Extending AuToBI to prominence detection in European Portuguese

Helena Moniz[1,2], *Ana Isabel Mata*[2], *Julia Hirschberg*[3],
Fernando Batista[1,4], *Andrew Rosenberg*[5], *Isabel Trancoso*[1,6]

[1]Spoken Language Systems Lab - INESC-ID, Lisbon, Portugal
[2]FLUL/CLUL, Universidade de Lisboa, Portugal
[3]Department of Computer Science, Columbia University, United States
[4]ISCTE - Instituto Universitário de Lisboa, Portugal
[5]Computer Science Department, Queens College (CUNY), United States
[6]IST, Universidade de Lisboa, Portugal

{`helenam;fmmb;imt`}`@l2f.inesc-id.pt`, `aim.@fl.ul.pt`, `julia@cs.columbia.edu`

Abstract

This paper describes our exploratory work in applying the Automatic ToBI annotation system (AuToBI), originally developed for Standard American English, to European Portuguese. This work is motivated by the current availability of large amounts of (highly spontaneous) transcribed data and the need to further enrich those transcripts with prosodic information. Manual prosodic annotation, however, is almost impractical for extensive data sets. For that reason, automatic systems such as AuToBi stand as an alternate solution. We have started by applying the AuToBI prosodic event detection system using the existing English models to the prediction of prominent prosodic events (accents) in European Portuguese. This approach achieved an overall accuracy of 74% for prominence detection, similar to state-of-the-art results for other languages. Later, we have trained new models using prepared and spontaneous Portuguese data, achieving a considerable improvement of about 6% accuracy (absolute) over the existing English models. The achieved results are quite encouraging and provide a starting point for automatically predicting prominent events in European Portuguese.

Index Terms: prosody, automatic prosodic labeling system, spontaneous speech.

1. Introduction

The role of detecting prosodic events is becoming more and more pervasive in different automatic speech processing tasks. The detection of prosodic events has proved to be useful for, *e.g.*, improving speech summarization[1], in ASR models[2, 3, 4], in predicting ASR recognized turns in dialogue systems[5], in predicting structural metadata events[6, 7, 8, 9], in improving unit-selection synthesis[10] or in detecting phonological units for expressive speech synthesis[11], and in identifying paralinguistic events[12].

The literature has documented a set of acoustic and visual correlates of prominence - duration, intensity, pitch, voice quality, and visual cues[13, 14, 15, 16, 17, 18, 19, 20, 21]. Systems built to predict prominence based on acoustic correlates are being used for cross-language studies. One such example is the **Au**tomatic **ToBI** annotation system (AuToBI) for Standard American English (SAE) by[17, 22]. AuToBI is a publicly

available tool[1], which detects and classifies prosodic events following SAE intonational patterns. AuToBI relies in the fundamentals of the ToBI system, meaning it predicts and classifies tones and break indices using the acoustic correlates - pitch, intensity, spectral balance and pause/duration. AuToBI uses a modular architecture, which allows to perform six tasks separately and provides English trained models for spontaneous and read speech (for further details, *vide*[17] and references therein). The six tasks correspond to: i) detection of pitch accents; ii) classification of pitch accent types; iii) detection of intonational phrase boundaries; iv) detection of intermediate phrase boundaries; v) classification of intonational phrase boundary tones; and vi) classification of intermediate phrase boundary tones. Previous work on prosodic event detection using AuToBI[17, 22, 23, 24] have shown that prominence and phrase boundaries can be predicted in a cross-language (American English, German, Mandarin Chinese, Italian and French) context, albeit with language specific properties. Those studies also found little support for the hypothesis that language families are useful for cross-language prosodic event identification. Taking into account the described results, the aim of this work is to extend the AuToBI prosodic event detection system from English to Portuguese in two stages. First, English models are used to predict prosodic events in European Portuguese (detection and classification of pitch accents). Second, AuToBI capabilities are adapted using a relatively small amount of annotated data to train Portuguese models.

This paper is organized as follows. Section 2 comprises the description of the corpus used in the experiments. Section 3 presents the process of converting of tone inventories from P-ToBI to SAE, a pre-requirement to apply AuToBI to European Portuguese. Section 4 describes the experiments conducted and the main results achieved. Section 5 presents our conclusions and future work.

2. Corpus

This study uses a subset of CPE-FACES [25, 26], an European Portuguese Corpus Spoken by Adolescents in School Context. The corpus consists of spontaneous and prepared unscripted speech from 25 students (14-15 years old) and 3 teachers, all

---

[1]http://eniac.cs.qs.cuny.edu/andrew/autobi/

Corpus subset →	train	test	total
total time (minutes)	33.4	10.9	44.2
useful time (minutes)	20.6	6.6	27.2
number of pitch accents	2061	717	2778
number of phrasing units	1382	489	1871
number of words	4361	1456	5817

Table 1: Properties of the CPE-FACES subsets.

Figure 1: Schematic F0 contours. Thick red lines indicate the stressed syllable.

speakers of standard European Portuguese (Lisbon region), totaling approximately 16h. It was designed to represent some of the speech tasks that are common in school context and it was collected in the last year of compulsory education (9th grade), in three Lisbon public high schools. It was recorded in a natural setting – the speakers classroom of Portuguese as L1 – in different communication situations: two dialogues (both spontaneous) and two oral class presentations (one spontaneous and another one prepared, but unscripted). In the spontaneous presentation, students and teachers were unexpectedly asked to relate an (un)pleasant personal experience. It was assumed that the involvement of speakers on topics related to their personal interests and day-to-day life would manifest in the naturalness and spontaneity of their talks [27]. The prepared situation corresponds to typical school presentations; the presentation was about a book the students must read following specific programmatic guidelines. For students, a variety of presentations on Ernest Hemingway's "The Old Man and the Sea" and on Gil Vicente's "Auto da Índia" was recorded. As for the teachers, all prepared presentations were related to the study of "Os Lusíadas" by Luís de Camões, and two address the same episode - the lyric-tragic episode of Inês de Castro.

Basically, spontaneous and prepared presentations differ in the degree of planning involved, the type of information communicated, the speakers' attention to the speech task and effort to speak clearly. In spontaneous presentations, the speakers can talk freely about any topic of their choice; they can change topic and move on to another topic whenever they feel like it. As far as typical (prepared) oral presentations at school are concerned, it was argued before that "more than talking about a pre-determined theme, an oral presentation presupposes the capacity to individually produce a greater amount of utterances, organizing the information that is given to the public in a clear structured form" [25].

The recordings of the two female teachers and all the students were done with an UHER 400 Report Monitor recorder with a BASF LPR 35 magnetic tape and a SENNHEISER MD 214 V-3 worn suspended from the neck microphone. These recordings were latter digitized at 44.1 kHz, using 16 bits/sample and afterwards downsampled to 16 kHz. CPE-FACES was recently extended with the recordings of a male teacher, using a TASCAM HD-P2, a Portable High Definition Stereo Audio Recorder, and a head-mounted microphone Shure, a Sub-miniature Condenser Head-worn Microphones, model Beta 54. The sound was recorded in mono, with 16-bit at samples rates of 44.1kHz, and afterwards downsampled to 16 kHz.

The subset of the corpus used in this study comprises 9 spontaneous presentations and 9 prepared unscripted presentations, from 6 teenage students (balanced by gender) and 3 teachers (2 female and 1 male). The data was split into train and test subsets, where the training part corresponds to about 75% of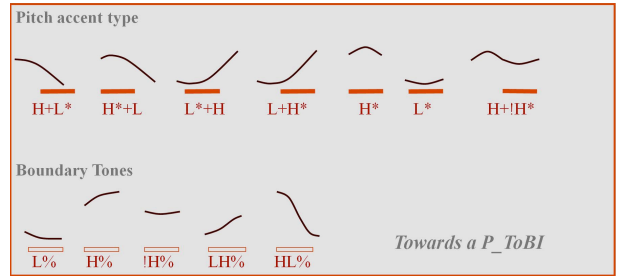 each speech file and the test part corresponds to the remaining 25%. Both subsets include portions of each speech file, which reduces the bias of training and testing using different conditions. Even so, one should note that the test set of our data is harder to automatically process, mostly due to the fact that by the end of a presentation interlocutors tend to interrupt more, leading to more overlapping speech. These rich interactions result in harder stretches of speech to process. The overall statistics about the subset of the corpus are presented in Table 1, where "total time" correspond to useful time and silences. The useful time is similar to other corpora subsets used by [23] for cross-language detection of prosodic events with AuToBI, namely the Italian subset (25 minutes of read speech by a single speaker). Having an equivalent amount of data (at this point a small sample of 27 minutes) allow us to make more direct comparisons with previous studies applying AuToBI.

The data was annotated by two expert linguists using the ToBI prosodic system adapted to European Portuguese (Towards a P_ToBI by [28]), in order to conduct experiments on automatic ToBI-labeling in European Portuguese, as part of the ongoing project[2] that funded this research. All the pitch accents (H+L*, H*+L, L*+H, L+H*, H*, L*, H+!H*) and the final boundary tones (L%, H%, !H%, LH%, HL%) that are covered in the Towards a P_ToBI proposal were used in this subset (see the schematic F0 contours in Figure 1). For further details on the annotation of the corpus, *vide* [26, 29].

3. Converting tone inventories

As previously said, AuToBI relies on the fundamentals of SAE. Therefore, as a pre-requirement to apply AuToBI to European Portuguese, the tonal inventory of P_ToBI, displayed in Figure 1, had to be converted converted to the SAE inventory. Since this work is our first step towards fully automatic prosodic labeling, at this stage no feature adaptation (pitch, intensity, spectral balance and pause/duration) was done. The conversion process was jointly carried out with one of the American co-authors of this paper. This task involved the analysis of a considerable amount of examples for each pitch accent and boundary tone in the P_ToBI inventory. Table 2 summarizes the conversions that have been made.

The main differences concern falling pitch accents and single boundary tones at intonational phrases (IPs). First, H+L* - a very frequent nuclear accent in European Portuguese (associated with declaratives and wh- question), which is absent in the SAE inventory. Second, H*+L - relatively uncommon in the subset analyzed, which is also absent in the SAE inventory. Third, H% and L% - single boundary tones at intonational

---

[2]PTDC/CLE-LIN/120017/2010.

P_ToBI	SAE
H+L*	H+!H*
H*+L	H*
L%	L-L%
H%	H-H%
	H-L%
HL%	!H-L%
!H%	H-L%
H*	H*
L*	L*
L+H*	L+H*
L*+H	L*+H
H+!H*	H+!H*
LH%	L-H%
L-	L-
H-	H-
!H-	!H-

Table 2: Inventory conversion from P_ToBI to SAE. In the first part the events changed and in the second part the ones kept.

EN models PT.all	Accented			Unaccented			
	Prec	Rec	F	Prec	Rec	F	Acc
Overall	80.8	64.9	72.0	70.0	84.1	76.4	**74.4**
Students	73.7	47.6	57.8	64.3	84.7	73.1	67.1
Teachers	83.4	73.8	78.3	74.3	83.8	78.8	78.5

Table 3: Prominence prediction, using EN models and PT data.

EN models PT.test	Accented			Unaccented			
	Prec	Rec	F	Prec	Rec	F	Acc
Overall	74.0	70.8	72.3	67.9	71.3	69.6	**71.0**
Students	61.1	63.1	62.1	55.8	53.8	54.8	58.8
Teachers	82.4	75.3	78.7	74.0	81.5	77.6	78.2

Table 4: Prominence prediction, using EN models and PT test set.

phrases, as proposed by [30, 28]. In [30, 28], two different levels of phrasing are equated both to the intonational phrase: the major IP and the minor IP, in line with [31]. Minor and major IPs are marked with breaks 3 and 4, respectively, and the diacritics "-" and "%" represent the different strengths of the IP boundaries. Although [32] proposes a reanalysis marking both IP levels as "%", the labels "-" and "%" were kept in the subsets of CPE-FACES. Table 2 also displays the tones shared by both intonational systems (H*, L*, L+H*, L*+H, H+!H*, LH%, L-, H-, !H-).

4. Results

This section presents our first efforts towards the automatic detection of prominent prosodic events in European Portuguese. We have started by identifying prominent events using exclusively English models. The evaluation of such models was initially performed using the Portuguese data as a whole and later using the test set exclusively. Finally, we have trained new Portuguese models and verified their impact on prominence detection. Portuguese models were trained firstly using the whole training set (which includes students and teachers data) and later using exclusively teachers data. In all steps, results were evaluated using standard performance metrics [33]: precision, recall, f-measure, and accuracy, which can be expressed in terms of true positives (*tp*), true negatives (*tn*), false positives (*fp*), and false negatives (*fn*) as follows:

$$precision = \frac{tp}{tp+fp},$$

$$recall = \frac{tp}{tp+fn},$$

$$accuracy = \frac{tp+tn}{tp+fp+fn+tn},$$

$$Fmeasure = \frac{2 \times precision \times recall}{precision+recall}$$

4.1. Prominence detection using English models

After the conversion process described in Section 3, existing English (EN) models were applied to the Portuguese (PT) data. The English models used correspond to the 1.3 version, which

includes training material from three corpora: Boston Directions Corpus, Boston University Radio News Corpus, Columbia Games Corpus ([17] and references therein). Building upon previous studies using AuToBI [23, 24], we hypothesized that prominence could be fairly detected by using English models on Portuguese data, with no further adaptations of the system rather than the conversions of tonal inventories.

Table 3 presents the results of prominence prediction after applying the English models to the Portuguese data, discriminating between students and teachers. Results show that overall prominent events are detected with 72.0% of f-measure (80.8% of precision and 64.9% of recall). The accuracy of prominence prediction using English models is 74.4%. Thus, the results provide evidence that a considerable percentage of prominent events may be predicted from English to Portuguese, supporting previous research using AuToBI [23] for West-Germanic and Romance Languages, Portuguese not included. Table 3 displays a striking difference of around 11% accuracy between speakers (67.1 for students *vs.* 78.5% for teachers). Results are clearly better for teachers than for students, evidencing an age/status dependent effect on the prominence detection tasks.

English models were also applied to the test set of the Portuguese data exclusively. It is important to emphasize that, as previously stated, the test set of our data is harder to automatically process due mostly to very lively interactions between interlocutors. We believe that this is the core reason for the poorer results presented in Table 4 when compared with the ones in Table 3: the overall accuracy decreases by about 3% in the Portuguese test set. This result is mostly related to the overall prediction for students. For teachers the results are quite similar, with a slight improvement in prominence prediction (from 78.3% to 78.7% f-measure). To sum up, using exclusively the Portuguese test set, the overall accuracy decreases from 74.4% to 71% and the main differences between speakers still stand.

4.2. Prominence detection using Portuguese models

Previous results suggest that the prediction of prominence in Portuguese is fairly accounted for with English trained models. In this section, we target the training of Portuguese models and their impact on prominence detection. We expect that Portuguese trained models improve the overall accuracy of

PT models	Accented			Unaccented			
PT.test	Prec	Rec	F	Prec	Rec	F	Acc
Overall	78.8	78.6	78.7	75.4	75.6	75.5	**77.2**
Students	73.8	66.9	70.2	65.6	72.7	69.0	69.6
Teachers	81.3	85.4	83.3	82.1	77.3	79.6	81.6

Table 5: Prominence prediction, using PT models and PT test set.

Classified as →	Accented	Unaccented
Accented	613	165
Unaccented	167	511

Table 6: Confusion matrix for prominence prediction.

prominence detection, as language-dependent models usually enhance a system performance.

At this point, we have extended AuToBI capabilities to Portuguese by training Portuguese models with the Portuguese train set. No feature adaptation was performed, since we aim at evaluating the impact of language specific data integration in the overall accuracy. Table 5 summarizes the results achieved, showing an overall accuracy increase of around 6% in prominence prediction (77.2%), when compared with previous results obtained with English models (71.0%, see Table 4). Moreover, better performances are achieved for both students and teachers, with accuracy improvements of 11% for students and 3% for teachers. Comparatively, prominence detection show a higher gain for students than teachers, however, there are still striking inter-speaker differences in terms of age/status. Results show that training with language specific data improves prominence detection, in line with our expectations.

The confusion matrix in Table 6 shows prominence (mis)detection, displaying a higher percentage of non-prominent events classified as prominent (24.6% *vs.* 21.2%).

Along this work we have been pointing out striking differences between speakers, showing that teenage data is clearly more difficult to deal with than teachers data. Taking such differences into account, we have trained a new model using exclusively teachers training data. Such model allows us to perform more direct comparisons with previous experiments for other languages, which are commonly based on adult speech, either read or spontaneous speech (task-oriented). The training material for this final evaluation corresponds to 18.8 minutes (silences included) of teachers data only, aiming at producing a more homogeneous training data. In this final evaluation, the overall accuracy decreases to 59.7% (53.7% for students and 63.2% for teachers), which clearly demonstrates that the amount of training material is insufficient. Moreover, these results also show that is preferable to have large amounts of training material from English than having small language dependent samples (see Table 4). This also suggests that increasing the amount of Portuguese training data will likely lead to even better results than the ones already achieved (see Table 5).

The comparison between language dependent or cross-language models applied to the same test set is on its own very informative. Namely, we could verify that using models trained with large amounts of English data captures a considerable amount of prominent events. Our results for European Portuguese are closer to the ones presented in [23], which reported an accuracy ranging from 62.9% to 82% for West-Germanic and Romance languages, as French and Italian. This may suggest that a considerable amount of acoustic information is shared amongst different typological languages.

5. Conclusions and future work

This paper presented our first steps towards the extension of the AuToBI prosodic event detection system to European Portuguese. The first step concerned the prediction of prominent prosodic events based on models trained for English. In the second step the AuToBI models were retrained with a moderate sample of Portuguese ToBI annotated data. This training corpus includes both spontaneous and prepared unscripted speech from both adults and teenagers (teachers and students). This is not the typical starting point for such cross language experiments, but could not be avoided. As expected, the results showed poorer performance in prominence prediction for teenagers, contrasting with state-of-the-art results achieved for adults' data, regardless of the models used. This shows a clear age/status dependent effect present in our data.

Regarding the experiments conducted using the English models on Portuguese data, results showed an overall accuracy of 74.4% of prominence detection, comparable to state-of-the-art results [23]. This result further supports previous predictions using the same English models, which show that prominence detection is fairly accounted for one language to another [23]. When adapting AuToBI capabilities to train Portuguese models, as expected, the prediction of prominent events further improved in about 6% absolute. We may interpret those results as pointing out to two main directions: first, a considerable amount of prominent prosodic events may be cross-language predicted, even when tackling a sample of different typologic languages; second, albeit the reasonable cross-language prediction rate, there are language specific traits captured only with language-dependent trained models. This raises several research questions, *e.g,* on cross-language universal acoustic correlates of prominence vs. language dependent acoustic correlates; on genre free or genre dependent prosodic properties; on tonal density across languages; or even on informational structure. With this work we hope to contribute to prominence detection research, still scarcely studied.

Our preliminary results for European Portuguese are quite encouraging and a starting point to further enrich large amounts of (highly spontaneous) transcribed data available for our language. Therefore, future work will tackle the prediction of other prosodic events (phrasing and tonal boundaries), and the evaluation of the trained Portuguese models for different genres (university lectures, map-task dialogues, broadcast news, broadcast interviews, meetings, etc.). This will also allow us to address the study of age-specific and status-specific properties.

6. Acknowledgments

7. References

[1] S. Maskey and H. Hirschberg, "Comparing lexical, acoustic/prosodic, structural and discourse features for speech

summarization," in *Proc. of Interspeech*, Lisbon, Portugal, 2005.

[2] K. Chen and M. Hasegawa-Johnson, "How prosody improves word recognition," in *Proceedings of Speech Prosody 2004*, Nara, Japan, 2004, pp. 583–586.

[3] E. Shriberg, "Spontaneous speech: How people really talk, and why engineers should care," in *Proc. Eurospeech*, Lisbon, Portugal, 2005, pp. 1781 – 1784.

[4] E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, and A. Stolcke, "Modeling prosodic feature sequences for speaker recognition," *Speech Communication*, vol. 46, no. 3-4, pp. 455–472, 2005.

[5] H. Hirschberg, D. Litman, and M. Swerts, "Prosodic and other cues to speech recognition failures," *Speech Communication*, no. 43, pp. 155–175, 2004.

[6] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1526–1540, 2006.

[7] M. Ostendorf, B. Favre, R. Grishman, D. Hakkani-Tür, M. Harper, D. Hillard, J. Hirschberg, H. Ji, J. G. Kahn, Y. Liu, S. Maskey, E. Matusov, H. Ney, A. Rosenberg, E. Shriberg, W. Wang, and C. Wooters, "Speech segmentation and spoken document processing," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 59–69, 2008.

[8] F. Batista, H. Moniz, I. Trancoso, and N. Mamede, "Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts," *Transactions on Audio Speech and Language Processing*, no. 20, pp. 474–485, 2012.

[9] H. Moniz, F. Batista, I. Trancoso, and A. I. Mata, "Automatic structural metadata identification based on multilayer prosodic information," 2013, DISS 2013.

[10] V. Strom, A. Nenkova, R. Clark, Y. Vazquez-Alvarez, J. Brenier, S. King, and D. Jurafsky, "Modelling prominence and emphasis improves unit-selection synthesis," in *Proc. of Interspeech*, Antwerp, Belgium, 2007.

[11] G. Anumanchipalli, A. Black, and L. Oliveira, "Data-driven intonational phonology," *The Journal of The Acoustical Society of America*, vol. 134, pp. 4237–4237, 2013.

[12] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, "Paralinguistics in speech and language - state-of-the-art and the challange," *Computer Speech and Language*, no. 27(1), pp. 4–139, 2013.

[13] S. Jun, *Prosodic typology: the phonology of intonation and phrasing*. Oxford University Press, 2005.

[14] P. Wagner, "Great expectations - introspective vs. perceptual prominence ratings and their acoustic correlates," in *Proc. of Interspeech*, Lisbon, Portugal, 2005.

[15] F. Tamburini and C. Caini, "An automatic system for detecting prosodic prominence in american english continuous speech," *International Journal of Speech Technology*, vol. 8, pp. 33–44, 2005.

[16] M. Swerts and E. Krahmer, "Facial expression and prosodic prominence: effects of modality and facial area," *Journal of Phonetics*, vol. 36, pp. 219–238, 2008.

[17] A. Rosenberg, "Automatic detection and classification of prosodic events," Ph.D. dissertation, University of Columbia, 2009.

[18] J. Cole, Y. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence," *Laboratory Phonology*, vol. 1, pp. 425–452, 2010.

[19] A. Windmann, I. Jauk, F. Tamburini, and P. Wagner, "Prominence-based prosody prediction for unit selection synthesis," in *Proceedings of Interspeech 2011*, Florence, Italy, 2011.

[20] P. Prieto, M. Vanrell, L. Astruc, E. Payned, and B. Post, "Phonotactic and phrasal properties of speech rhythm. evidence from catalan, english, and spanish," *Speech Communication*, vol. 54(6), pp. 681–702, 2012.

[21] M. Mehrabani, T. Mishra, and A. Conkie, "Unsupervised prominence prediction for speech synthesis," in *Proceedings of Interspeech 2013*, Lyon, France, 2013.

[22] A. Rosenberg, "Autobi – a tool for automatic tobi annotation," in *Interspeech 2010*, 2010.

[23] A. Rosenberg, E. Cooper, R. Levitan, and J. Hirschberg, "Cross-language prominence detection," in *Proc. of Speech Prosody*, Shangai, China, 2012.

[24] V. Soto, E. Cooper, A. Rosenberg, and J. Hirschberg, "Cross-language phrase boundary detection," in *Proc. of ICASSP*, Vancouver, Canada, 2013.

[25] A. I. Mata, "Para o estudo da entoação em fala espontânea e preparada no Português Europeu," Ph.D. dissertation, University of Lisbon, 1999.

[26] A. I. Mata, H. Moniz, F. Batista, and J. Hirschberg, "Teenage and adult speech in school context: building and processing a corpus of European Portuguese," in *Proceedings of LREC 2014*, Reykjavik, Iceland, 2014.

[27] W. Labov, *Sociolinguistique*. Paris: Minuit, 1976.

[28] M. C. Viana, S. Frota, I. Falé, I. Mascarenhas, A. I. Mata, H. Moniz, and M. Vigário, "Towards a P_ToBI," in *Unpublished Workshop of the Transcription of Intonation in the Ibero-Romance Languages, PaPI 2007*, 2007.

[29] A. I. Mata, H. Moniz, T. Móia, A. Gonçalves, F. Silva, F. Batista, I. Duarte, F. Oliveira, and I. Falé, "Prosodic, syntactic, semantic guidelines for topic structures across domains and corpora," in *Proceedings of LREC 2014*, Reykjavik, Iceland, 2014.

[30] S. Frota, *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation*. New York: Garland Publishing, 2000.

[31] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 1986.

[32] S. Frota, "The intonational phonology of European Portuguese," in *Prosodic Typology II*, Sun-uh, Ed. Oxford: Oxford University Press, 2009.

[33] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in *Proc. of the DARPA Broadcast News Workshop*, Herndon, VA, Feb. 1999.