

Prosodic and syntactic boundaries in spontaneous English and Finnish speech

Anna Dannenberg¹ & Stefan Werner²

¹*University of Helsinki (Finland)*, ²*University of Eastern Finland (Finland)*
anna.dannenberg@helsinki.fi, stefan.werner@uef.fi

In this paper, we examine prosodic and syntactic boundaries in English and Finnish spontaneous speech corpora. Comparing prosodic and syntactic units can help to understand the interaction between prosody and syntax in both speech production and perception.

For analysing the prosodic structure of spoken language, especially of spontaneous speech, effective automatic applications have unfortunately been rare. A Continuous Wavelet Transform (CWT) based method [1] applies the weighted sum of f_0 , energy and segmental durations to represent prosodic signals in a two-dimensional time-scale plane akin to spectrograms. The results can be further enhanced with lines of maximum amplitude to produce a visual representation of the prosodic hierarchies of speech.

In this study we have used a CWT based tool to detect prosodic boundaries in spontaneous speech. The results have been compared with grammatical analysis of the same data to examine the relation of prosodic and syntactic units in spoken language.

Our English dataset is extracted from the Buckeye corpus (Ohio State University) [2]. Our corpus was comprised of a total of 10076 words of spontaneous English speech, consisting of samples from informal interviews of five female speakers of American English.

For comparison, we used data from Finnish dialect interviews conducted by the Institute for the Languages of Finland [3]. Our sample contains 2217 words of spontaneous Finnish speech from two speakers, one female and one male, both native speakers of different dialects of Finnish.

In preprocessing, comments of the interviewer were deleted, and the speech was divided into turns based on speaker changes and obvious pauses.

The data was prosodically segmented using a wavelet based tool. The weighted sum of normalized f_0 , energy and segmental durations was used as an input signal for the CWT. Prosodic boundaries were determined by tracking minima across scales in the resulting scalograms, lines of minimum amplitude. For the English data, the whole segmentation process was fully automatic and unsupervised. For the Finnish data, prosodic boundaries were tagged manually based on how the words were grouped into branches in the prosodic tree structure produced by the tool.

The syntactic segmentation of the English data was performed by 20 informants, all of them native Finnish speakers having good to excellent skills in English language and grammar. They were asked, demonstrated by some examples, to tag every syntactic sentence and clause boundary in the text, and in ambiguous cases, make their own interpretations according to the context. The informants had no access to the spoken data but only the transcription, so they had to perform the task without help of any acoustic cues. Each sample was segmented by four informants independently of each other. The Finnish corpus was syntactically segmented according to the same principles, but due to its smaller size, the segmentation was performed by only one skilled native speaker of Finnish.

The wavelet based prosodic segmentation resulted in 1700 prosodic boundaries in the English data and 703 in the Finnish data. The syntactic segmentation resulted in 2147 and 457 syntactic boundaries respectively. The mean length of a prosodic unit was thus 5.9 words in the English data and 3.2 words in the Finnish data, and for syntactic units, 4.7 and 4.9 words respectively. The remarkable difference in the length of prosodic units may be due to the fact that the Finnish speakers were rather aged, therefore speaking more slowly and taking more pauses that were interpreted as prosodic breaks.

Of all the boundaries marked in the English data, 906 were co-occurrences of prosodic and syntactic ones. This is 53.3% of all the prosodic boundaries and 42.2% of the syntactic ones. In the Finnish data, the number of co-occurrences was 313, the percentages being 44.5% and 68.5%.

It is thus clear that prosodic and syntactic boundaries tend to co-occur, since in both languages their co-occurrences are significantly more common than a random distribution. It is interesting, though, to take a look at the exceptions: prosodic boundaries without a syntactic one or vice versa.

For a syntactic boundary without a prosodic one, the most common instance in both English and Finnish data was a new syntactic main clause, either independent or coordinate, beginning in the middle of a prosodic unit. Almost 60% of solitary syntactic boundaries in the English data and a little less than 30% in the Finnish data were such occasions. They included a lot of half-grammaticalized syntactic elements like *I think* or *you know* produced as a part of a larger structure without need to separate them prosodically. A little less common were subordinate clauses beginning in the middle of a prosodic unit, with the percentages of ca. 30% and 20%.

- (1) a solitary syntactic boundary after *y(ou) know* (P = prosodic boundary, S = syntactic boundary)
PS but it's n P ah PS yknow S it's not a backwater either PS

The most significant difference between English and Finnish data was related to cases where a solitary syntactic boundary preceded a conjunction, either a coordinate or a subordinate one, and a prosodic boundary followed the same conjunction. These cases corresponded to 42% of all the solitary syntactic boundaries in the Finnish data but only 5% in the English data; in English, it was much more common to have a prosodic break both before and after a conjunction. In spontaneous speech, a conjunction at the end of an utterance or a speaker turn is a common phenomenon since it often shows an intention to continue despite the prosodic break.

Prosodic boundaries in both English and Finnish were only seldom situated between a subject and a predicate, or between an auxiliary and a main verb. Only ca. 15% of solitary prosodic boundaries in both English and Finnish data were situated in either one of these locations. On the contrary, more than 50% of these boundaries in Finnish and almost 50% in English were situated either between a verb and its arguments other than subject, or between more remote constituents.

- (2) a solitary prosodic boundary between a verb and its object (P = prosodic boundary, S = syntactic boundary)
PS oh they well they knew P all d all different things PS

It was also not uncommon to have a prosodic break in the middle of a noun phrase, especially in cases where an NP consisted of several words and the situation included some kind of hesitation.

Our results thus show that in both spontaneous English and Finnish speech, prosodic and syntactic boundaries typically tend to co-occur. Even where they do not, prosodic boundaries only seldom break nuclear elements of a syntactic clause; rather they are situated on the peripheral areas of a clause. The fact that prosodic boundaries do not tend to break fixed syntactic elements corroborates the assumption that prosody and syntax serve a common purpose in structuring spoken language.

References:

- [1] Suni, A., Šimko, J., Aalto, D., & Vainio, M. (2017). Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech & Language*, 45, 123-136.
- [2] Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech*. 2nd release. Columbus, OH: Department of Psychology, Ohio State University (Distributor), www.buckeyecorpus.osu.edu.
- [3] *Samples of spoken Finnish: an online speech corpus* (2014). Version 1.0. Helsinki: Institute for the Languages of Finland. <http://urn.fi/urn:nbn:fi:lb-1001100134>.