible directions for the future research may include a more particular reanalysis of the productions, both for categorical classification of the boundaries, but also additional methods of measurement of pre-boundary segments.

References:

Anderson, C., and Carlson, K. (2010). Syntactic structure guides prosody in temporarily ambiguous sentences. *Language and Speech*, 53, 472-493.

Bates, D., Maechler, M. and Bolker, B. (2012) lme4: Linear mixed-effects models using S4 classes. URL: http://lme4.r-forge.r-project.org/

Becker, M. and Levine, J. (2010): Experigen, an online experimental platform. [Computer Program]

Beckman, M. E., and Elam, G. A. (1997). Guidelines for ToBI labelling, version 3.0. Manuscript and accompanying speech materials. Columbus, OH: Ohio State University.

Beckman, M. E., and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255-309.

Boersma, P. and Weenink, D. (2012): Praat: doing phonetics by computer. [Computer program]

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. Phonetica, 57, 3-16.

Byrd, D., and Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. Journal of Phonetics, 26, 173-199

Carlson, K., Frazier, L., and Clifton, C., Jr. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, 119, 1066–1082

Carlson, K., Clifton, C. Jr., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. Journal of Memory and Language, 45(1), 58-81.

Clifton, C., Jr., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech,* 45, 87-114.

Clifton, C. J., Frazier, L., and Carlson, K. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. Psychonomic Bulletin and Review, 13(5), 854-861.

Cooper, W. E. and J. Paccia-Cooper (1980). Syntax and Speech. Cambridge, MA: Harvard University Press.

Cooper, WE, and Sorensen, JM (1981). *Fundamental frequency in sentence production.* New York/Berlin: Springer-Verlag.

Dehé, Nicole. 2009. Clausal parentheticals, intonational phrasing, and prosodic theory. Journal of Linguistics, 45, 569-615.

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

D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., and Vigário, M., (2005). Intonational Phrasing in Romance: The role of prosodic and syntactic structure. In *Prosodies*, eds. Sonia Frota, Marina Vigário and M.J. Freitas, 59-97. 54 Berlin: Mouton de Gruyter.

Edwards, J., Beckman, M. E., and Fletcher, J. (1991). The articulatory kinematics of final lengthening. The Journal of the Acoustical Society of America, 89, 369-382.

Ferreira, F. (1988). Planning and timing in sentence production: The syntax-to-phonology conversion. Unpublished doctoral dissertation, Amherst, MA: University of Massachusetts.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, 30 (2), 210-233.

Ferreira, F. (1993) Creation of prosody during sentence production. *Psychological Review 100*: 233-253

Gee and Grosjean (1983) Performance Structures: A Psycholinguistic and Linguistic Appraisal. Cognitive Psychology 15, 411-458.

Jackendoff, R. S. (1977). X'-syntax: A study of phrase structure. Cambridge, MA: MIT Press.

Keating, P. and S. Shattuck-Hufnagel (2002). A prosodic view of word form encoding for speech production. UCLA Working Papers in Phonetics, 101, 112-156.

Kisseberth, Charles. 1994. On domains. In *Perspectives in Phonology*, eds. Jennifer Cole and Charles Kisseberth, 133-166. Stanford, CA: CSLI Publications.

Kjelgaard, M. M., and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. Journal of Memory and Language, 40(2), 153-194.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics, 3, 129-140.

Krivokapic, J. and Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetic*s. Volume 40, 3, 430–442.

Kubozono, H. 1993. *The organization of Japanese prosody*. Tokyo: Kurosio.

Ladd, D. R. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology* 3:311-340.

Ladd, D. R. 1988. Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* 84:530-544.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. Glossa, 7, 107-122.

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

McCarthy, John. 2003. OT constraints are categorical. Phonology, 20,75-138.

Meyer, A. S., Belke, E., Häcker, C., and L. Mortensen (2007). Use of word length information in utterance planning. Journal of Memory and Language, 57, 210-231

Nespor, M. (1993) Fonologia. Bologna: il Mulino.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Foris: Dordrecht.

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. Journal of the Acoustical Society of America, 9, 2956-2970.

Prince, Alan and Paul Smolensky (2004). Optimality Theory: constraint interaction in generative grammar. Malden, Mass. and Oxford: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science. [Available at: http://roa.rutgers.edu/.]

Robert, C. and Casella, G. (2004): Monte Carlo Statistical Methods.

Schafer, A. J., Speer, S. R., Warren, P., and White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.

Selkirk, E., (1978/1981). On prosodic structure and its relation to syntactic structure. In *Nordic Prosody II: Papers from a symposium*, ed. Thorstein Fretheim, 111-140. Trondheim: TAPIR.

Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In *Prosody: Theory and experiments*, ed. Merle Horne, 231-262. Dordrecht: Kluwer.

Selkirk, E. (2005). Comments on Intonational Phrasing in English. In Sonia Frota, Marina Vigario, Maria Joao Freitas, eds. 2005. *Prosodies.* Mouton de Gruyter, Berlin.

Selkirk, E. (2008). Contrastive focus, givenness and the unmarked status of "discourse-new". *Acta Linguistica Hungarica* 55:331-346.

Selkirk, E., (2009). The syntax-phonology interface. In *The Handbook of Phonological Theory*, eds. John Goldsmith, Jason Riggle and Alan Yu. Oxford: Blackwell.

Selkirk, E. (2012). Prosodic headedness and prosody typology: effects of the constraint family Hd(π) tone. Talk given at the Fifth European Conference on Tone and Intonation (TIE5), Oxford: England (handout).

Selkirk, E., and Tateishi, K. (1988). Constraints on minor phrase formation in Japanese. In *Proceedings of the 24th annual meeting of the Chicago Lingusitics Society*, eds. M.G. Larson and Diane Brentari, 316-336. Chicago IL: Chicago Linguistics Society.

Selkirk, E., Shinya, T., and Kawahara, S. (2004). Phonological and phonetic effects of phrase length on F0 in Japanese. In *Proceedings of the second international conference on speech prosody 2004*, eds. Bernard Bel and Isabelle Marlien, 183-186. Nara, Japan.

Snedeker, J., and Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. Language and Cognitive Processes, 25, 1234-1264.

Snedeker, J., and Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. Journal of Memory and Language, 48, 103-130.

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature Neuroscience, 2(2), 191–196.

Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. Journal of Psycholinguistic Research, 25(2), 193-247.

Swerts, M., Strangert, E., and Heldner, M. (1996) '*F*0 declination in spontaneous and read-aloud speech. Proc. International Conference on Spoken Language Processing, Philadelphia, October 1996, pp. 1501–1504.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. Linguistic Inquiry, 30,219-256.

Truckenbrodt, H. (2002). Upstep and embedded register levels. Phonology, 19(1), 77_120.

Turk, A. E., and Shattuck-Hufnagel, S. (2007) "Multiple Targets of Phrase-final Lengthening in American English Words." *Journal of Phonetics* 35 (4): 445-472.

Watson, D. G., and Gibson, E. A. (2004a). Making sense of the sense unit condition. Linguistic Inquiry, 35, 508-517.

Watson, D. G., and Gibson, E. A. (2004b). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713-755.

Watson, D., and Gibson, E. A. (2005). Intonational phrasing and constituency in language production and comprehension. Studia Linguistica, 59(2-3), 279-300.

Wheeldon, L., and Lahiri, A. (1997). Prosodic units in speech production. Journal of Memory and Language, 37, 356_381

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America, 92, 1707-1717.

Appendix A: Stimuli Used in the Experiment

<u>Stimuli Sentences</u>:

1. HS: Sarah will definitely try to evaluate some data she obtained from her experiment later this week.

   LS: Sarah will definitely try to evaluate some data she obtained from her experiment earlier this week.

   HL: Sarah will definitely try to evaluate some data she obtained from her experiment later this week after the start of spring semester.

   LL: Sarah will definitely try to evaluate some data she obtained from her experiment earlier this week at the start of spring semester.

2. HS: The knights will effectively eliminate some enemies that managed to sneak into their residence an hour from now.

   LS: The knights will effectively eliminate some enemies that managed to sneak into their residence an hour ago.

   HL: The knights will effectively eliminate some enemies that managed to sneak into their residence an hour from now when the gates will open.

   LL: The knights will effectively eliminate some enemies that managed to sneak into their residence an hour ago when the gates were open.

3. HS: The evil aliens will mercilessly annihilate six planets they have spotted through the telescope one time cycle from now.

   LS: The evil aliens will mercilessly annihilate six planets they have spotted through the telescope one time cycle ago.

   HL: The evil aliens will mercilessly annihilate six planets they have spotted through the telescope one time cycle from now as they enter the star system.

   LL: The evil aliens will mercilessly annihilate six planets they have spotted through the telescope one time cycle ago as they entered the star system.

4. HS: Clark Kent will probably be assigned to interview several cops who helped to arrest a dangerous criminal early next Monday.

LS: Clark Kent will probably be assigned to interview several cops who helped to arrest a dangerous criminal early last Monday.

HL: Clark Kent will probably be assigned to interview several cops who helped to arrest a dangerous criminal early next Monday in front of the Mayor's house.

LL: Clark Kent will probably be assigned to interview several cops who helped to arrest a dangerous criminal early last Monday in front of the Mayor's house.

5. HS: The police will attempt to incarcerate a criminal that several witnesses identified an hour from now.

LS: The police will attempt to incarcerate a criminal that several witnesses identified an hour ago.

HL: The police will attempt to incarcerate a criminal that several witnesses identified an hour from now when he'll walk into a bank.

LL: The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

6. HS: The Hilton will definitely begin to renovate some rooms which the drunken tourists vandalized next April.

LS: The Hilton will definitely begin to renovate some rooms which the drunken tourists vandalized last April.

HL: The Hilton will definitely begin to renovate some rooms which the drunken tourists vandalized next April during the slow tourist season.

LL: The Hilton will definitely begin to renovate some rooms which the drunken tourists vandalized last April during the busy tourist season.

7. HS: Sally will eagerly begin to implement a plan she proposed to the church committee at the start of next week.

LS: Sally will eagerly begin to implement a plan she proposed to the church committee at the end of last week.

HL: Sally will eagerly begin to implement a plan she proposed to the church committee at the start of next week when the holidays end.

LL: Sally will eagerly begin to implement a plan she proposed to the church committee at the end of last week when the holidays began.

8. HS: The university will go ahead and compensate all female scientists it unfairly ostracized later this month.

   LS: The university will go ahead and compensate all female scientists it unfairly ostracized earlier this month.

   HL: The university will go ahead and compensate all female scientists it unfairly ostracized later this month after the start of spring semester.

   LL: The university will go ahead and compensate all female scientists it unfairly ostracized earlier this month at the end of spring semester.

<u>Filler sentences</u>:

1. The scientist observed the mice through the transparent partition.
2. The soloist learned the song with the help of a recording.
3. The prisoner revealed the names were the part of the code.
4. As Louisa undressed the baby started to cry.
5. While the thief hid the elegant jewelry sparkled in the light.
6. While Eric played the piano the harp snapped a string.
7. While Rupert ate the soup the main dish grew cold.
8. Elaine discovered her neighbor had flown planes in the war.

<u>Reading practice sentences</u>:

1. Sue found the marinara after the search through the freezer.
2. The witness described the security guard of the minister who arrived in the late night flight.
3. While the violinist practiced the concerto the symphony blared from the neighbor's radio.

Audio demo sentence:

1. The software will quickly and flawlessly synchronize each file that the user has modified.

Appendix B: screenshots from the software used in the experiment, detailing the instructions given to the participants and the course of producing one experimental item.

Welcome!

In this experiment you will see various sentences, which you will be asked to read.

The procedure is as follows:
1. You will see a sentence and read it silently to yourself.
2. You will answer three questions about the sentence.
3. You will then pronounce the sentence out loud twice, each time in a different manner.

All instructions will appear on the screen as the experiment progresses.

If you have any questions at this point, please ask the experimenter.

If you're ready, please press the button and proceed to the training.

continue

Please read the sentence silently to yourself:

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

continue

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

Question 1: What will the police do?

- Attempt to incarcerate.
- Begin to execute.

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

Question 2: Who will the police attempt to incarcerate?

- A protester that hit a policeman.
- A criminal that several witnesses identified.

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

Question 3: When did the witnesses identify the criminal?

- Three hours ago when he contacted his mother.
- An hour ago he walked into a bank.

Please read the sentence out loud ***in your usual manner****.*

You may repeat the sentence if you feel that you didn't read it in a manner that you intended.

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

continue

Please read the sentence out loud, ***in a more deliberate manner****, as if you are giving a speech on stage in front of an audience*

You may repeat the sentence if you feel that you didn't read it in a manner that you intended.

The police will attempt to incarcerate a criminal that several witnesses identified an hour ago when he walked into a bank.

continue

Appendix C: results from the analysis of pitch reset.

The analysis of pitch reset data revealed that it main effects did not correspond to the main effects discovered for the word durations, and in fact, did not show any regular patterns. The only interesting result that was found was its interaction with the speech manner in the high attachment structures, which can be seen in Figure 8 below:

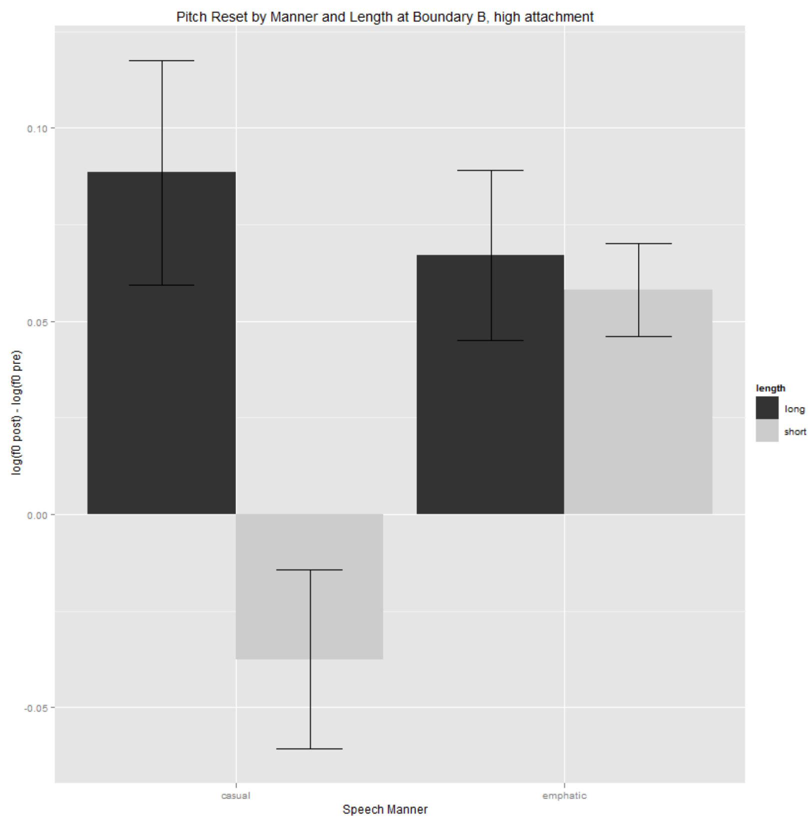

Figure 8: interaction between the speech mode and the length of adverbial at high attachment condition at boundary B. Pitch reset means are indicated by the bars, while the standard error is indicated by the whiskers.

The figure shows that when the manner of speech was casual, and the adverbial was long the pitch reset was relatively large; when the adverbial was short the pitch reset was negative – meaning that there was no interruption in the declination of pitch level across the boundary site. Therefore it is possible to conclude that in the casual speech, the parsing of material into prosodic constituents is affected by the length of the utterance. It seems then that when the temporal adverbial was short, it was parsed into the same prosodic constituent as the preceding material, and therefore no boundary was realized at B. On contrary then, when the temporal adverbial was long, it was parsed into a separate constituent. Although this explanation seems probably, it is still unclear why this method of parsing is restricted to (1) pitch reset (2) casual speech (3) high condition only.

It is known that the English speakers can realize more than one specific tonal pattern, including the boundary tones, which can be either high or low (for the full inventory of English tones, please see Pierrehumbert 1980, Beckman and Pierrehumbert 1986). Should half of the participant realize the pre- boundary tone as L-/L% and the post boundary tone as H*, and the second half of the participants to reverse this pattern, the cumulative patterns of speech would be evened out. Related to this fact, an additional possible acoustic measurement of pitch reset would be to measure the difference between the H* peaks as close as possible to the boundary (before and after it, Backman and Pierrehumbert 1986) This might help to account for the fact that the pitch reset might not start from the very first segments after the boundary, but might be starting from the topmost high tone, as demonstrated in Figure 9 below. This figure clearly shows, that although the next constituent after the pause seem to start at the word *that* (the next word after the break), the highest pitch is not placed there, but on the next word, *several*. A possible reason for this shift could be *that* being a function word, with the f0 target preferring a content word for its realization.

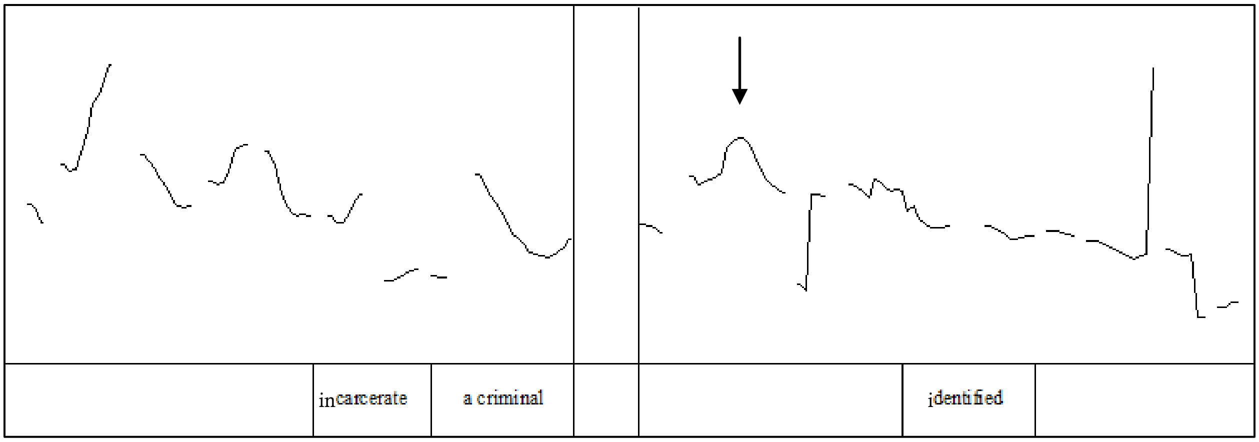

Figure 9: pitch track from the same file as in Figure 8, illustrating the pitch reset mismatch. The vertical lines mark the location of the pause at boundary A1 between words *criminal* and *that*. The vertical arrow indicates the location of the post boundary f0 peak.

Appendix D: analysis of proportions of word durations A/ word durations B.

This analysis examined the possibility to reveal the influence of length of boundary A on the length of the subsequent boundary by examining the proportions between word durations A and word durations B. The rationale was that one should expect to see the following in the results: word durations A/ word durations B < 1 for high short attachment condition; word durations A/ word durations B > 1 for the low attachment conditions. Ideally, these proportions should repeat themselves in the by-participants and by-items data, since this is supposed to be something that people constantly do, and also hold for casual and emphatic speech styles. In long attachment condition it is unclear how the boundary B should be scaled with respect to boundary A.

The figure below shows side by side the proportions of high short and low short conditions by participants. The horizontal line indicates the value of 1, to which the proportions are compared.

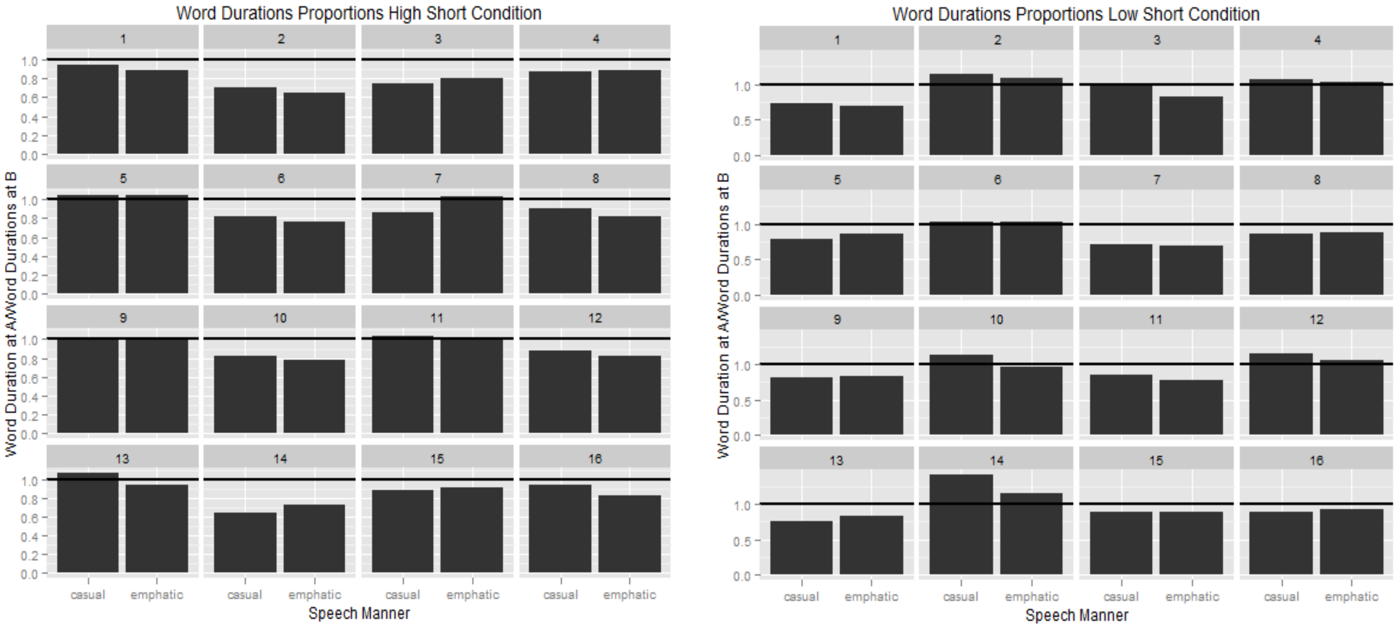

Figure 10: mean proportions by participants of word durations A/ word durations B; high short condition and low short condition across the casual and emphatic manners of speech. Columns higher than the horizontal line show instances in which that word durations at boundary A were bigger than word durations at boundary B; columns lower than the horizontal line indicate the opposite result.

The ideal picture that the figure above could portray is that for high short attachment condition all the columns would be below the black line, while for the low short condition the columns would be below the black line. However, this is not the case. Although the patterns do look very similar across casual and emphatic speech styles, no opposite trends between the two conditions can be found. The comparison of the same two conditions by item (Figure 11 below) also does not show any clear patterns. In fact, it seems that there was a pattern for each item that was preserved in both conditions:

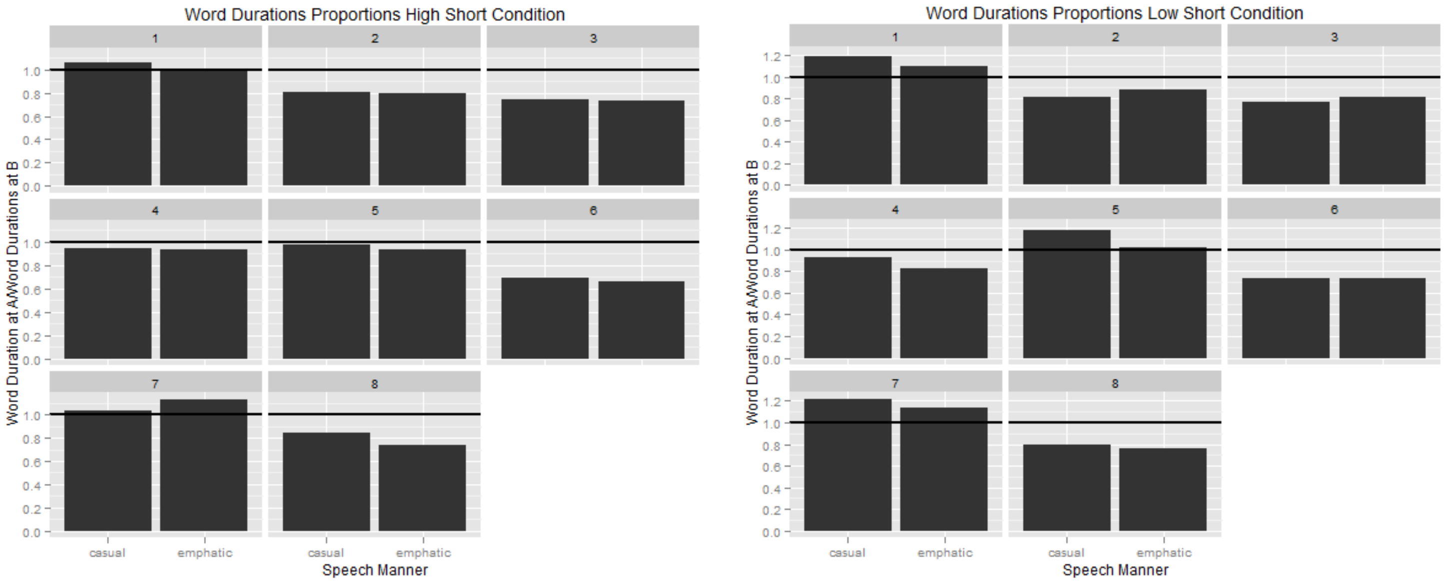

Figure 11: mean proportions by items of word durations A/ word durations B; high short condition and low short condition across the casual and emphatic manners of speech. Columns higher than the horizontal line show instances in which that word durations at boundary A were bigger than word durations at boundary B; columns lower than the horizontal line indicate the opposite result.

There were no predictions with respect to the scaling that can be found in the long condition. Nonetheless, it was examined to see whether there will be differences in trends between the high and the low attachment conditions. The examination of the long condition also did not reveal any differences in both by subject and by item views. The figures for the former and the latter are presented below:

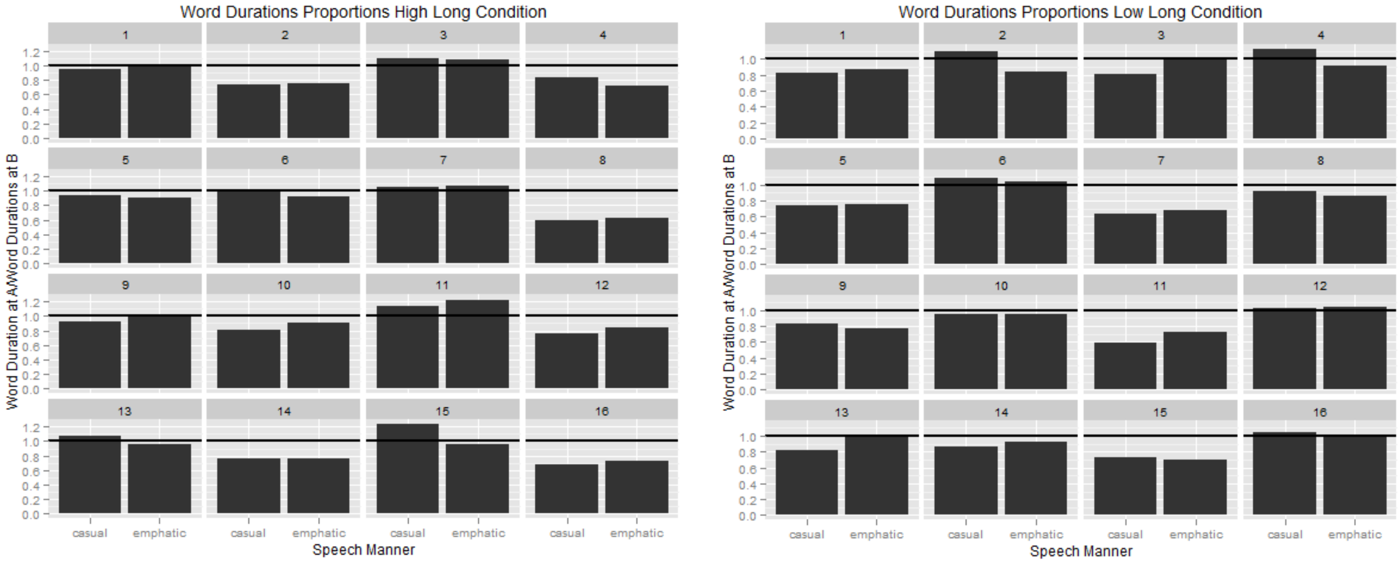

Figure 12: mean proportions by participants of word durations A/ word durations B; high long condition and low long condition across the casual and emphatic manners of speech. Columns higher than the horizontal line show instances in which that word durations at boundary A were bigger than word durations at boundary B; columns lower than the horizontal line indicate the opposite result.

Considering that the main indicator of the difference in trends is whether some proportion is smaller or bigger than one, as also for the short condition, no clear trends or clear differences seem to occur between two conditions. The by item view in Figure 13 below does not shed any further light either.

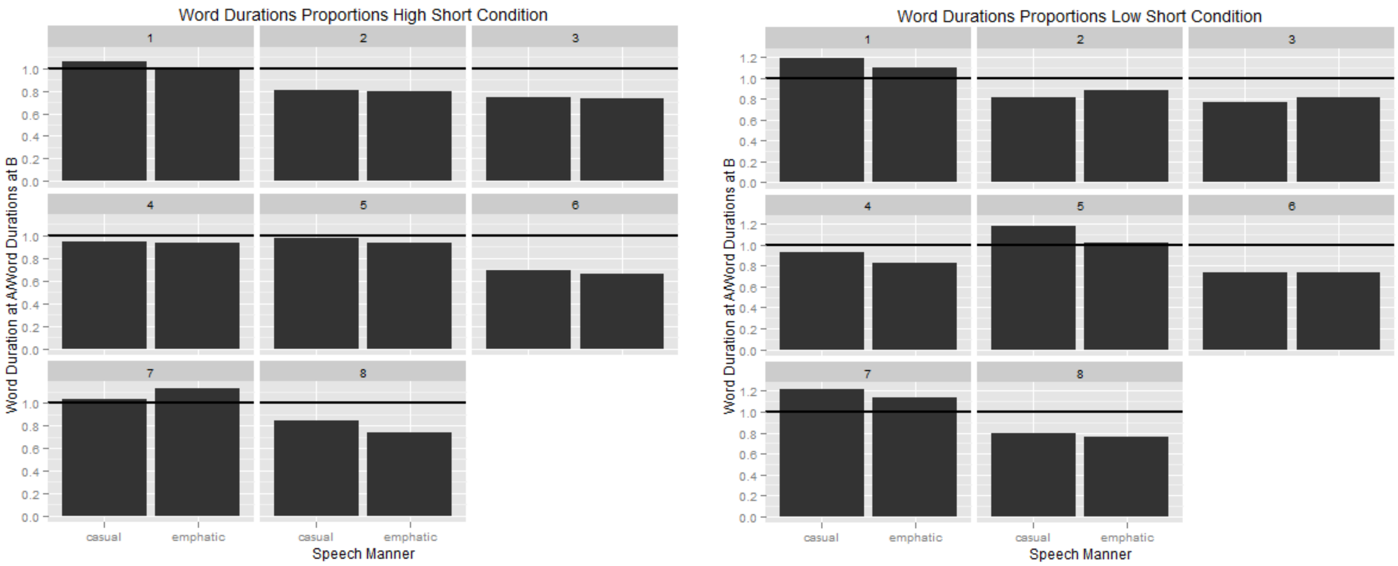

Figure 13: mean proportions by items of word durations A/word durations B; high short condition and low long condition across the casual and emphatic manners of speech. Columns higher than the horizontal line show instances in which that word durations at boundary A were bigger than word durations at boundary B; columns lower than the horizontal line indicate the opposite result.

Appendix E: full analysis of correlations

This section summarizes the Pearson's R correlation analysis performed on the word durations before boundaries A and B, called Numerator A and Numerator B in Figure 14 below. The purpose of this analysis was to test a hypothesis that English speakers could scale the pre- boundary word durations as predicted in the Rational Speaker Hypothesis by keeping the word durations at boundary A constant **but** adjusting the word durations at boundary B **only** such that in the high short attachment conditions, Numerator A < Numerator B, while at the long short attachment condition Numerator A > Numerator B. Therefore, in the case of high short attachment, the **ideal** situation would be to find a strong positive correlation between the values of Numerator A and Numerator B. Another condition would be that similar correlations to these found in the high short attachment condition would **not** be found between Numerator A and Numerator B for the low short attachment condition and also between Numerator A and other chunks of the sentence.

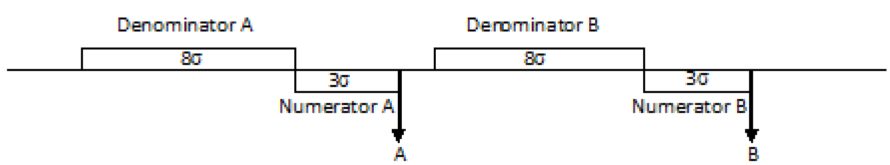

Figure 13: the schema for the word durations in various parts of the sentence. Numerators A and B are the three syllables measured for the word durations, while denominators A and B are the eight syllables preceding each of the numerators.

Therefore, Pearson's R correlations between the following chunks of the sentence were performed (see Figure 13 for terminology):

Numerator A and Numerator B to test the correlation between pre boundary words.

Numerator A and Denominator B to test the correlation between pre- boundary A word and a later chunk of the sentence.

Denominator A and Denominator B to test correlation between two portions of the sentence other than pre- boundary words.

The analyses were performed on each of the conditions: high short attachment, low short attachment, high long attachment and low long attachment. The short attachment condition is of the major interest, since it should be the one where the correlation between the pre boundary words (numerators A and B) should not be influenced by other factors such as length.

The results of Pearson's R tests are presented in the table below.

	NumA and Num B	NumA and DenomB	DenomA and DenomB
High Short	**0.78**	0.42	0.28
Low Short	0.49	0.41	0.37
High Long	0.45	0.47	0.28
Low Long	0.57	0.44	0.16

Table 3: Pearson's R correlation results. Correlations between portions of the sentences other than the pre-boundary words proved to be significantly smaller than the high attachment short adverbial condition, which is the main indication for the scaling of boundaries A and B.

It is possible to see that the R correlation for the high short condition (marked in red) is the highest among the rest of the correlations. The table also shows that the correlations between the Numerators in the short conditions are the highest than the correlations between the Numerators and Denominators (Figure 13) and between the two Denominators. Therefore, one might say that the correlations between the pre boundary words are not a function of the speech rate of some local sentence chunk.

The denominators were initially measured to provide an additional scaling of the word durations was made in order to account for the rate of speech for each individual production. In order to do the scaling, the length of 8 preceding the pre- boundary words of A and B syllables was measured. Then two measures of scaling were performed: the first was dividing the pre- boundary word by its respective denominator. The second was to divide the pre- boundary words by the mean duration of the 8 syllables preceding the pre- boundary A and the pre- boundary B words.

The fractions from both scaling were then arcsine transformed, an action which is supposed to normalize the fractions. However, the examination of the distribution of the pre- arsine numbers revealed that they already were normally distributed. Same statistic tests as before were performed on the scaled word durations: both arcsine transformed and not. The results were nearly identical to these of the raw words durations: the only difference was that the effects of speech style which would often disappear.

Appendix F: coefficients and intercepts for the linear models

	Data subset	Linear Model Factors		
Boundary A	Short adverbial condition		COEFFICIENT	*p*
		Intercept	468.466	<0.0001
		height	-1.241	0.777
		manner	2.108	0.631
	Long adverbial condition		COEFFICIENT	*p*
		Intercept	469.073	<0.0001
		height	6.777	0.07
		manner	-12.311	0.001
	High attachment condition		COEFFICIENT	*p*
		Intercept	473.251	<0.0001
		length	2.677	0.486
		manner	-6.756	0.080
	Low attachment condition		COEFFICIENT	*p*
		Intercept	464.365	<0.0001
		length	-1.992	0.641
		manner	-6.796	0.114
		length*manner	-7.705	0.07
Boundary B	Short adverbial condition		COEFFICIENT	*p*
		Intercept	537.04	<0.0001
		height	17.37	0.0004
		manner	-10.76	0.027
	Long adverbial condition		COEFFICIENT	*p*
		Intercept	544.471	<0.0001
		height	5.098	0.287
		length	-11.780	0.014
	High attachment condition		COEFFICIENT	*p*
		Intercept	551.990	<0.0001
		length	-2.422	0.618
		manner	-12.160	0.01
	Low attachment condition		COEFFICIENT	*p*
		Intercept	529.522	<0.0001
		length	9.851	0.034
		manner	-10.385	0.026

The abbreviations used are as follows: height for height of attachment, length for length of adverbial, manner for speech manner. The signs of the coefficients imply the following: height of attachment – positive coefficient means that the word durations in the high attachment condition were longer than in the low attachment condition; length of adverbial – positive coefficient means that the word durations in the long adverbial condition were longer than in the short adverbial condition; speech manner – positive coefficient means that the word durations were longer when the sentence was pronounced in emphatic manner than when it was pronounced in the casual manner.