



USING PHONOLOGICAL PHRASE
SEGMENTATION TO IMPROVE AUTOMATIC
KEYWORD SPOTTING FOR THE HIGHLY
AGGLUTINATING HUNGARIAN LANGUAGE

Gyorgy Szaszak Andras Beke

Idiap-RR-23-2013

JUNE 2013

Using Phonological Phrase Segmentation to Improve Automatic Keyword Spotting for the Highly Agglutinating Hungarian Language

György Szaszák¹, András Beke²

¹Idiap Research Institute, Martigny, Switzerland

²Research Institute of Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

gyorgy.szaszak@idiap.ch, beke.andras@nytud.mta.hu

Abstract

This paper investigates the usage of prosody for the improvement of keyword spotting, focusing on the highly agglutinating Hungarian language, where keyword spotting cannot be effectively performed using LVCSR, as such systems are either unavailable or hard to operate due to high OOV rates and poor N-gram language modelling capabilities. Therefore, the applied keyword spotting system is based on confidence scores computed as a ratio of acoustic scores obtained in two ways: firstly, by decoding with an universal background model; and secondly, by decoding with a keyword model embedded into filler models. Prosody is used to perform an automatic phonological phrase alignment for speech, proven to be useful for automatic partial word boundary detection in fixed stress languages. Several features deduced from the phonological phrase alignment are investigated to rescore baseline confidence scores both in a rule-based and in a data-driven manner. Results show that in relevant operating points of the system, a false alarm reduction of 10% - 40% can be reached by the same miss probability rates.

Index Terms: spoken term detection, keyword spotting, prosody, agglutinating languages

1. Introduction

Keyword Spotting (KWS) or as recently referred to, Spoken Term Detection (STD) has become a basic task in audio indexing and information retrieval. Keyword spotting can be implemented using various approaches. A basic form of keyword spotting system can be obtained by using pure phoneme (mono-phoneme or triphone) models to build models for keyword terms, and the same phoneme models can then be used in a hooked-like architecture to implement both filler and background models. This kind of KWS is sometimes referred to as *acoustic* keyword spotting [1]. The authors will also use this notation for this kind of KWS system in this paper. The greatest advantages of such an acoustic KWS system are its simplicity, flexibility (regarding vocabulary for spoken terms) and fast operation (online, real-time operation is without problems).

Another, very effective way of keyword spotting is to use Large Vocabulary Continuous Speech Recognizers (LVCSR) to generate a word lattice, and run keyword detection algorithm on the lattice. This type of keyword spotting is often referred to as STD [1]. As a lattice features several concurring hypotheses, keywords have more chance to pass detected. The keyword detection algorithms used on the lattice range from a simple text grep to more flexible methods, especially if not a word, but a sub-word (for example syllable) or phoneme lattice is generated, and spoken term detection is carried out on these, usually using the forward-backward algorithm. The advantage of an

LVCSR based KWS is its better performance (in overall measured performance usually outperforms the acoustic systems), its drawback is its complexity and slow operation: as real-time spoken term detection is most often problematic with LVCSR, only off-line operation is possible. In case of LVCSR based KWS, out-of-vocabulary (OOV) and rare (low probability in language models) words are disfavoured or impossible to detect. Although LVCSR systems usually perform better, this is not true for all operating points and conditions: acoustic KWS is still a better choice in tasks where OOV rates are high or in applications requiring flexible or real-time operation, or for applications requiring low miss probability rates [2].

Some hybrid systems were also proposed to try to combine the advantages of acoustic and LVCSR-based keyword spotting [2], yielding sometimes considerable improvement over baselines.

For highly agglutinating languages (like Hungarian, Finnish, Turkish or Arabic languages), LVCSR task is problematic and less effective due to vocabulary and language model size / coverage problems [3], [4]. For such languages, LVCSR is characterized by high OOV rates, language modelling is also impaired by the size of vocabulary. This also means that LVCSR based spoken term detection is hard to implement for such languages.

As in current experiments, false alarm rates constitute less problem than missed keyword occurrences, and as for the very highly agglutinating Hungarian, LVCSR system is not yet available, the authors decided to use an acoustic KWS system and analyse to what extent prosodic information can improve performance of the KWS system.

Previous research on speech prosody revealed that prosody can be usefully exploited to perform automatic phrasing (a segmentation for phrases) of input speech, where phrases can be prosodic phrases [5], phonological phrases [6], or even further down in the prosodic hierarchy, prosodic words [7]. The depth of this phrasing in terms of the prosodic hierarchy is to some extent language specific [6]. In several studies, such prosodic phrasing (or boundary information based on prosody) was successfully used to improve speech recognition [8] or understanding [9].

Using prosody in KWS was mainly proposed so far for the tonal languages, like Mandarin [10], or the non-tonal, but prosodically very rich Japanese [11]. Yamashita and Mizoguchi [12] modelled F0 contours of Japanese keywords and used these patterns to calculate a prosodic penalty for each keyword candidate. If prosodic similarity was poor, it resulted in a high penalty and the keyword was rejected. Their approach reduced the number of false alarms considerably, however, it strongly relied on the rich prosody of Japanese, often reflecting even

the word structure of an utterance. Since in the majority of languages, prosody does not yield so rich information about word boundaries, and as keywords, even if reflected somewhat by prosody, may occur in different grammatical role and hence with quite different prosodic realization, this approach does not seem applicable in our case. The proposed approach in this paper consists in exploiting phonological phrasing information, especially as previous research has shown that - at least for fixed-stress languages, like Hungarian - (phonological) phrase boundary information can be powerfully used to perform partial word-boundary detection [6], [8].

This paper is organized as follows: Section 2 presents the setup of the baseline acoustic KWS system, Section 3 describes the automatic prosodic phrasing approach. Section 4 presents two approaches, a rule-based and a data-driven one to improve KWS performance using prosodic phrase boundary information. In Section 5, results are presented and evaluated. Finally, conclusions are drawn.

2. KWS setup

The chosen baseline KWS system is an acoustic one, that is, filler and background models are monophone phoneme loops, and keyword models are constructed from the monophone models according to keyword pronunciation(s). The operation of such a system is explained in details in [1]. The reason for using an acoustic KWS - as already mentioned in the Introduction - is twofold: first, as the minimization of the ratio of missed keyword occurrences is more important than reducing radically false alarm rates, an acoustic KWS can be a better choice [2]; secondly, Hungarian is a highly agglutinating language, which makes LVCSR based KWS problematic and less efficient (high OOV rates, poor word probability predictions based on language model due to the extreme high number of different word forms). The more flexibility and faster operation of the acoustic KWS are additional advantages.

The agglutinating nature of the Hungarian language poses problems in acoustic KWS, too. For example, in Hungarian, a single noun might have several hundreds of inflected forms, most of them being acoustically very similar. The detection of another form of the same keyword is a false alarm in the strict sense, but from the user's point of view, it could be even preferred that words reflecting other inflected forms of the same keyword become detected. Therefore, such a keyword spotting system could have two operation modes: in one of them, keyword spotting is performed in the strict sense, in the other, keywords are stemmed before keyword spotting. This latter approach has the drawback that the detection of shorter keywords can be less accurate (causing more false alarms), but solves the problems emerging from the variability of inflected forms. In the present setup evaluated in this paper, keywords were stemmed. This was done manually, but some tools are also available to perform stemming in automatic way [20].

Phoneme models used in the KWS system are 3-state left-to-right monophone Hidden Markov Models (HMM), state emissions are modelled with 32 component Gaussian Mixture Models (GMM). Acoustic features are 13 MFCC with first and second order deltas appended. Models are trained on the Hungarian Reference Speech Database [13] (phonetically rich corpus, 332 Hungarian speakers, mostly read speech).

Keyword spotting is performed using two parallel threads: on one thread, a simple phoneme recognition is executed with the phoneme loop background model, on the other, the chosen keyword is forced to occur in the utterance embedded be-

tween filler models. The confidence score is computed as the difference of the log-scores obtained from the two threads, normalised for the length of the keyword.

The KWS test corpus consisted of an extract of a medical audio archive, containing reports of gastro-esophageal endoscopy for 105 patients. The length of the corpus was approx. 1.3 hours. The selected 50 keywords covered mostly expressions linked to certain gastro-esophageal diseases or conditions. The investigated keywords occurred in the test corpus 728 times.

3. Automatic phonological phrasing

Automatic Phonological Phrasing (APP) is a segmentation process, which identifies phonological phrases (PP) from speech based on prosodic-acoustic features. The approach for Hungarian was presented in details in [6]. Here, a short and basic overview is provided about the operation of the system, but the reader is referred to the above reference for more detailed information.

3.1. Predicting syntax from prosody

APP is based on features derived from F0 and energy, and yields a phonological phrase alignment for speech utterances. According to the *prosodic structure hypothesis* of Selkirk [14], speech utterances are composed of *intonational phrases* (IP), which can be divided into *phonological phrases*. Phonological phrases are built up from *prosodic words* (PW). This means that PPs are placed between IPs and PWs in the prosodic hierarchy.

The correspondence between prosodic and syntactic structure - that is, between words in written text and PPs or PWs in spoken utterances - is close, even if not univocal: mapping prosody to syntax or vice versa is sometimes ambiguous, but generally feasible and usable in speech technology applications. For example, in unit selection TTS, generation (prediction) of prosody based on syntax works well and gives good quality synthesized speech in terms of prosody [15], [16]. Mapping prosody to syntax is less often used, but has also been shown to improve performance or add extra information in speech recognition and understanding applications, [5], [6], [7]. Neuroscience research has also confirmed this close relation between prosody and syntax, as regions of human brain responsible for the perception or production of syntax and prosody collaborate closely and overlap with each other [17], [18].

All these results mean that prediction of syntax can be possible based on prosody, indeed, the used APP in this study has already been proven to be able to detect a high ratio of syntactic word boundaries in speech [6] and also to improve automatic speech recognition [8]. The goal of the present experiments is to use APP output to improve KWS.

3.2. Setup for the prosodic phrasing

The APP system used in this study is based on the one presented in [6]. PPs modelled are given in Table 1.

The theoretic prototype of a PP (*ms*) in Hungarian shows a smart rise of F0 at the stressed syllable, then a slightly descending contour follows. As Hungarian is a fixed-stress language (stress if present can almost always be found on the first syllable), location of the stress within the phonological phrase is not a distinctive feature for this language. Differences between the modelled PPs reflect the strength of stress and higher - IP or utterance - level impact, such as for example continuation rise (*cr*), resulting in a high ending. Higher prosodic level impacts

Table 1: *Phonological phrases modelled for APP.*

Prosodic label	Description
co	Clause onset PP
ss	Strongly stressed PP
ms	Medium stressed PP
ce	Low clause ending PP
cr	High ending (continuation rise) PP
ls	Low-stress PP

influence the prosodic properties of PPs, hence, PP modelling should handle these, and train separate models for such PPs if needed.

Feature extraction, acoustic-prosodic pre-processing and the training of PP models is described in details in [6].

3.3. Automatic alignment of phonological phrases

APP operates theoretically like a Hidden Markov Model based automatic speech recognizer used in unforced segmentation mode (but features are prosodic ones and the models are those of the phonological phrases). All these mean that this tool performs automatic segmentation for PPs: detects hypothesized PP boundaries and classifies the PPs.

During APP, a sophisticated PP sequential model can be used, which has an identical role to a language model in speech recognition. For the present study, the rule-based model presented in [6] was used.

4. Improving KWS using prosody

The basic idea in using PP alignment in KWS is to try to extract some information which can be used to recompute/improve baseline keyword confidence scores. A similar approach is presented in [12], where a prosodic score is computed and then combined to the baseline acoustic score. Indeed, this approach is closely related to lattice rescoring used in speech recognition [21], for example to include prosodic information [8], [9], and theoretically could be used in LVCSR based KWS too.

Yamashita and Mizoguchi argued against a prosodic phrasing based approach in favour of a direct contour modelling for each keyword, as (i) in Japanese, it cannot differentiate between acoustically similar, but prosodically strictly different words, (ii) long compound words might be split into different phrases and (iii) it does not need global processing [12]. For the present paper, a phrasing based approach is still considered more advantageous, as ad (i), the majority of languages (including Hungarian) is neither tonal nor preserves rich word level prosody like the Japanese; ad (ii), prosodic phrasing for phonological phrases is unlikely to split long compounds (whereas phrasing for prosodic words usually splits them) [6]; ad (iii) global prosodic processing is very fast (RTF ~ 0.01 on standard processors), and allows for flexible modelling, i.e. there is no need for previously prepared prosodic models for the keywords. Further on, keywords may occur in different places within the sentence, especially in Hungarian, characterized by free word order, which influences the prosodic properties of keywords too.

Based on the PP alignment, a set of new features was selected and analysed whether it helps KWS. Features are based mostly on phonological phrase boundaries (PPB) and phonological phrase labels (PPL), presented in Table 2.

Table 2: *Features derived for confidence rescoring.*

Feature	Description
<i>KWCS</i>	KW confidence score (from acoustic KWS)
ΔT_s	Frame shift betw. PPB and supposed KW onset
ΔT_e	Frame shift betw. PPB and supposed KW end
<i>BPPL</i>	Label (type) of the best overlapping PP
<i>OLR</i>	Part of KW covered by BPPL (%)
<i>IOLR</i>	Part of BPPL covered by KW (%)
<i>SILR</i>	Part of KW covered by prosodic silence (%)
<i>PPCS</i>	Prosodic confidence score of BPPL

4.1. Feature selection

Feature selection is used to identify features with large discrimination power and also those which can be ignored as they can be regarded irrelevant in the given decision or classification task. For this research, selection of features was based on Receiver Operation Characteristic (ROC) curves and the Area Under Curve (AUC) measure, calculated for the ROC. AUC is a strong predictor of performance, especially in imbalanced data classification problems, where it can be used for feature ranking. The measurement of the ROC AUC is very simple and fast compared to other feature selection algorithms [23]. Each proposed feature was hence analyzed in terms of its keyword discrimination power in order to discover features that can contribute to statistically significant improvement in KWS performance.

Discrimination power measured by ROC AUC is given in Table 3 for features computed from APP, and also for the original baseline confidence score used in the acoustic KWS (*KWCS*). Features ΔT_e , *OLR*, *PPCS* were not found to discriminate between keywords and non keywords significantly better than chance, and hence are not listed in Table 3.

Table 3: *Discrimination power of APP derived features (AUC).*

Feature	ROC AUC	Remark
<i>KWCS</i>	0.931	baseline
ΔT_s	0.630	labels mapped to scores
<i>BPPL</i>	0.588	
<i>IOLR</i>	0.652	
<i>SILR</i>	0.808	

Not surprisingly, the highest AUC was found for *KWCS*, that is, the keyword confidence score of the baseline acoustic KWS system. Out of the features reflecting phonological phrasing information, *SILR* was found to be of good discrimination power. This feature reflects what percent of a keyword candidate is covered by silence detected on prosodic features as part of the phonological phrasing. Performing a deeper analysis revealed that a considerable part of false keyword candidates were supposed to occur around pauses, PP or rather IP endings or parts of speech corrupted by speaker or background noises, as these are the parts of the audio signal where keywords can be aligned at the lowest cost by the acoustic KWS. APP is a robust silence detector, moreover, often truncates glottalized, low-energy phrase endings and detects silence instead [6]. This means that this property, disadvantageous in other applications, might contribute to unveil a part of false alarms in KWS.

Regarding the coincidence between PP boundaries and keyword endings, assessed by feature ΔT_e , it is possible that the

applied keyword stemming is responsible for the weak discrimination power, as in case of stemming, only the initial part of a word is regarded to be the keyword, while phonological phrasing, if it is really a good predictor for word-boundaries as found in [6], still refers to the end of the original, inflected form of the word. In contrast, ΔT_s showed some, although not really high discrimination power. This can also be linked to the word-boundary prediction capabilities of PP boundaries, however, the uncertainty of the word-boundary prediction in terms of exact localization in time, equivalent typically to 1-3x average phoneme length seems to be more disturbing in KWS than in standard speech recognition (cf. [8]). The similar discrimination power of feature *IOLR* can be explained also with coincidence between phonological phrasing and word boundaries.

Although any of the features computed from APP is a weaker standalone classifier than *KWCS* alone, combining them can lead to KWS performance improvement if keywords, or a part of keywords, missed by one feature can be detected by another; or if a false keyword detected based on *KWCS* can be rejected based on them. This is investigated in next sections.

4.2. Rule based approach

As the number of features is low, a rule-based approach was also considered, even if it has some risk of overspecialisation. Features found discriminative were linearly combined to *KWCS* with tunable weight parameters. Weights were tuned on a development set, separated from test data, representing approx. 25% of all utterances.

4.3. Logistic regression based approach

Logistic regression is a type of regression suitable for predicting the actual value for a binary dependent variable, based on one or more independent variable(s) [24]. In contrast to linear regression, logistic regression uses the logit function, characterized by an 'S' shape (similar to sigmoid function), which makes it a better predictor if the dependent variable is binary and hence not normally distributed. In KWS, this can be used to predict whether a keyword is present in the speech or not, that is, to generate a confidence score based on features derived from prosodic phrasing. Adding the baseline keyword confidence score from the acoustic KWS to the set of independent variables X will result in a combination of the baseline score and prosodic measures. The estimated new confidence score, $\Pi(X)$ is then computed from the following equation:

$$\Pi(X) = \frac{1}{1 + e^{-\sum_{k=0}^K x_k \beta_k}}. \quad (1)$$

Logit parameters ($\beta_0 \dots \beta_K$) are usually estimated by ML approach, however, in case of multicollinearity or few data, this estimation is not convenient. As an alternative, ridge regression estimation method can be used [25]. This latter was used in the presented experiments too. Logit parameters were estimated on approx. 25% of the data.

5. Results

Results are evaluated using DEtection Trade-off (DET) curves [22], commonly used for the evaluation of KWS systems. Equal Error Rates (EER) were also determined. DET plots are shown in Fig. 1, whereas EER for the investigated setups were as follows: 13.19% for the baseline system, 12.15% and 12.04% for the improved systems for the rule based and the logit approach,

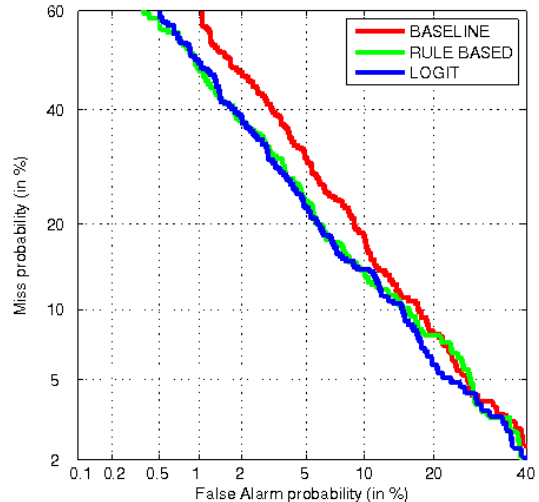


Figure 1: *Detection Trade-off curves for baseline and improved KWS.*

respectively. Analyzing the DET curves show that phonological phrasing based features significantly improved KWS performance, both in the rule based and the logit approach. The logit approach performed better although the difference compared to the rule based one was not significant. DET curves show significant improvements for operating points defined by miss probability $P_{miss} > 0.05$, where a false alarm reduction ranging 10-40% was reached.

6. Conclusions

This paper addressed the use of phonological phrasing to improve keyword spotting. As the target language was the highly agglutinating Hungarian, where LVCSR based spoken term detection is problematic, an acoustic KWS system was used. Several features computed based on the phonological phrasing were analyzed in terms of their discrimination power for KWS. Features related to phonological phrase start points compared to keyword candidate start points, keyword candidate coverage by the phonological phrasing and partial coverage of candidates by prosodically detected silence were found to be discriminative features. Using these to recompute confidence scores gave significant performance improvement in KWS. There was no significant difference between the rule based and the logistic regression based rescoring approach for the selected task. Results showed up to 40% reduction in false alarm probability in operation points defined by the same miss probability. As results of previous studies proved that phonological phrasing based lattice rescoring yields improvement in speech recognition and as current findings prove that phonological phrasing can contribute to improve KWS, a similar approach used for LVCSR lattice rescoring in KWS may also improve performance, however, this issue could not be investigated for Hungarian given the lack of a real LVCSR system.

7. Acknowledgements

This work was done with the help of the Laboratory of Speech Acoustics at the Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics.

8. References

- [1] Szöke, I., Schwarz, P., Matejka, P., Burget, L., Karafiát, M., Fapso, M. and Cernocky, J., "Comparison of Keyword Spotting Approaches for Informal Continuous Speech", Proc. of Interspeech 2005, Lisbon, Portugal, pp. 633-636, 2005.
- [2] Motlicek, P., Valente, F. and Szöke, I., "Improving Acoustic Based Keyword Spotting Using LVCSR Lattices", Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP) 2012, Japan, pp.4413-4416, 2012.
- [3] Carki, K., Geutner, P. and Schultz, T., "Turkish LVCSR: Towards Better Speech recognition for Agglutinative Languages", Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP) 2000, Istanbul, Turkey, pp. 3688-3691. 2000.
- [4] Mihajlik, P., Fegyó, T., Tüske, Z. and Ircing, P., "A morpho-graphemic approach for the recognition of spontaneous speech in agglutinative languages - like Hungarian", Proc. Interspeech 2007, Antwerp, Belgium, pp. 1497-1500, 2007.
- [5] Gallwitz, F., Niemann, H., Nöth, E. and Warnke, W., "Integrated recognition of words and prosodic phrase boundaries", Speech Communication, 36:81-95, 2002.
- [6] Szaszák, G. and Beke, A., "Exploiting Prosody for Syntactic Analysis in Automatic Speech Understanding" Journal of Language Modelling, 0(1):143-172, 2012.
- [7] Hirose, K. Minematsu, N. Hashimoto, Y. and Iwano, K., "Continuous Speech Recognition of Japanese Using Prosodic Word Boundaries Detected by Mora Transition Modeling of Fundamental Frequency Contours", Proceedings of ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, Red Bank, NJ, USA, pp.61-66, 2001.
- [8] Vicsi, K. and Szaszák, G., "Using prosody to improve automatic speech recognition", Speech Communication 52(5):413-426, 2010.
- [9] Nöth, E., Batliner, A., Kiessling, A., Kompe, R. and Niemann, H., "Verbmobil: the use of prosody in the linguistic components of a speech understanding system", IEEE Trans, ASSP, Vol. 8, pp. 519-532. 2000.
- [10] Chen, Yeou-Jiunn J., Wu, Chung-Hsien H. and Yan, Gwo-Lang L., "Utterance verification using prosodic information for Mandarin telephone speech keyword spotting" IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. vol 2. pp- 697-700, 1999.
- [11] Ida, M. and Yamasaki, R., "An evaluation of keyword spotting performance utilizing false alarm rejection based on prosodic information", Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia, 1998.
- [12] Yamashita, Y. and Mizoguchi, R., "Keyword spotting using f0 contour matching", Proceedings of 5th Conference on Speech Communication and Technology (Eurospeech '97), volume 1, pp. 271-274, 1997.
- [13] Vicsi, K., Kocsor, A., Teleki Cs. and Tóth, L., Beszédadatbázis irodai számítógép-felhasználói környezetben, Second Conference on Hungarian Computational Linguistics (MSZNY 2004), Szeged, Hungary, p. 315, 2004.
- [14] Selkirk, Elisabeth, "The Syntax-Phonology Interface", in Smelser, N.J. and Baltes, P.B. (Eds), International Encyclopaedia of the Social and Behavioural Sciences, Oxford: Pergamon, pp. 15407-15412. 2001.
- [15] Chen, K., Hasegawa-Johnson, M. and Cohen, A., "An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model", Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP), (1):509-512, 2004.
- [16] Koutny, I., Olasz, G. and Olasz, P., "Prosody prediction from text in Hungarian and its realisation in TTS conversion", International Journal of Speech Technology, (3-4):187-200, 2000.
- [17] Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A. and Wildgruber, D., "The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax", Journal of Neurolinguistics, (15):59-90, Elsevier, 2002.
- [18] Strelnikov, K.N., Vorobyev, V.A., Chernigovskaya, T.V. and Medvedev S.V., "Prosodic clues to syntactic processing – a PET and ERP study", NeuroImage (29)4:1127-1134. 2006.
- [19] Roach, P. et al., "BABEL: An Eastern European multi-language database", Proc. of the 4th International Conference on Speech and Language Processing, Philadelphia, USA, Vol. 3. pp. 1892-1893. 1996.
- [20] Halácsy, P. and Trón, V., "Benefits of Resource-Based Stemming in Hungarian Information Retrieval", In: Evaluation of Multilingual and Multi-modal Information Retrieval. 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain; Revised Selected Papers: Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M. (Eds.), Vol. 4730, 2007, ISBN 978-3-540-74998-1, Softcover, pp. 101-106, 2006.
- [21] Ljolje, A., Pereira, F. and Riley, M., "Efficient general lattice generation and rescoring", Proc. of Eurospeech'99, Budapest, Hungary, pp. 1251-1254, 1999.
- [22] Martin, A., Doddington, G., Kamm, T., Ordowski, M. and Przybocki M., "The DET Curve in Assessment of Detection Task Performance", In: Proc. of Eurospeech97. Rhodes, Greece, pp. 1895-1898, 1997.
- [23] Rui Wang, R. and Tang, K. "Feature Selection for Maximizing the Area Under the ROC Curve" 2009 IEEE International Conference on Data Mining Workshops, pp.400-405, 2009.
- [24] Hosmer D.W. and Lemeshow S., "Applied logistic regression", New York: Wiley, 1989.
- [25] Schaefer, R., Roi, L. and Wolfe, R., "A ridge logistic estimator", Communications in Statistics: Theory and Methods 13, pp. 99-113. 1984.