Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb25. Syntactic predictability influences duration

Claire Moore-Cantwell*

***Corresponding author's address: Linguistics, University of Massachusetts Amherst, Amherst, Massachusetts 01003, cmooreca@linguist.umass.edu**

Building on work by Gahl and Garnsey, 2004, this paper demonstrates that speakers "buy time" during the planning of upcoming low-probability syntactic structures by producing prosodic boundaries with longer duration before low-probability than before high-probability structures. Subject extraction cleft sentences ("It was Edward who (t) loved Lucy.") are more common in corpora than object extraction cleft sentences ("It was Edward who Lucy loved (t).") (Roland et al., 2007), and are also easier to process (e.g. Gibson, 1998). The duration of the clefted constituent ("Edward") was measured in planned productions of subject- and object-extraction clefts in English. In order to disentangle the probability of each structure from its difficulty level, the probabilities were manipulated within the experiment through training. Participants read aloud: First, two SE and two OE clefts; second, eight SE or eight OE clefts; and finally, another two SE and two OE clefts. Before training, the clefted constituent was longer in OE clefts (mean 407ms) than in SE clefts (370ms, t=2.4, p=.02). This difference was no longer present after OE training (OE: 385ms, SE: 397ms), but was still present after SE training (OE: 448ms, SE: 388ms).

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4801075]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

There is strong evidence that the frequency of lexical items, word-to-word transitional probabilities, and even probabilities over large groups of words (Arnon and Miller, 2011; Arnon and Snider, 2010) influence the phonetic duration of words. This paper demonstrates that probabilities over syntactic structures likewise influence the duration of words within those structures. I hypothesize that speakers 'buy time' at prosodic boundaries for the planning of upcoming low-probability structures.

This hypothesis (called the "Structural Probability-Boundary Strength hypothesis" (SPBS)) builds on the "Left hand side/Right hand side Boundary hypothesis" (LRB) put forth by Watson and Gibson (2004). The LRB assumes that a chunk of syntactic structure gets planned at a high-level prosodic boundary (Intonational Phrase or Phonological Phrase) immediately preceding it. The LRB states that the longer and more complex the chunk of structure being planned, the more likely the speaker is to use a larger prosodic boundary. Since larger prosodic boundaries are implemented as increased duration on the phrase-final word (Wightman *et al.*, 1992), they give the production system more time to plan the difficult structure. The SPBS states additionally that low-probability structures are more difficult to plan than high-probability structures (perhaps because the planning mechanism gets less practice with them), and therefore that speakers implement larger prosodic boundaries before low-probability structures than before high-probability structures.

Gahl and Garnsey (2004); Gahl *et al.* (2006) argue that probabilities over syntactic structures are in fact explicitly stored as part of a speaker's knowledge of syntax. They use verbs which vary in the probability with which they take a particular type of complement. For example, the verb 'decide' occurs much more often with a clausal complement in a corpus than with a direct-object or noun-phrase complement. On the other hand, the verb 'print' occurs much more often with a noun-phrase (NP) complement than with a clausal complement. Participants read aloud sentences whose complements matched and mismatched the verb's bias, as in Table 2.

TABLE 1: Verbs like 'decide' are strongly biased to take a particular type of object

Clausal complement bias:	
Matching:	The experienced judge *decided the appeal* should be started right away.
Mismatching:	The experienced judge *decided the appeal* on the merit of the case.
NP complement bias:	
Matching:	The journal editor *printed the article* with the footnotes at the end.
Mismatching:	The journal editor *printed the article* had been slanderous to him.

They measured the durations of participants' productions of (1) the verb, and (2) the NP following the verb (the italicized material). The verb and the immediately following NP are phonologically and lexically identical between the matching and mismatching conditions, and differ only in the structural role that the NP is playing. In one case, the NP is the complement of the verb, while in the other case it is the subject of an embedded clause which is the complement of the verb.

They found that words in bias-mismatching contexts (low probability structures) were longer than their counterparts in bias-matching contexts (high-probability structures). Whether that difference showed up on the NP or on the verb varied with bias type. For clausal complement biased verbs, it was the following NP that was longer in the mismatching context (an NP-complement) than in the matching context (a clausal complement). For NP-bias verbs, it was the verb that was longer in the mismatching context (a clausal complement) than in the matching complement (an NP complement). They conjecture that the two bias types behave differently because duration is being manipulated just at unpredictable prosodic boundaries.

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

Gahl and Garnsey claim that their results show that there is explicit representation of probability in the syntactic grammar, but they do not rule out the possibility that syntactic probabilities are only affiliated with particular lexical items. That is, upon looking up a lexical item (say, 'claim') speakers access one of two lexical 'subentries': $\textsc{Claim}_1$, which requires an SC-complement, and $\textsc{Claim}_2$, which requires an NP-complement. Each of these subentries might have a lexical frequency affiliated with it.

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

TABLE 2: Cleft structures

Subject-extraction (SE):	It was Edward who (t) scammed Melvin out of some money.
Object-extraction (OE):	It was Edward who Melvin scammed (t) out of some money.

Subject- and Object- extraction cleft sentences are very similar syntactically, and differ only in gap position. However, subject-extraction is more frequent in a corpus than object-extraction. In both cleft types, we might expect a large prosodic boundary at the right edge of the moved constituent (here 'Edward), as it is the last thing before an embedded CP (beginning with 'who'). Any duration difference evident between these two structures is therefore likely to manifest on the extracted constituent.

In addition to being less frequent, object-extraction is also more complex in that it contains a longer filler-gap dependency. Evidence for thinking of sentences with longer dependencies as more syntactically complex comes from experiments with relative clauses, showing that object extraction is generally more difficult to process than subject-extraction (Gibson, 2000; Warren and Gibson, 2002, among many others). In order to distinguish the effects of probability from the effects of complexity, I manipulate the probability of each structure locally within the experiment. One group of subjects encountered many examples of object-extraction, and few examples of subject extraction; a different group encountered many examples of subject extraction and only a few examples of object extraction. If a structure's probability is represented distinctly from its complexity in the production system, and that representation can be updated in a fairly short amount of time, we would expect the groups to behave differently according to how probable each structure was in the recent set of sentences they were exposed to.

The group for which object extraction is highly probable should produce words in object extraction structures shorter than those same words in subject extraction structures. The group for which subject extraction is highly probable should do the opposite: they should produce words in subject extraction structures shorter than those same words in object extraction structures. If the groups do not differ in their behavior, then we can conclude that probability does not explicitly influence production, and the complexity of a structure is a better predictor of the duration of words in it.

METHODS

Participants

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

Items

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

All items followed the form 'It was NP_1 who (NP_2) Verbed (NP_2) PP'. Some items has more than one prepositional phrase after the verb or NP_2, and those varied in whether they were an argument of the verb or not. The complete list of items (16 total) is given in the appendix. All sentences were presented visually, along with a context sentence of the form 'Did NP Verb NP PP ?'.

TABLE 3: Example items

SE:	Did John scam Melvin out of some money?
	It was Edward who scammed Melvin out of some money.
OE:	Did Melvin scam John out of some money?
	It was Edward who Melvin scammed out of some money.

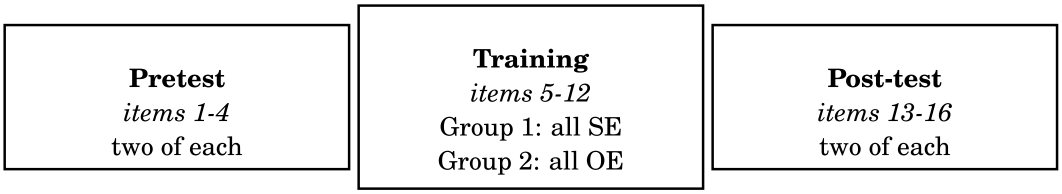
Procedure

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

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

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

FIGURE 1



Analysis

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

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

TABLE 4: Regions of interest measured

It was	Edward	who	scammed	Melvin	out of some money.
	1	2	3	4	

Several utterances contained disfluencies, and several utterances did not match the target sentence given to the participant. Counts of each type of error were recorded for each condition, but those utterances were removed from the duration analysis.

RESULTS

Duration differences

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

The clefted constituent

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

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

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

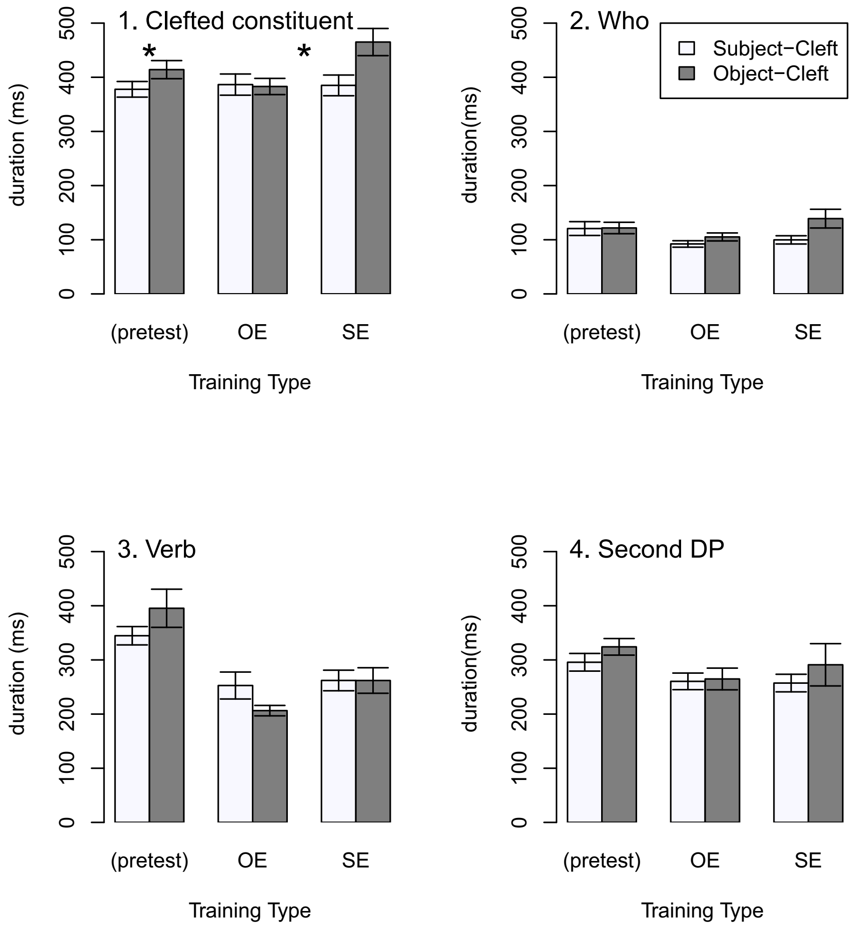
The Verb

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

Other elements

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

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



Accuracy differences

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

TABLE 5: Errors by training condition and error type

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

DISCUSSION

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

Duration differences did not occur on every word in the sentence, but rather were localized to the clefted constituent, the constituent most likely to be followed by a large prosodic boundary. Since duration is known to correlate strongly with prosodic boundary strength, I hypothesize that this clefted constituent is always followed by a prosodic boundary, whose strength is modulated by the probability of the structure of the upcoming phrase (that is, its extraction type).

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

TABLE 6: The LRB plus the SPBS

Watson and Gibson, 2004, pg. 732, augmented:The LRB weight at a word boundary between w_1 and w_2 is defined to be the sum of:	
a)	the size of the LHS constituent terminating in w_1, in terms of phonological phrases
b)	the projected size of the RHS constituent in phonological phrases starting at w_2, if this is not an argument of w_1;
c)	1, if w_1 marks the end of a phonological phrase.
d)	**1 minus the probability of the upcoming structure**

Upcoming syntactic structures take longer to plan the longer they are, and the more complex they are. Additionally, they take longer to plan the less probable they are, where probability is calculated over some recent, relatively short period of time. Low-probability structures may be more difficult to plan than high-probability structures simply because the production system gets less practice with them.

The need for extra planning time on low-probability structures is fulfilled via prosodic structure building. Specifically, placing a prosodic boundary or increasing the strength of its phonetic realization increases the amount of time available to the production system in which to plan an upcoming chunk of syntactic structure. Therefore, the production system aids itself by placing more boundaries with stronger phonetic implementations while planning long, complex, and low-probability structures.

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

ACKNOWLEDGMENTS

I wish to thank Lyn Frazier, Michael Wagner, Anne Pycha and Kristine Yu for invaluable advice and direction. I also wish to thank Florian Jaeger, Roger Levy, the UMass third-year seminar and poster-viewers at CUNY for helpful questions and comments. This work was supported in part by a National Science Foundation Graduate Research Fellowship to the author, and is dedicated to the glory of God.

REFERENCES

Arnon, I. and Miller, S. (**2011**). “More than words: phonetic duration is reduced in more frequent chunks”, Poster presented at CUNY.

Arnon, I. and Snider, N. (**2010**). “More than words: Frequency effects for multi-word phrases”, Journal of Memory and Language **62**, 67–82.

Boersma, P. and Weenink, D. (**2011**). “Praat: doing phonetics by computer [computer program].”, in *http://www.praat.org/*.

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

Gahl, S., Garnsey, S. M., Fisher, C., and Matzen, L. (**2006**). “"That sounds unlikely": Syntactic probabilities affect pronounciation”, in *28th Annual Conference of the Cognitive Science Society*.

Gibson, E. (**2000**). “The dependency locality theory: a distance-based theory of linguistic complexity”, in *Image, Language, Brain.*, edited by A. Marantz, Y. Miyashita, and W. O’Neil (MIT Press).

Gorman, K., Howell, J., and Wagner, M. (**forthcoming**). “Prosodylab-aligner: a tool for forced alignment of laboratory speech.”, (The annual conference of the Canadian Acoustical Association).

Roland, D., Dick, F., and Elman, J. L. (**2007**). “Frequency of basic english grammatical structures: A corpus analysis”, Journal of Memory and Language **57**, 348–379.

Wagner, M. and Watson, D. (**2010**). “Experimental and theoretical advances in prosody: a review”, Language and Cognitive Processes **25**, 905–945.

Warren, T. and Gibson, E. (**2002**). “The influence of referential processing on sentence complexity”, Cognition **85**, 79–112.

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

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 **91**, 1707–1717.