AN EMPIRICAL APPROACH FOR COMPARING SYNTAX AND PROSODY DRIVEN PROMINENCE MARKING

György Szaszák[1] and András Beke[2]
[1]Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics,
[2]Research Institute for Linguistics, Hungarian Academy of Sciences
[1]szaszak@tmit.bme.hu, [2]beke.andras@mta.nytud.hu

Abstract

In theory, the primary role of prosody is often described as reflecting and/or disambiguating syntax in speech, and eventually signaling semantic, pragmatic or other paraverbal content. Speech understanding or speech-to-speech translation applications often require the identification of key or emphasized elements in speech. One of the basic interests of this paper is whether these elements can be recovered by syntax, prosody or both? A statistical analysis is performed on a large representative Hungarian corpus to compare syntax- and prosody-based prominence marking approaches. Additionally, the signaling of focus by basic text and prosody is also evaluated. Results show that 50% more prominence was revealed by prosody over syntax. The overlap between text- (syntax) and prosody-based prominence marking is around only 30%. This finding suggests that prosody provides information about word boundaries and prominence in conjunction with syntax. We conclude that robust prominence detection needs both prosodic and syntactic analysis.
Keywords: prosody, syntax, prominence, focus

1 Introduction

The idea that speech prosody represents the underlying syntactic structure of an utterance is utilized in several speech technology applications, mainly in speech synthesis, where prosody prediction is usually based on models capturing syntactic regularities or characteristics either in a rule-based or in a data-driven manner (van Santen, 1997; Romportl, 2006). The input of such prosody generation models usually consists of the pure textual representation, eventually accompanied by syntactic-analytic information, such as Part-of-Speech (POS) tagging, syntactic phrasing etc. These models yield acceptable Text-to-Speech (TTS) quality, although their limits are known (Obin, *et al.,* 2010). However the fact that predicting prominence based on text is feasible suggests that syntax and prosody are closely linked. On the other hand, when moving toward spontaneous speech or expressing attitudes or emotions, the limits of these text-based approaches become more evident (Jiang *et al.,* 2005).

On the other hand, when prosody analysis is required to either support syntactic analysis (Szaszák & Beke, 2012) or at least to identify prominent syntactic elements (words or syntactic phrases) directly from speech (Ananthakrishnan & Narayanan, 2007), the syntax-phonology interface seems to be exploitable to some extent. However, when mapping prosody to syntax, ambiguity is much more problematic than in the case of syntax to prosody

Szaszák, György & András Beke: An empirical approach for comparing syntax and prosody driven prominence marking. *The Phonetician,* 114. 2017. 46-57.

mapping. Indeed, for a given syntactic representation, several grammatically correct prosodic phrasing and accentuation patterns may exist. It depends on the speaker's semantic and pragmatic intentions which of these patterns gets realized during the speech production process. Whereas by mapping syntax to prosody, we can choose a canonical form of prosodic representation and we can use the text as input. In the reverse direction, we can rely only on acoustics or eventually on an Automatic Speech Recognition (ASR) output, containing more or fewer errors.

There is a wide range of different theoretical interpretations that attempt to explain the role of prosody. The most common (and oldest) one is that prosodic structure results from syntactic constraints and hence syntax maps to prosody fairly well (Truckenbrodt, 1999). Another approach supposes an auxiliary role of prosody in reflecting syntax: when pure syntax is ambiguous, prosody resolves this ambiguity (Allbritton *et al.,* 1996; Snedeker & Truswell, 2003); in the lack of ambiguity, prosodic markedness may be omitted. The third group of theories have a more complex view about prosody: along with syntax, other linguistic and para-linguistic factors influence the prosodic structure (Watson & Gibson, 2004). Language dependency is also a crucial factor regarding the functions of prosody in speech, i.e. some languages (such as English) need more intervention from prosody to express emphasis. Whereas in other languages (typically, morphologically rich languages where word order can vary, such as in Hungarian), syntax is more independent and has a rich repertoire to convey the information structure itself, allowing prosody to adopt functions not driven by syntax.

Our basic interest in this paper is to compare the above modelling paradigms in an empirical way. Prosodic phrase prediction based on text (syntax) and prosodic phrasing recovered from the pure acoustic speech signal without any kind of semantic or pragmatic interpretation are compared. We use a practical approach as we focus on future exploitation in automatic speech recognition or understanding applications: the cue we are looking for is some kind of acoustic markedness, which can be thought of a prominent prosodic event. These acoustic markers define the prosodic phrases in our current interpretation (each prominent segment is regarded to be an individual phonological phrase). With respect to grammatical functions, acoustic prominence can reflect word prominence, sentence prominence or the marking of semantic focus, *etc.* As in their acoustic manifestations, these different functions are very similar and are often treated together in speech technology (TTS or prosodic phrase detection from the acoustic speech signal).

An interesting and actual exploitation of prosodic phrasing is necessary in machine-based speech-to-speech translation. Thereby, the transfer of prosody allows for an adequate translation in terms of synthesizing the original prosody, which allows for taking into account deviations from canonical prosody in the source language and makes it possible to transfer these into the target language on the corresponding words. As speaker adaptation in TTS technology is already available, the transfer of prosody can highly contribute to a multi-modal, more complete translation, capable of transferring emotions, atypical patterns or contrastivity – that is semantic and pragmatic information that is largely unavailable from the text (syntax).

Inspired probably by listening to read or standard formal speech, our naive expectation could be that where syntactic grouping requires, an acoustic prominence is necessarily produced in

prosody. However, as mappings between syntax and prosody are far from unambiguous, and as a one-to-one like syntax-prosody mapping would provide extensive and often useless redundancy, we may suspect that prosody sometimes really reflects syntax, and sometimes syntax unambiguous enough to let prosodic markedness go. On the other hand, as syntax itself has limited capabilities in expressing semantic and pragmatic functions alone, prosody can take such functions. In our research we try to quantify these alternatives.

Our experiments will be based on a read corpus, designed for a TTS inventory, consisting of formal style utterances. We use the written form to derive syntactic phrasing and an assignment of prominence, as syntax is the same in spoken and written language. This prominence annotation was provided by human experts, who are proficient native users of Hungarian. The acoustic component was analyzed by an automatic phonological phrase alignment and prosodic prominence detection application, which, based purely on speech acoustics, detects prosodic phrase boundaries with high precision and resists micro-prosodic alternations. We are mainly interested in noting the overlap of these two approaches. In addition, a third alternative is also analysed: Hungarian as a topic-prominent language (Li & Thompson, 1976) (in contrast to subject-prominence in English), Hungarian syntax is built around the information structure which is believed (at least by theory) to be also reflected by prosody. Unlike the two previous approaches, the third one operates at the intonational phrase level (and not at the phonological phrase level).

This paper is organized as follows: first materials and the prominence marking methods to be compared are presented. We then compare syntax- and acoustic-driven prominence alternatives hypothesizing significant differences in prominence patterns of the compared utterances. Then automatic focus marking is investigated compared to the other two methods. We hypothesize a high correlation between focus and both syntax- and acoustic-based prominence marking systems. Speaker dependency on the acoustic prominence marking method is also analysed hypothesizing significant differences among speakers. Next, the correlation between Part-of-Speech (POS) and prominence marking is investigated. Finally, conclusions are drawn.

2 Methods

The test corpus was the parallel, Precisely Labelled Hungarian corpus (Olaszy, 2013), which with 12 times 1948 sentences containing 28,944 words read by 12 native speakers of Hungarian (6 males and 6 females, between the ages 26 and 60 years). One of the female speakers was a professional broadcaster. The other speakers were non-professionals, all living in Budapest with various professional backgrounds. The corpus was designed to be phonetically rich and consisted of sentence (almost exclusively declarative sentences and questions). Both single and compound sentences occurred in the corpus. The corpus was initially designed primarily for TTS purposes. This means that the corpus contains standard, formal style, neutral pronunciations. The size of the corpus makes it possible to derive general conclusions for read speech.

The prominence labelling alternatives are presented in the next section. Their common point is that prominence is assigned at the word level. Hence, comparisons will be made by aligning prominence patterns for a given word sequence obtained with each of the methods. Statistics will be used to test for significant differences.

3 Prominence labelling alternatives

This section presents the two prominence marking approaches that will be compared. Focus marking is also added.

3.1 Text-based prominence marking

This is the simplest labelling scheme. Native Hungarian annotators were asked to read written sentences and mark the words they would expect to have any kind of notable acoustic prominence if the sentence were read aloud. The expert was given a response sheet with all 1948 sentences used in the experiment and he was instructed to mark the words he thought should be stressed. For each sentence, he was allowed to mark as many words as he wanted. As Hungarian is a fixed stress language (stress is assigned to the first syllable of a word), a pure marking of the word was expected to carry sufficient prominence. There was no distinction regarding the strength of prominence, but a simple binary decision was made (prominent or not prominent). This labelling was carried out by a single male annotator for the whole corpus, accompanied by two others (1 male, 1 female) on a smaller subset (on 10% of randomly chosen data) to evaluate inter-annotator agreement. Agreement was determined to be 86.2%.

3.2 Prosody-based prominence marking

Prosody based prominence marking is carried out by automatic phonological phrase (PP) alignment and prosodic prominence detection application. The latter operates purely on the acoustic speech signal (relying mainly on F_0 and overall energy as acoustic features) and did not have any linguistic input. A complete description of this system is available in (Szaszák & Beke, 2012). This procedure resulted in a dataset, which was suitable for the analysis of the research hypotheses. The position and type of the stressed words were automatically inserted into a sheet representing utterances from all speakers.

The utterances are fed into the machine which segmented the sentences into phonological phrases. Given first syllable fixed prominence in Hungarian, respective PPs revealed prominence and hence acoustic prominence at their left edge. The precision (PRC) and recall (RCL) of this system was evaluated on a separate set of sentences using a 100 ms detection tolerance interval and is shown in Fig. 1 (Szaszák & Beke, 2012). In these experiments, the system was used with a 200 ms tolerance interval. For simplicity, we evaluated the system in the operating point where recall and precision were equal (81.2%). These numbers are close to the inter-annotator rates reported in other studies of perceptual labelling of prominence (Wightman *et al.,* 2000). There was no upper limit for the number of words stressed within a sentence. The lower limit was one word stressed per sentence.

The training phase of the PP models used in the automatic prominence detection tool may implicitly encode some syntactic knowledge (PPs are annotated in its training corpus such that they coincide with word-boundaries). On the other hand, the prosody-based prominence marking system only has the acoustic speech signal available for processing.

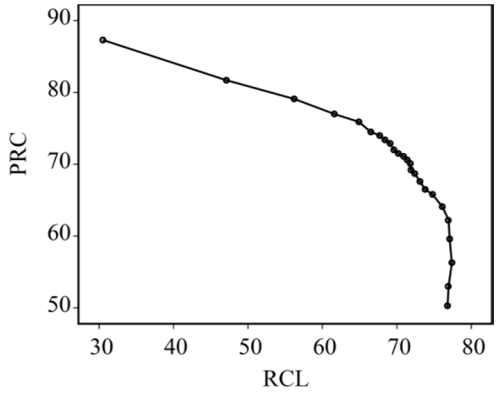

Figure 1. Operation curve of the PP alignment system in the precision and recall space by 100 ms tolerance.

3.3 Focus marking

Focus marking is run automatically and is based on a deep syntactic analysis performed using only textual features within "magyarlánc" tool (Zsibrita *et al.,* 2013). Unlike the two previous approaches, focus is interpreted at the intonational phrase (IP) level, and hence will not recover all prominent words, but only the most prominent among them for each IP (IPs can be thought of as clauses in this carefully read, formal speech corpus). This system was trained on BEA (the Hungarian Spontaneous Database, Neuberger *et al.,* 2014).The detailed description of the focus detection algorithm can be found in (Beke & Szaszák, 2012).

The decision for incorporating focus-based prominence marking as an alternative was motivated by the peculiarities of the Hungarian language and the relevance of focus-based prominence to the syntax/prosody interface. The focus of a clause is a constituent containing the main prominence of the intonational phrase (Szendrői, 2003). Hungarian is a morphologically rich language where cases are frequently reflected by suffixes unlike English, where the positions within the sentence identify cases (*i.e.,* the order of the subject, the object and the verb are fixed within the sentence). As case endings already function to reflect grammatical roles, word order is less constrained in Hungarian. Word order takes another function, which results in a different focus marking strategy for Hungarian: In English, an acoustically present prosodic prominence is needed to assign the focus to a given word in the bound word sequence, whereas in Hungarian, word order has the capability of identifying the focus (preverbal position), even without explicit acoustic marking.

It is a common impression of Hungarian speakers and an expectation described by several Hungarian linguists (Kálmán & Nádasdy, 1994) that focus is also marked acoustically by prosody in parallel with word order. However, as focus position is clearly marked by syntax in most cases, prosodic marking of the focus is a redundancy, which theoretically could be omitted. Since prominence is perceived as focus by most listeners, it is possible that this results from interpretation and meaning and not from acoustics. Therefore, one of our goals in this paper is to explore to what extent prosodic prominence accompanies syntactic focus marking in Hungarian.

4 Results and Discussion

As Hungarian is fixed stressed on the first syllable, we conducted the analysis at the word level. Eventual secondary or rhythmic stress within the word or in compound words realized on other syllables than the first one is out of the scope of the current study.

4.1 Similarity between the different approaches

The first research question assessed the ratio of prominent words with the three tested alternatives, that is, the text-based (TB), the prosody-based (PB) and the focus detection-based (FB) prominence marking approaches. Results are shown in Table 1.

Table 1. Ratio of prominent words from all words [%].

Prosody (PB)	Text (TB)	Focus (FB)
48.4	33.1	13.7

The highest prominence marking ratio was produced by the PB method, which identified almost every second word as prominent. The TB method had a weaker ratio, around 1/3. The difference in prominence marking frequency between these two methods is statistically significant (McNemar test: $\chi^2 = 10.119$; $p < 0.001$).

The fact that prosody revealed more prominent words could be an argument for its predominant role in marking syntactic prominence in speech. This finding can still be linked to a syntactic function, the signalling provided by word or rather clitic group boundaries (Vicsi & Szaszák, 2010).

The FB approach was expected to identify prominence at the IP level (focus). Its hit rate, between 1/7 and 1/8, is exactly in line with average sentence length measured in the corpus: on average, a sentence contains 7 words (±2 words standard deviance). Fig. 2 shows the overall distribution of prominent words by keeping as its basis only the words marked as prominent by any of the three methods.

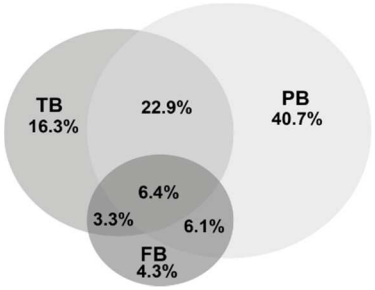

Figure 2. Overall distribution of prominent words.

4.2 Overlap among the prominence marking approaches

Given that Fig. 2 shows less overlaps than expected, it is interesting to carefully evaluate the overlap between the different approaches that is the "inter-annotation ratio. For this pairwise evaluation (PB *vs.* TB, TB *vs.* FB and PB *vs.* FB), words marked as prominent by at least one of the methods currently compared are kept as the basis, the remaining words

(not marked as prominent by any of the methods) were removed. The overlap between the different approaches is shown in Fig. 3.

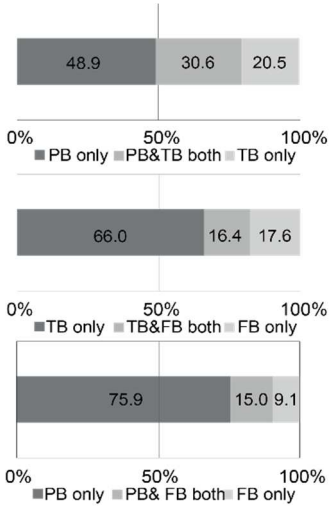

Figure 3. Pairwise overlaps between prominence marking methods. The basis consists of the words marked as prominent by any of the two methods actually compared.

Table 2 shows overlaps in prominence marking normalized for each method. This process assumed a selected method marked a word as prominent and then determined what percentage of these words were also marked by another method.

Table 2. Normalized overlapping in prominence marking between methods [%]. Pairwise comparison.
Basis: 100%=all words marked as prominent by a selected "primary" method (row-wise). Column-wise: given the primary method of marking prominence, what is the ratio of prominence marking considering the second method (*i.e,*. this is like a conditional relative frequency).

Primary prominence marked by	**Text (TB)**	**Prosody (PB)**	**Focus (FB)**
Text (TB)	100.0	59.9	19.9
Prosody (PB)	38.5	100.0	16.5
Focus (FB)	51.8	62.2	100.0

Results of the comparison between the TB and the PB methods at the first glance reflect the ambiguity seen in theory; *i.e.,* whether prosody reflects syntax or is it more autonomous. The fact that if any of these two methods marks the prominence, the other one has some tendency to co-mark it is clearly identifiable. However, the overlap between these two methods remains still relatively weak: only 30.6% of the words marked by one of the methods are marked simultaneously by the other one. This is obviously not sufficient if we hypothesize that prosody should consequently reflect the written form (*i.e.,* including spaces between words) and hence the TB prominence markers. Approaching results from this

aspect, 40.1% of the words marked by TB are missed by PB. This finding suggests that if prominence is syntactically obvious, its prosodic marking can be missed by the speakers, even in carefully planned read speech.

Summarizing the results and supposing that syntax is the main governing factor in TP prominence marking, we may observe that

(i) syntax often (20.4%) marks prominence alone without the intervention of prosody;
(ii) prosody often marks prominence, which is unavailable in human syntax motivated prominence annotations, although the eventual role of prosody in signalling word boundaries is an alternative (and likely see Vicsi & Szaszák, 2010; Szaszák & Beke, 2012) hypothesis here. However, the fact that prosody accounts for almost half of the prominence markings alone and that automatic word boundary recovery based on prosody could not reach more than 78% precision (Szaszák & Beke, 2012) justifies that prosody has a function in human speech which is beyond syntax.
(iii) often (up to 30%) both syntax and prosody mark a prominence. This does not necessarily justify an explanation like "prosody is there to reflect syntax and hence the information structure", but may also mean that prosody further strengthens an already syntactically signalled emphasis.

Although results cannot prove or disprove any prominence theory, they confirm the hypothesis that prominence predictions based on acoustics and syntax are significantly different. This is an important outcome considering automatic speech recognition, speech-to-speech translation or speech understanding applications, where it seems to be crucial to include both acoustic and textual analysis if content analysis or extraction of the meaning is targeted. For example, in speech-to-speech translation, a pure text-based regeneration of the prominence on the target language side would result in losing all prominence marked only by prosody and hence the information conveyed in this manner. These findings favour a theoretical approach of prominence, whereby prosody is not only an auxiliary player subordinated to syntax, but rather has an individual role in further structuring of the information and providing cues for perception orientation and for the decoding of paraverbal information. The role of prosody may also include some "rescoring" or re-weighting of the syntactically marked emphasis or information structure in general. Indeed, this is the difference which makes human speech richer than text alone with the same word sequence. These findings primarily concern languages where word order is less constrained and hence theoretically prosody is not necessary to mark emphasis since it is already marked by syntax (*i.e.,* word order).

Regarding focus, TB and PB methods marked the focus position as prominent by 51.8% and 62.2%, respectively. Combined together (*i.e.,* when the TB and the PB methods are used together), focus recall augments to 78.5%. (If we disregard focus marking, 64.8% of the words are marked as prominent by TB or PB methods together.) This means that focus marking by prominence can be observed almost 4 times from 5 cases (78.5%). However, as focus is a syntactically identifiable position, the recall value (overlap in prominence marking) yielded by the TB method, believed to be primarily syntax driven, is still regarded to be weak. This means that our original hypothesis was disproved. Probably a comparison

to a syntactic analysis prior to the manual TB prominence marking (which is based on human judgement without explicit syntactic analysis) could help to obtain a clearer overview in this regard, however, for the present study, this kind of prominence marking is unavailable. Focus marking by prosody has already been investigated for Hungarian (Markó, 2005; Mády, 2012). In (Mády, 2012) the authors found that focus marking by prosody cannot be confirmed as a general phenomenon in Hungarian. The results obtained in the present study using a larger and representative corpus are consistent with hers (Mády, 2012).

4.3 Speaker dependency

The TB and FB methods using textual input are relatively speaker independent, but theoretically, prosody and hence the PB method may show some speaker dependency. This is investigated in this subsection. We use the Jaccard similarity measured pairwise between the PB and TB and the PB and FB methods. Results showed minimal fluctuation in similarities between the methods depending on the speaker (Fig. 4). This means that in carefully read speech, prominence producing strategies are relatively speaker independent, so our hypothesis was not confirmed. We believe this is due to the nature of the corpus we used, as variation in prosody has been observed in other studies (Chappel & Hansen, 1998). So we have to emphasize that we found this for a TTS corpus, which contains a rather strict formal style of speech and the utterances are individual, without context, which limits the number of possible interpretations of meaning which may limit fluctuations in prominence across speakers. In addition, the data recorded from the professional broadcaster does not show differences in correlation compared to the other non-professional speakers.

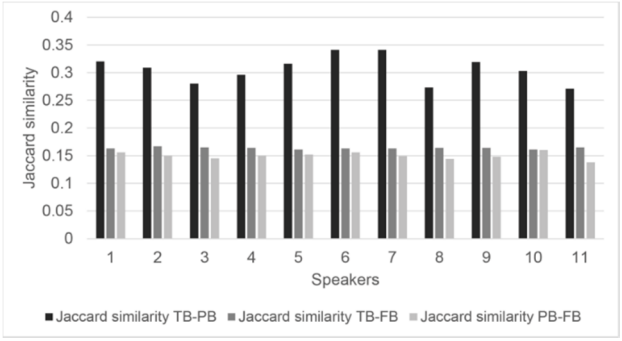

Figure 4. Jaccard similarity measured pairwise, for different speakers.

4.4 Prominence distribution on the utterance level

Regarding the position of prominent words within the sentence, Fig. 5 shows the distributions with the different methods. The shape of the curves is quite similar, showing a preference for the first half of the sentences. The peak on the second word seen by the TB method is unique.

Additional results show that words marked prominent by all methods were sentence initial in 37% of all cases (basis: words co-marked by all the three methods) and they are among the first 4 words of the sentence in 73% of all cases. Indeed, prominence marked on words positioned in the second half of the sentence occurred typically for complex or compound sentences consisting of at least two clauses (that is, in the second or last clause). To sum up,

there is no major difference among the methods, all of them favour the left edge of utterances in Hungarian.

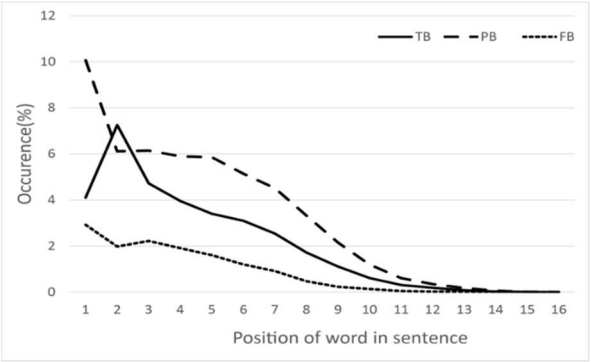

Figure 5. Distribution of prominent words within the sentences.

4.5 Part of speech versus prominence

We also investigated eventual overlaps between prominence marking and part of speech tags. These results are presented in Table 3. It seems to be method independent, that most often nouns and verbs are prominent. In the TB prominence marking scheme, adjectives account for more prominence markings than with the two other methods. The FB method disfavoured adverbs. This is understandable, as these words rarely fill the focus position, however, if they occur, they usually add important information (when, where, *etc.*). Analysing these cases more closely showed that text-based human annotation often moved prominence to the noun premodifiers, if present. However, despite this difference, pairwise correlations between the prominence marking methods are quite strong in terms of part of speech and type of the prominent element: $r = 0.968$, $p < 0.001$ for TB *vs.* PB; $r = 0.946$, $p < 0.001$ for TB vs FB; and $r = 0.981$, $p < 0.001$ for PB vs FB.

Table3. Distribution [%] of POS for the words marked as prominent, separated by methods.

POS	Text (TB)	Prosody (PB)	Focus (FB)
Noun	32.0	34.9	36.0
Verb	17.7	17.8	19.2
Adjective	19.5	12.9	11.7
Adverb	13.6	13.7	7.8
Pronoun	8.2	7.6	5.9
Numeral	3.6	2.2	1.7
Conjunction	2.3	6.0	4.0
Article	2.6	3.4	1.1
Postposition	0.1	1.2	2.5
Other	0.4	5.1	10.1

5 Conclusions

This paper investigated prominence in Hungarian, especially the interconnection between syntax and prosody in prominence marking, in a large corpus of read speech. Results confirmed some correlation between syntax and prosody, but in cases where both the text- and the prosody-based methods marked prominence simultaneously, only 1/3 of the prominent words were identified. Based on this, the role of prosody as a pure reflector of the syntactic structure can be discarded. The fact that almost half of the prominence is marked only by prosody suggests its function beyond syntax (even if partly parallel to the signaling of word boundaries). Prominence marking in the different methods did not show speaker dependency. The distributions of the POS tags of words marked as prominent by the different methods did not show major differences between methods either. Overlap between text or prosody and focus-based prominence marking was weak, which suggests a limited applicability of the focus theory in current automatic speech understanding applications. However, we can conclude that speech technology applications, such as content analysis, speech-to-speech translation *etc.,* could benefit from both syntactic and prosodic analysis.

Acknowledgements

The authors would like to thank the support of the Hungarian National Research, Development and Innovation Office (NKFIH) under contracts PD-112598, Automatic Phonological Phrase and Prosodic Event Detection for the Extraction of Syntactic and Semantic/Pragmatic Information from Speech; and K-108762, Annotation and development of BEA Hungarian spoken language database.

References

Allbritton, D., McKoon, G. & Ratcliff, R. 1996. Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 714-735.

Ananthakrishnan, S. & Narayanan, S. 2007. Improved speech recognition using acoustic and lexical correlates of pitch accent in a n-best rescoring framework. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2007.* (Vol. 4.) 873-876.

Beke, A. & Szaszák, Gy. 2012. Combining NLP techniques and acoustic analysis for semantic focus detection in speech. In: *Proceedings of the 5th IEEE International Conference on Cognitive Infocommunications.* 493-497.

Chappell, D. T. & Hansen, J. H. L. 1998. Speaker-specific pitch contour modelling and modification. Acoustics, Speech and Signal Processing. *In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 2). 885-888.

Jiang, D., Zhang, W., Shen, L. & Cai, L. (2005) Prosody Analysis and Modeling for Emotional Speech Synthesis. In: *Proceedings of ICASSP 2005.* 281-284

Kálmán, L. & Nádasdy, A. 1994. A hangsúly [the prominence]. In: *Strukturális magyar nyelvtan [Structural Hungarian Grammar].* Budapest: Akadémiai Kiadó. 393-467

Li, C. & Thompson, S. 1976. *Subject and topic: A new typology of language.* New York: Academic Press. 465-489.

Mády, K. 2012. Prosodic marking of focus in read and spontaneous speech. In: *Beszéd, adatbázis, kutatások. [Speech, data base, research].* Budapest: Akadémiai Kiadó. 91-107.

Markó. A. 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége. [On Some Prosodic Features of Spontaneous Speech].* PhD thesis, Eötvös Loránd Tudományegyetem, Budapest.

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M. & Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. In *Proceedings of International Conference on Text, Speech, and Dialogue.* Cham: Springer.

Obin, N., Lanchantin, P., Avanzi, M., Lacheret-Dujour, A. & Rodet, X. 2010. Towards improved HMM-based speech synthesis using high-level syntactical features. In: *Proceedings of Speech Prosody 2010.*

Olaszy, G. 2013. Precíciós, párhuzamos magyar beszédadatbázis fejlesztése és szolgáltatásai [Development and services of a Hungarian precisely labelled and segmented, parallel speech database]. *Beszédkutatás 2013 [Speech Research],* 261-270.

Romportl, J. 2006. Structural data-driven prosody model for TTS synthesis. In: *Proceedings of Speech Prosody 2006.* (Vol. 4). 549-552.

van Santen, J. 1997. Prosodic modelling in text-to-speech synthesis. In: *Proceedings of EuroSpeech1997.* KN19-28.

Snedeker, J. & Truswell, J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48(1), 103–130.

Szaszák, G. & Beke, A. 2012. Exploiting prosody for automatic syntactic phrase boundary detection in speech. Journal of Language Modeling 0(1), 143-172.

Szendrői, K. 2003. A prominence-based approach to the syntax of Hungarian focus. *e-Linguistic Review*, 20. 37-38.

Truckenbrodt, H. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry,* 30, 219-255.

Vicsi, K. & Szaszak, G. 2010. Using prosody to improve automatic speech recognition. *Speech Communication* 52(5), 413-426.

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

Wightman, C., Syrdal, A., Stemmer, G., Conkie, A. & Beutnagel, M. 2000. Perceptually based automatic prosody labeling and prosodically enriched unit selection improve concatenative speech synthesis. In: *Proceedings of International Conference on Spoken Language Processing.* (Vol. 2.), 71-74.

Zsibrita, J., Vincze, V. & Farkas, R. 2013. magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: *Proceedings of RANLP.* 763-771.