Two Corpus-Based Empirical Studies on Syntax-Prosody Interface

Guohao Dou

CS@Illinois; Summer Research Program at NUS

Contact Information:
Computer Science
University of Illinois Urbana-Champaign
Daniels Hall 0350A
1010 W. Green St.
Urbana, IL

Phone: +1 (217) 550 1454
Email: gdou2@illinois.edu

Abstract

It is intuitively true that something in the speech sound wave helps us disambiguate conversations. Theoretical linguists have long been working on an optimality-theoretic approach to study the alignment of prosodic and syntactic constituents. Also, empirical studies have been conducted intensively in lab environments. In this work I conducted two large-scale corpus-based empirical studies to explore the interface in an incremental manner. I found that results of previous empirical studies remain valid in spite of the richness of textual data.

Introduction

This topic has been an active field of research for a long while. Klatt [4] observed the co-occurrence of vowel phoneme lengthening and phrase-finals. Cooper et al. [3] attempted to mine the syntactic information in speech acoustics in 1980, finding that some phonological phenomena are closely related to syntactic constituents in statistical terms. Prince et al. [6] designed their experiments to show that, apart from acoustic statistics, subjects of their experiments can indeed utilize prosodic cues to disambiguate sentences. Also, they discovered that the usefulness of prosodic cues may vary across different types of ambiguity.

Optimality Theory (OT) [5] has played a crucial role in phonological research as a framework and optimality-theoretic approaches have been proposed to formulate the interface. Selkirk [7] proposed the Match theory as a set of universal constraints in the OT, in order to explain the correspondence of syntactic and phonological hierarchy. As stated by the Match theory, major syntactic constituents should be aligned with major prosodic constituents, which are usually marked by some outstanding cues of disjuncture.

However, before validating the theory, we need a reliable measure of the "majorness" of syntactic constituents, a topic that has also been widely studied. One of the algorithms, the one proposed by Watson et al. [8], is adopted both for convenience of implementation and domain-related reasonableness. A brief illustration of the algorithm is presented in the following section.

Syntax Tree Made Quantitative

Syntactic ambiguity is ubiquitous. The following is a syntactic counterpart of "hello world", demonstrating an age-old instance of *PP attachment*.

I saw a girl with a telescope

And this sentence has two legitimate parses.

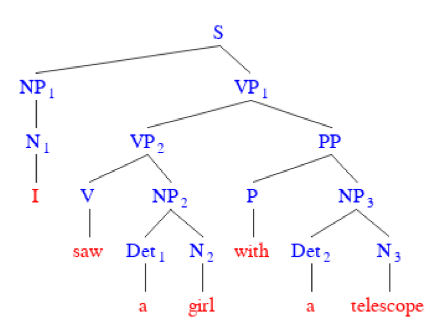

Figure 1: "with a telescope, I saw a girl."

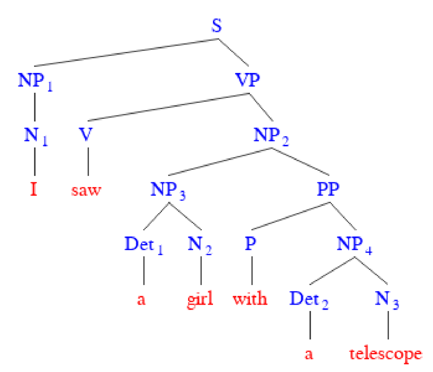

Figure 2: "I saw a girl and the girl is equipped with a telescope."

To conduct large-scale empirical study on the interface between syntax and other systems of human language, we first need a corpus[2] with the desired layers of annotation. We also need an algorithm with psycholinguistic ground that extracts numerical features from the tree.

Watson et al. [8] examined algorithms from previous study with controlled experiments. They discovered that even though all of them produced reasonable predictions and exhibited explanatory power, it is possible to switch to a much simpler algorithm without trading off statistical significance. They hypothesized that the appearance of disjuncture depends on both **LHS** and **RHS** of a word boundary, where **LHS** could be defined as following:

Denote the word immediately to the left of the boundary as w_1. **LHS** of the word boundary is the word count of the largest subtree of which w_1 is the right-most leaf.

while the definition of **RHS** could be easily mirrored.

As a small example to work on, consider the word boundary between "girl" and "with" in Figure 1 and Figure 2. Both have a RHS size of 3, since "with" is the left edge of PP in both cases. However, in Figure 1 "girl" is the right edge of the VP "saw a girl", while in Figure 2 it is the right edge of the NP "a girl". In conclusion, we should expect a more perceptible disjuncture after "girl" in Figure 1 than 2.

According to the initial statement of the LRB hypothesis proposed by Watson et al., the likelihood of phonological boundary depends on the both LHS and RHS of a word boundary. The greater the word counts are, the more perceptible the prosodic disjuncture should be.

Syntax v.s. ToBI break index

We study the correlation between prosodic break index and syntax structure, where we use ToBI annotation as the indicator for prosodic disjuncture, and LHS/RHS(described in the previous section) as the indicator for syntactic features.

For each word boundary, we compute its LHS/RHS value. The word boundaries are then grouped by their break indices. If the LRB hypothesis stated in the previous section holds, we should be able to observe that word boundaries with more outstanding disjuncture also receive larger LHS/RHS values. The result is shown in Table 1 and Table 2.

To interpret the grids of statistical significance, an entry in the i^{th} row and the j^{th} column stands for the p-value of the test that the i^{th} disjuncture type has a smaller LHS value than that of the j^{th} disjuncture type.

Break Index	Total	0	1-	1	2-	2	3-	3	4-	4
Count	61616	3192	2774	37193	309	1027	1079	2735	1744	11563

Table 1: Break index v.s. unsanitized LHS value: sample size

	0	1-	1	2-	2	3-	3	4-	4
0	x	0.401	0.327	0.0284*	7e-06**	0.0**	0.0**	0.0**	0.0**
1-	x	x	0.431	0.0318*	6e-06**	0.0**	0.0**	0.0**	0.0**
1	x	x	x	0.0323*	2e-06**	0.0**	0.0**	0.0**	0.0**
2-	x	x	x	x	0.476	0.0**	0.0**	0.0**	0.0**
2	x	x	x	x	x	0.0**	0.0**	0.0**	0.0**
3-	x	x	x	x	x	x	0.921	8e-06**	0.0**
3	x	x	x	x	x	x	x	0.0**	0.0**
4-	x	x	x	x	x	x	x	x	0.0**
4	x	x	x	x	x	x	x	x	x

Table 2: Break index v.s. unsanitized LHS value: statistical significance

Syntax v.s. Phoneme Lengthening

Setup

Preboundary lengthening is one of the cues for ToBI break index labelling[1]. Now that the category of ToBI break index has predictive power on the placement of syntactic boundaries, it is advisable to make one step towards raw speech sound wave and study the relationship between phoneme lengthening in word-final syllables and syntactic constituents.

Klatt [4] presented his research on how much of vowel duration is syntactically determined. According to his experiment on a small corpus, whether a vowel is in a word/phrase-final syllable or not turns out to explain 16.2% of the variance, ranked only next to the inherent duration of a vowel among the factors studied (vowel type, phrase/word-final syllable, postvocalic consonants, syllable position/syllable count).

In this study both vowel type and postvocalic consonants are controlled by normalization, leaving syllable position/count uncontrolled in that it explains only 1% of the variance in Klatt's work.

Normalization is carried out in the following scheme:

$$d(v) = \frac{d(v,c) - \bar{d}(v,c)}{\sigma(v,c)}$$

where v denotes a vowel and c is its postvocalic consonant. $d(v)$ is the normalized duration of v. Note that c may fall into a special category called "NONE" when the syllable does not have a coda. $\bar{d}(v,c)$ denotes the mean duration of vowel v when postvocalic consonant is v. $\sigma(v,c)$ denotes the standard deviation of vowel v duration when postvocalic consonant is v.

The LHS value is computed in exactly the same way as the previous section.

Observation

16 vowel phonemes and 26 consonant phonemes are present in the corpus.

Regression is done between LHS value and normalized vowel duration. The result is as shown below in Table 6.

Phoneme	Slope	R^2	Phoneme	Slope	R^2
ae(æ in bat)	0.0815	0.0303	uh(ʊ in book)	0.106	0.0599
ah(ʌin butt)	0.0641	0.0154	aw(aʊ in bout)	0.0770	0.0674
er(ɛ in bird)	0.0517	0.0364	ax(ə in about)	0.0179	0.00186
aa(ɑ in bot)	0.0890	0.0552	ay(aɪ in bite)	0.0699	0.0233
ih(ɪ in bit)	0.0270	0.00852	iy(i in beat)	0.0506	0.0239
uw(u in boot)	0.0892	0.0441	ey(eɪ in bait)	0.0729	0.0505
ao(ɔ in bought)	0.0769	0.0514	eh(e in bet)	0.0685	0.0418
ow(oʊ in boat)	0.0259	0.00701	oy(ɔɪ in boy)	0.0382	0.0152

Table 3: a linear regression between LHS value and normalized vowel duration

Parameter is significantly nonzero with a p-value less that 10^{-6} for all vowel phonemes(two-tailed test). A one-tailed test of slope being positive will give us the same conclusion.

Conclusions

Empirical evidence about syntax-prosody interface can be reproduced in a large corpus with many speakers, rich vocabulary, and a wide range of topics.

References

[1] Mary E Beckman and Gayle Ayers. Guidelines for tobi labelling. *The OSU Research Foundation*, 3:30, 1997.

[2] Sasha Calhoun, Jean Carletta, Jason M Brenier, Neil Mayo, Dan Jurafsky, Mark Steedman, and David Beaver. The nxt-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language resources and evaluation*, 44(4):387–419, 2010.

[3] William E Cooper and Jeanne Paccia-Cooper. *Syntax and speech*. Number 3. Harvard University Press, 1980.

[4] Dennis H Klatt. Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3):129–140, 1975.

[5] John J McCarthy. *A thematic guide to Optimality Theory*. Cambridge University Press, 2002.

[6] P. J. Prince, M. Ostendorf, S. Shattack-Hufnagel, and C. Fong. The use of prosody in syntactic disambiguation. 1991.

[7] Elisabeth Selkirk et al. The syntax-phonology interface. *The handbook of phonological theory*, 2:435–483, 2011.

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