Theses of Doctoral (PhD) Dissertation

Automatic extraction of prosodic features and their employment in the analysis of speech corpora

István Szekrényes

Supervisor: Dr. Hunyadi László

UNIVERSITY OF DEBRECEN

Doctoral Shool of Linguistics

Debrecen, 2019

1 Introduction

Certain properties of prosody – sometimes referred to as the "musical means of the expression of speech" and "non-linguistic elements of communication" can, in their pejorative sense, easily make the impression that, as compared to the more abstract linguistic means of information they represent additional, esthetic–rhetoric elements of a peripheral area of linguistics. When focusing on the primary, constative, logical content of the linguistic sign, the musical organization of speech can be ignored as just a formal addition, its "coloring". But, due to some internal and external factors in linguistics the above assumed status of prosody has by now greatly changed.

One of the important "internal" factors is Speech Act Theory (Austin, 1962), the birth of a discipline representing the performative turn in pragmatics, within which the utterance representing the fundamental units of communication is no longer just the carrier of the propositional content, but as an acts with illocutionary force gets into the focus of research. In the identification and typology of illocutionary forces, in addition to other, especially contextual information the role and significance of prosody also gains weight. It does not mean that prosody does not play an important role in determining the conventional meaning of linguistic structures in formal semantics. We come across with several examples where structural ambiguity (the scope of logical operators, the positions of focus and topic) can only be disambiguated by considering the distinctive role of the suprasegmentals. Even though this module is secondary to that of syntax in generative models of language, there are approaches that adhere to the independence of the prosodic structure and its role in the proper interpretation of logical form (Jackendoff, 2002; Hunyadi, 2002). The significance of this approach is that prosody cannot only be taken as being the pragmatic aspect of language use, but it also becomes an inherent part of grammar contributing to the formation of the surface structure. It is then not surprising that in formal accounts of the phonology of Hungarian, stress/accent and intonation have received due importance (Kornai & Kálmán, 1988; Varga, 1994).

As "external" factors one can consider the emergence and spread of those new technologies and methodologies that make it straightforward to measure and analyze by computational means the acoustic structure of speech and its suprasegmentals (fundamental frequency and intensity). As a result, the observations and hyphotheses of linguists about prosody, mostly supported by introspection became scientifically verifiable and extend-

able, and more easily accessible for interdisciplinary research, and applicable in language technology. There is a marked tendency in the development of modern communication technologies for creating more effective procedures – the application of linguistic and linguistics-related interdisciplinary basic research in the joined work of scientists in the fields of communication theory, information technology, linguistics and psychology.

The assumption that prosodic information can be usefully applied in tasks regarding speech segmentation and analysis (such as speech and emotion recognition) has been proved several times (Shriberg et al., 2000; Szaszák, 2010). One of the important, perspective and goal-oriented aim of the related research is to increase the effectiveness and naturalness of human-machine communication. This can make, as an example, the treatment of turn taking more successful if the suprasegmental properties (speech melody, intensity, tempo etc.) are also considered in the interpretation of the utterances of the user, or in the generation of their responses (Hirschberg, 2002).

There are studies of the prosody-pragmatics interface in a so far less researched but prospectively valuable field, which are aimed at finding correlation between the thematic and prosodic structure of spoken social interactions (Nakajima & Allen, 1997; Zellers & Post, 2009). Nakajima & Allen (1997) studied the possible accompanying prosodic features of the thematic structure pointing out a correlation between the fundamental frequencies of utterances at their beginning and end, on the one hand, and their propositional relation to the adjacent utterances (introduction of new information, introduction of new information within a given topic, elaboration of a previous utterance, continuation of a topic). At the beginning of utterances introducing a new topic the fundamental frequency was found higher and it changed within a wider range, as compared to similar measurements with utterances continuing the same topic. Zellers & Post (2009) included further prosodic features and established that in order to discover correlations with the thematic structure one needs to simultaneously, rather than discretely study the given features. According to their view, even doing so certain prosodic patterns will not have unambiguous propositional or pragmatic interpretation. In their case study, Baiat & Szekrényes (2012) observed that subjects can perceive the change of the topic of a conversation surprisingly well carried out in a language unfamiliar to them (Hungarian vs. Persian) – solely relying on the dynamic and prosodic properties of the dialogue.

2 The aim of the research

Keeping in line with previous studies described above, the aim of the present research is to contribute to the study of spoken language by offering new methodologies and experiments in which the prosodic properties of speech alone are viewed as an informative acoustic reflection of the process of communication both in perception and the analysis of communication. The experiments carried out within the frameworks of the present dissertation are based on the following intuitive hypothesis:

> Prosody in speech is organized not only at the level of the utterance, it does not only modulate its semantic interpretation, but, offering a "flexible", quasi musical environment for the interaction, it conveys independent, underspecified, non-lexical information, which has perceptional relevance in the interpretation of the interaction even if the concrete verbal content is, for some reason, unaccessible.

This underspecified information content expressed by prosody becomes especially interesting if it represents an adequate starting point for the reconstruction of the actual properties and process of the given social interactions; by which I do not, naturally, mean the denotation of the concrete conversation universe, rather, those global properties associated with the relation between and roles of actors of communication, and with the speech context, reflected in the various intonation patterns, the changes of the dynamic and the rhythm of turn taking, and those local events of the process of conversation (such as topic change, and further speech acts), which can be categorized in form of binary decisions (whether it is formal or informal, whether the topic has changed or not), without knowing the specific content.

As for the methodology of the dissertation, its most important goal was to elaborate a model of a simple algorithmic annotation that, integrated with the annotation scheme of the HuComTech project (Hunyadi et al., 2012) could extend the description of the non-verbal content of the corpus with the prosodic properties of spoken language. These kinds of information were relevant to the tasks of the research group with the aim to extend research of the nature of multimodal communication into the formal basic structure of interactions and the cooperation between the various modalities (Hunyadi, 2011). In designing the automatic prosodic annotation it was considered important to only rely on

the acoustic analysis of the sound recordings and the manually annotated structure of turn taking. It is not only motivated by the eventual fact that during corpus development the word or syllable level segmentation of the transcriptions was not yet carried out; in any case, with the aim to be theory neutral, I wanted to analyze the prosodic structure independently of any syntactic or phonological judgement that could only be reproduced by manual analysis and that would presuppose the dependency of the suprasegmentals on other linguistic levels. As another important requirement, I wished to assure that the output of the annotation should represent the physically measurable changes of prosody with perceptional relevance, in the form of contours labeled using discrete categories. Since, following these principles, the model of annotation was aimed to offer a descriptive, non language-specific, contour-based representation of prosody, and by doing so it gave me the opportunity to implement an algorithm applying a rule-based approach – instead of machine learning requiring a teaching material.

The aim of the experiments carried out within the dissertation, both those involving perception and those applying neuronal networks, was to test the main hypothesis of prosody having informativeness independently of conversation and verbal content. Accordingly, the primary goal of tests with neuronal networks was not to create a solution by using all available information and guaranteeing maximal effectiveness, but to test what results can be achieved by relying solely on the prosodic information, excluding the verbal semantic content and the visual information.

3 Material and methods

As the primary source of data for both research and testing I used the recordings of the formal and informal dialogues of the HuComTech multimodal corpus and their related annotations, based on interviews with 111 subjects, published for research purposes in The Language Archive and the eMeta-Share.

In the first perception experiments I followed (Baiat & Szekrényes, 2012) using the conversation scheme of the HuComTech corpus but having dialogues in Persian and some newer versions of these dialogues, where the speech material was re-synthesized, void of its segmental content, preserving only its prosodic structure. The use of Persian, a language that the subjects were unfamiliar with was justified by the aim to exclude the

verbal content. In the first experiment I wished to know what impressions the subjects would have about the given prosodic features of the interactions without knowing the language, relying solely on prosody. The second test was aimed at learning about the mental representation of the prosodic structure: to what extent the subjects could follow the changes of pitch, intensity and speech tempo.

In the design of a further test I applied László Varga's system for the phonological description of Hungarian intonation (Varga, 1994). The test aimed at measuring the distinctive power of the character tones of the system: whether it is possible to identify the attitudinal content of utterances isolated from the original context, solely based on intonation. I carried out the experiment under three conditions: using the original recordings of the utterances, using sinus sounds re-synthesized on the basis of fundamental frequency values, and, finally, using a stylized version of the previous two.

The development of an application for the automatic annotation of prosody, implemented in the script language of Praat (Boersma & Weenink, 2016) was mainly inspired by the method later called as Prosogram (Mertens, 2004), originally designed by Cristophe D'Alessandro and Piet Mertens for the psychoacoustic stylization of tonal contours (Hart, 1976), but regarding the categorization of the modulations of the fundamental frequency it also learned from the parameters of the Tilt model (Taylor, 2000). The essence of the previous procedure is that it extracts the actually perceivable modulations of the tone from the changes of the frequency values measurable only by the computer considering speech intensity and certain threshold values determined by musicology. The resulting representation contains the movements of fundamental frequency stylized at syllable level (the vectors of the original data reduced to its value pairs). In my algorithm I reached beyond the syllable level and attempted to more holistically match the tune onto a "tonal event", integrating eventually more than one syllable, where the resulting annotation labels represent the form of tonal movement (rising, falling, descending etc.) and the relative (starting and ending) pitch placed in the individual pitch range of the subject. For the calculation of the threshold values in the categorization of tonal forms I considered the pitch range of the subjects, their average pitch and pitch variation, but the results of the perception experiments also served as orientation.

I extended the applied methodology to a similar study of intensity and speech tempo. The algorithm was developed and tested on the sound recordings and transcriptions of the

HuComTech corpus, but certain corrections and improvements that resulted in the present flexibility of the program were made possible through the cooperation with the SegCor project, in the analysis of the FOLK corpus (Schmidt, 2016) that contained German recordings of 2-14 subjects, under a wide range of conditions. The output of the program is the Praat supported TextGrid, but it is also available in XML format. The visualization of the tone contours is done as SVG graphic, following the XML to XSLT transformation of the files. These visualizations were proved especially useful in the subjective evaluation of the output of the analyses, since based on our earlier research (Hunyadi et al., 2014, 2015) focusing on the connection between hearing and vision in perception we could expect that due to the simultaneous interpretation of the two modalities (as congruent vs. incongruent), the visual output of the automatic annotation can be more easily compared with the audible intonation. In the experiment I used some recordings received from the archive of the Department of Ethnology of the University of Debrecen.

The output of the prosodic annotation was used in two experiments using neuronal networks (Szekrényes & Kovács, 2017; Kovács & Szekrényes, 2019), the aim of which was to automatize the classification tasks similar to the experiments on perception. The goal of the first experiment was to separate the formal and informal dialogues of the HuComTech corpus that we performed on the basis of three groups of features: the sequences of turn taking, raw F0 data, and the labels of tonal contours as output of the prosodic annotation. The goal of the second experiment with neuronal networks was to localize the questions of interviews as unmotivated topic change. To teach the neuronal networks we again used the sequences of turn taking, and again, the numeric versions of the output of Prosotool, using speech tempo and intensity normalized by the individual subjects.

4 Research results

The first partial results related to the initial development of the automatic prosodic annotation were first presented at the 8th Conference of Hungarian Computation Linguistics (Szekrényes et al., 2011) as new features of the analysis of the HuComTech corpus. Later publications primarily focus on the theoretical and practical background of the development of the procedure for the automatic annotation of the speech melody (Szekrényes, 2014) and the details of its technical realization (Szekrényes, 2015). The current ver-

sion of the application was made available within the e-magyar project (Váradi et al., 2017) for specialists in the field as a module based on the output of speaker diarization (Kornai & Szekrényes, 2017). The algorithms for the analysis of changes of speech loudness and tempo are still under development. Scientific results based on the methodology of the prosodic annotation were published as three book chapters (Hunyadi et al., 2016a,b; Kovács & Szekrényes, 2019), two publications in conference proceedings (Kovács & Váradi, 2017; Szekrényes & Kovács, 2017) and a successfully defended doctoral dissertation (Abuczki, 2013).

On the interactive page, that I created for the subjective evaluation of the output of the algorithm for the analysis of fundamental frequency, 9 individuals have so far submitted their evaluation, comparing the visualization generated on the bases of the F0-values with their perception of the intonation of the utterances. According to these evaluations 82.9% of the automatic evaluation of the 25 utterances presented were compatible with the perceived sound. Further details of the methods and results are found in chapter 4.2.6. of the dissertation. The interactive page as a technology to visualize the stylized fundamental frequency contours can also be considered a separate methodological result. Although the relevant Praat scripts and XML templates are already available and can be downloaded from a github repository, it is among my future plans to make the generation of prosodic labels and the visualization of prosody available for testing on the web page of e-magyar, after uploading an arbitrary sound file and its textual transcription.

Regarding the experiments based on Persian recordings, due to the small amount of stimuli and problems with controlling some dependent variables, as well as the difficulty of the task, it is hard to present definitive, quantitative results. I could not either support or reject the hypothesis regarding the relevance of the underspecified information content of prosody, however, based on the data I find it substantiated to continue the tests under more carefully controlled conditions. Furthermore, regarding the tracking of the prosodic features of speech in real time, it can be considered an interesting result that, even though, due to their frequent changes to follow the features in real time proved to be extremely difficult in all conditions, the average number of signals (of changes of pitch levels) received from the subjects was the smallest when listening to the original versions of the recordings. This observation could suggest that, regardless of the effect of the foreign language, a higher level of processing might have started in the subjects

that hindered their concentration on the physical changes of the prosodic features. Accordingly, if they could isolate speech sounds and words in the recordings, the prosodic features could, instinctively, also be reflected upon as linguistic or communicative functions promoting the interpretation of the utterances and attitudes of the speakers which, as a result, reorganized the mental imagery about the physical reality and decreased their perceptual priority.

he experiments on the distinctive power of Varga's typological tonal contours (Varga, 1994) confirmed that the procedures used in the processing of the fundamental frequency contour (smoothing and stylization) do not cause a significant loss of the amount of information in the evaluation of the attitudinal content of the utterances. The subjects could identify the most suitable context for the given utterances as efficiently as in the case of stimuli generated based on the raw F0 values. In the identification of the "falling" and "descending" contours the relative pitch proved to be an important factor even in case of fewer patterns.

It was an unexpected but important result of the experiments with neuronal networks that the classification for the separation of the formal and informal dialogues of the HuComTech corpus was surprisingly effective based only on the sequences of turn taking. The classification of the test set containing 108 extracts of dialogues had a 81.5% accuracy. It further improved by 3.7% (meaning a 20% relative decrease of error rate), when the labels of the prosodic annotation (tonal contours and relative pitch levels) were added to the set of features. In contrast, the inclusion of the F0 values resulted in a 1.1% decrease of the results. The experiments aimed at the automatic detection of topic change ended with positive results regarding the usability of the prosodic features. These results are comparable with international research with similar goals and conditions (see chapter 6.2.).

5 The theses of the dissertation

Based on the current results of research done within the framework of the dissertation, the findings of can be summarized in the following theses:

1. The automatic labeling of the changes of prosody relevant to the perception and information content of speech with rule-based algorithms (stylization, analysis of

fundamental frequency relations) is not just significantly faster than machine learning and human annotation, but has similar accuracy.

2. We demonstrated that a descriptive, rule-based model of annotation without phonological analysis of contours is, for Hungarian data, suitable for the computational processing of prosodic changes relevant to perception. Since the methodology does not contain language-specific elements, I find it possible to apply it to other, even tonal languages as well.

3. The appropriate stylization of the fundamental frequency contour does not result in significant loss of information in the successful identification of the attitudinal content of the various tonal patterns.

4. The rule-based processing of the raw acoustic features and their representation in the form of categorical variables (tonal patterns) efficiently reduces the learning time of neural networks and does not result in such a loss of information that would deteriorate the success of the classification of spoken social interactions as formal or informal.

5. The rhythm of sequences of turn taking complete with pauses, overlapping speech and prosodic information can be effectively used in determining the global characteristics of the interaction regarding the relation and role of the partners as well as the context of the interaction.

6. In case of unmotivated turn takings the detection of topic change can be effectively implemented without the verbal content, relying solely on the prosodic structure, if, in the teaching of deep neuronal networks and classification, one does not only consider the utterances introducing the new topic (starting F0 value, pitch range), but also the intonation properties and patterns of their environment.

References

Abuczki, A. (2013). A mondjuk nem konceptuális használatának vizsgálata multimodális kontextusban. In T. Váradi (Ed.), *Alknyelvdok7* (pp. 303–316). Budapest: MTA Nyelvtudományi Intézet.

Austin, J. L. (1962). *How to do things with words.* Cambridge: Harvard University Press.

Baiat, G. E., & Szekrényes, I. (2012, Dec). Topic change detection based on prosodic cues in unimodal setting. In *2012 ieee 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (p. 527-530). doi: 10.1109/CogInfoCom.2012.6422037

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer [computer program]. version 6.0.22.* `http://www.praat.org/`. (retrieved 15 November 2016)

Hart, J. t. (1976). Psychoacoustic backgrounds of pitch contour stylisation. *IPO-APR*, *11*, 11-19.

Hirschberg, J. (2002). Communication and prosody: functional aspects of prosody. *Speech Communication*, *36*, 31–43.

Hunyadi, L. (2002). *Hungarian sentence prosody and universal grammar.* Frankfurt: Peter Lang Verlag.

Hunyadi, L. (2011). Multimodal human-computer interaction technologies. theoretical modeling and application in speech processing. *Arguemntum*, *7*, 240–260.

Hunyadi, L., Földesi, A., Szekrényes, I., Staudt, A., Kiss, H., Abuczki, A., & Bódog, A. (2012). Az ember–gép kommunikáció elméleti–technológiai modellje és nyelvtechnológiai vonatkozásai. In *Általános nyelvészeti tanulmányok XXIV: Nyelvtechnológiai kutatások* (pp. 265–309). Budapest: Akadémiai Kiadó.

Hunyadi, L., Kiss, H., & Szekrényes, I. (2016a, 01). Incompleteness and fragmentation: Possible formal cues to cognitive processes behind spoken utterances. In (p. 231-257). doi: 10.1007/978-3-319-21209-8_14

Hunyadi, L., Kiss, H., & Szekrényes, I. (2016b). Prosody enhances cognitive infocommunication: Materials from the hucomtech corpus. In A. Esposito & C. L. Jain (Eds.), *Toward robotic socially believable behaving systems - volume I: Modeling emotions* (pp. 183–204). Cham: Springer International Publishing.

Hunyadi, L., Szekrényes, I., Czap, L., & Sziklai, I. (2014). Seeing the sounds? *Argumentum*, *10*, 325–338.

Hunyadi, L., Szekrényes, I., & Sziklai, I. (2015). Vizuális percepció és nyelvi feldolgozás. *Beszédkutatás*, *23*, 186-208.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution.* Oxford University Press.

Kornai, A., & Kálmán, L. (1988). Hungarian sentence intonation. In H. Hulst van Der & N. Smith (Eds.), *Autosegmental studies on pitch accent* (pp. 183–195). Dordrecht: Foris Publications.

Kornai, A., & Szekrényes, I. (2017). e-magyar beszédarchívum. In A. Tanács & V. Vincze (Eds.), *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY2017)* (pp. 103–109). Szeged: JATEPress.

Kovács, G., & Szekrényes, I. (2019). Applying neural network techniques for topic change detection in the hucomtech corpus. In L. Hunyadi & I. Szekrényes (Eds.), *The temporal structure of multimodal communication: Theory, methods and applications* (pp. 147–162). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-22895-8_8` doi: 10.1007/978-3-030-22895-8_8

Kovács, G., & Váradi, T. (2017). A különböző modalitások hozzájárulásának vizsgálata a témairányítás eseteinek osztályozásához a hucomtech korpuszon. In A. Tanács & V. Vincze (Eds.), *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY2017)* (pp. 103–109). Szeged: JATEPress.

Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Proceedings of speech prosody.*

Nakajima, S., & Allen, J. F. (1997). Prosodic features of utterances in task-oriented dialogues. In *Computing PROSODY* (pp. 377–386). New York: Springer-Verlag.

Schmidt, T. (2016). Good practices in the compilation of folk, the research and teaching corpus of spoken german. In J. M. Kirk & G. Andersen (Eds.), *Compilation, transcription, markup and annotation of spoken corpora, special issue of the international journal of corpus linguistics [ijcl 21:3]* (pp. 396–418).

Shriberg, E., Stolcke, A., Hakkani-Tür, D., & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, *32*, 127–154.

Szaszák, G. (2010). A prozódia szerepe a beszédfelismerésben. In *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek* (pp. 390–392). Budapest: Akadémiai Kiadó.

Szekrényes, I., & Kovács, G. (2017). Classification of formal and informal dialogues based on turn-taking and intonation using deep neural networks. In A. Karpov, R. Potapova, & I. Mporas (Eds.), *Speech and computer* (pp. 233–243). Cham: Springer International Publishing.

Szekrényes, I. (2014). Annotation and interpretation of prosodic data in the hucomtech corpus for multimodal user interfaces. *Journal on Multimodal User Interfaces*, *8:(2)*, 143–150.

Szekrényes, I. (2015). Prosotool, a method for automatic annotation of fundamental frequency. In *6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 291–296). New York: IEEE.

Szekrényes, I., Csipkés, L., & Oravecz, C. (2011). A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei, automatikus prozódiai annotáció. In V. Vincze (Ed.), *VIII. magyar számítógépes nyelvészeti konferencia* (pp. 190–198). Szeged: JATEPress.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, *107(3)*, 1697–1714.

Varga, L. (1994). A hanglejtés. In *Struktúrális magyar nyelvtan 2. Fonológia* (pp. 468–549). Budapest: Akadémia Kiadó.

Váradi, T., Simon, E., Sass, B., Gerőcs, M., Mittelholcz, I., Novák, A., ... Vincze, V. (2017). Az e-magyar digitális nyelvfeldolgozó rendszer. In V. Vincze (Ed.), *Xiii. magyar számítógépes nyelvészeti konferencia (mszny2017)* (pp. 49–60). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.

Zellers, M., & Post, B. (2009). Fundamental frequency and other prosodic cues to topic structure. In *Proceedings of IDP* (pp. 377–386). Paris.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/367/2019.PL
Subject: PhD Publikációs Lista

Candidate: István Szekrényes
Neptun ID: SQGGTJ
Doctoral School: Doctoral School of Linguistics
MTMT ID: 10038180

List of publications related to the dissertation

Hungarian book chapters (2)

1. Hunyadi, L., Földesi, A., **Szekrényes, I.**, Staudt, A., Kiss, H., Abuczki, Á., Bódog, A.: Az ember-gép kommunikáció elméleti-technológiai modellje és nyelvtechnológiai vonatkozásai.
In: Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások. Szerk.: Kenesei I., Prószéky G., Várady T, Akadémiai Kiadó, Budapest, 265-309, 2012. ISBN: 9789630593083

2. **Szekrényes, I.**: Automatizált F0-mérések beágyazott mondatszerkezetek rekurziós sajátosságainak vizsgálatához.
In: Rekurzió a nyelvben I. Prozódiai megközelítés. Szerk.: Hunyadi László, Tinta Könyvkiadó, Budapest, 156-169, 2011, (Segédkönyvek a nyelvészet tanulmányozásához, ISSN 1419-6603 ; 123.) ISBN: 9789639902824

Foreign language international book chapters (3)

3. Kovács, G., **Szekrényes, I.**: Applying Neural Network Techniques for Topic Change Detection in the HuComTech Corpus.
In: The Temporal Structure of Multimodal Communication: Theory, Methods and Applications. Ed.: Hunyadi László, Szekrényes István, Springer International Publishing Ag, Cham, 147-162, 2019. ISBN: 9783030228941

4. Hunyadi, L., Kiss, H., **Szekrényes, I.**: Incompleteness and Fragmentation: Possible Formal Cues to Cognitive Processes Behind Spoken Utterances.
In: Intelligent Decision Technology Support in Practice. Ed.: Jeffrey W. Tweedale, Rui Neves-Silva, Lakhmi C. Jain, Gloria Phillips-Wren, Junzo Watada, Robert J Howlett, Springer International Publishing Ag, Cham, 231-257, 2016. ISBN: 9783319212081

5. Hunyadi, L., **Szekrényes, I.**, Kiss, H.: Prosody Enhances Cognitive Infocommunication: Materials from the HuComTech Corpus.
In: Toward Robotic Socially Believable Behaving Systems. Vol. I : Modeling Emotions. Ed.: Anna Esposito, Lakhmi C. Jain, Springer International Publishing Ag, Cham, 183-204, 2016. ISBN: 9783319310565

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Hungarian scientific articles in Hungarian journals (1)

6. Hunyadi, L., **Szekrényes, I.**, Sziklai, I.: Vizuális percepció és nyelvi feldolgozás. *Beszédkutatás. 23*, 186-208, 2015. ISSN: 1218-8727.

Foreign language scientific articles in Hungarian journals (2)

7. Hunyadi, L., **Szekrényes, I.**, Czap, L., Sziklai, I.: Seeing the sounds? *Argumentum (Debr.). 10*, 325-338, 2014. EISSN: 1787-3606.

8. Pápay, K., Szeghalmy, S., **Szekrényes, I.**: HuComTech Multimodal Corpus Annotation. *Argumentum (Debr.). 7*, 330-347, 2011. EISSN: 1787-3606.

Foreign language scientific articles in international journals (2)

9. **Szekrényes, I.**: Post-processing T-patterns Using External Tools From a Mixed Method Perspective.
Front. Psychol. 10, 1-12, 2019. EISSN: 1664-1078.
DOI: http://dx.doi.org/10.3389/fpsyg.2019.01680
IF: 2.129 (2018)

10. **Szekrényes, I.**: Annotation and interpretation of prosodic data in the HuComTech corpus for multimodal user interfaces.
J. Multimodal User Interfaces. 8 (2), 143-150, 2014. ISSN: 1783-7677.
DOI: http://dx.doi.org/10.1007/s12193-013-0140-1
IF: 0.797

Hungarian conference proceedings (2)

11. Kornai, A., **Szekrényes, I.**: e-Magyar beszédarchívum.
In: XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017). Szerk.: Vincze Veronika, Szegedi Tudományegyetem, Szeged, 103-109, 2017. ISBN: 9789633065181

12. **Szekrényes, I.**, Csipkés, L., Oravecz, C.: A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei. Automatikus prozódiai annotáció.
In: VIII. Magyar Számítógépes Nyelvészeti Konferencia : MSZNY 2011. Szerk.: Tanács Attila, Vincze Veronika, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 190-198, 2011. ISBN: 9789633061213

Foreign language conference proceedings (6)

13. Hunyadi, L., Váradi, T., Kovács, G., **Szekrényes, I.**, Kiss, H., Takács, K.: Human-human, human-machine communication: on the HuComTech multimodal corpus.
In: Selected papers from the CLARIN Annual Conference 2018. Ed.: Inguna Skadina, Maria Eskevich, Linköping University Electronic Press, Linköping, 56-65, 2018. (Linköping Electronic Conference Proceedings, ISSN 1650-3686) ISBN: 9789176850343

14. **Szekrényes, I.**, Kovács, G.: Classification of Formal and Informal Dialogues Based on Turn-Taking and Intonation Using Deep Neural Networks.
In: Speech and Computer : 19th International Conference, SPECOM 2017: Proceedings. Ed.: Alexey Karpov, Rodmonga Potapova, Iosif Mporas, Springer International Publishing Ag, Cham, 233-243, 2017. ISBN: 9783319664286

15. Hunyadi, L., Váradi, T., **Szekrényes, I.**: Language technology tools and resources for the analysis of multimodal communication.
In: Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH). Ed.: Erhard Hinrichs, Marie Hinrichs, Thorsten Trippel, University of Tübingen, Tübingen, 117-124, 2016. ISBN: 9784879747082

16. **Szekrényes, I.**: ProsoTool, a method for automatic annotation of fundamental frequency.
In: 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). Ed.: Baranyi Péter, Csapó Ádám, IEEE, New York, 291-296, 2015.

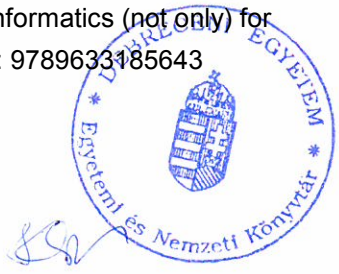
17. Hunyadi, L., **Szekrényes, I.**, Borbély, A., Kiss, H.: Annotation of spoken syntax in relation to prosody and multimodal pragmatics.
In: 3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2012 : Proceedings, December 2-5, 2012 Košice, Slovakia. Ed.: Péter Baranyi, IEEE, Danvers, 537-541, 2012. ISBN: 9781467351881

18. Esfandiari Baiat, G., **Szekrényes, I.**: Topic change detection based on prosodic cues in unimodal setting.
In: 3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2012 : Proceedings, December 2-5, 2012 Košice, Slovakia. Ed.: Péter Baranyi, IEEE, Danvers, 527-530, 2012. ISBN: 9781467351874

List of other publications

Foreign language Hungarian books (1)

19. Csűry, I., Hunyadi, L., Abuczki, Á., Esfandiari Baiat, G., Földesi, A., **Szekrényes, I.**: Elements of Electronic Information and Document Processing: An introduction to informatics (not only) for the humanities. Debrecen Univ. Press, Debrecen, 139 p., 2016. ISBN: 9789633185643



Hungarian scientific articles in Hungarian journals (2)

20. **Szekrényes, I.**: Egy apolitikusan felfogott élet esélyei.
Nagyerdei Almanach. 1, 1-8, 2010. EISSN: 2062-3305.

21. **Szekrényes, I.**: A király meztelen: avagy Egy boldogtalan kritikus "apológiája".
Forrás. 39 (7-8), 161-167, 2007. ISSN: 0133-056X.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Foreign language conference proceedings (1)

22. Hunyadi, L., Bertók, K., Németh, T. E., **Szekrényes, I.**, Abuczki, Á., Nagy, G., Nagy, N., Németi, P., Bódog, A.: The outlines of a theory and technology of human-computer interaction as represented in the model of the HuComTech project.
In: Proceedings of 2nd Cognitive Infocommunications International Conference. Ed.: by Baranyi Péter, IEEE Computer Society, New York, 249-255, 2011. ISBN: 9781457718069

Total IF of journals (all publications): 2,926
Total IF of journals (publications related to the dissertation): 2,926

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

06 November, 2019

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár