

Testing different discourse segmentation models with Portuguese data

This is an abstract of a full paper.

The present paper is set to describe one of the initial tasks of PRODIP, a project currently being developed at ILTEC, Lisbon. The main purpose of PRODIP is to investigate what prosodic features speakers use to mark the information structure of spoken discourse and which cues are most relevant for the listeners to identify this structure.

There have been no such studies concerning the Portuguese language. PRODIP can thus provide valuable information for computational linguistics, also allowing the comparison between Portuguese and other languages regarding macro-level prosody.

The data used in PRODIP have previously been collected for REDIP [20], a project that aims to study the language of Portuguese media, dealing mostly with radio and TV broadcasts. One of the reasons we are using the REDIP *corpus* is because it contains a large amount of spontaneous speech. The importance of using spontaneous speech in this kind of work has already been outlined in [1, 21].

Following the claims of several authors, PRODIP assumes that there is a relationship between discourse structure and prosodic features. Crucially, its goal is to explain how exactly that relationship holds in Portuguese.

It has been stated that if we want to identify the role of prosody in the structuring of information, we must compare it with an independently obtained discourse structure, in order to minimize the risks of circularity [21-25]. Previous work [4,19] on other languages has shown that there is no direct match between syntactic structure and prosodic constituency. Instead, prosody seems to be constrained by semantic and pragmatic aspects. Therefore, we should not rely on syntax for that matter, which would otherwise be the most immediate choice.

In order to have some sort of information structure against which prosody can be confronted, some authors elicit instruction monologues, a method which yields speech with a discourse structure determined *a priori* [22,23,25]. Others rely on discourse segmentations resulting from discourse analysis [6-13,17,18]. Both approaches thus assume that spoken discourse exhibits a structure similar to that of written texts, on what concerns the grouping of sentences into larger units like paragraphs, for instance.

PRODIP will not use the instruction monologues method, since we want to take advantage of the REDIP *corpus*, which is mainly composed of dialogues. So we opted for the second approach, which has the advantage of making it possible to study different speech styles as opposed to the instruction monologues method, with a very specific kind of data.

In order to obtain these discourse boundaries, we are now comparing two discourse segmentation models. We chose the models of Grosz and Sidner [7] and Litman and Passoneau [18], as these have been widely used and there is extensive research on them. Both models produce intention based segmentations, but the former generates a hierarchical structure while with the latter we get a linear one.

One of these two models will be chosen for the PRODIP research on prosody. This choice will be based on a test we are now conducting with the purpose of evaluating inter-coder agreement.

For this test, we have selected two excerpts from the REDIP *corpus*. These consist of interviews from the radio, and feature spontaneous speech, involving both male and female speakers. We have asked twenty subjects to annotate these two transcripts

according to the previously mentioned models. They all received an orthographic transcription of the selected texts, but only ten of them heard the original recordings. The subjects were also split into two different groups according to the model they were instructed to work with.

As we already said, the results from this task will be analysed statistically in order to find out which model displays the highest inter-coder agreement. Additionally, we want to check whether hearing the audio files had any effect on subject agreement. We will employ the kappa statistic, which recent work has considered to be the most well suited for this purpose [2,5]. We will also resort to other statistics that have been used in previous studies working with these discourse theories so that we can compare our results to theirs.

This test will first enable us to decide which of the considered models best suits our research, and whether it will be necessary to adapt it, considering the nature of the project.

In the next stage, it will be used to verify how the structures defined within each model relate to a specific prosodic feature: pauses. Subsequently, we will be analysing other prosodic variables, such as intensity, speech rate, pitch, etc. We plan on running perceptual tests as well. Obviously, one of our goals is also to see if the role of prosody in signaling large prosodic units in Portuguese differs from that in previously examined languages, namely English.

References:

- [1] Beckman, M. E. (1997) A Typology of Spontaneous Speech. In Y. Sagisaka, N. Campbell & N. Higuchi. *Computing Prosody: Computational Models for Processing Spontaneous Speech*. New York, Springer: 7-26.
- [2] Carletta, Jean (1996) Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics* 22 (2): 249-254.
- [3] Collier, R. (1993) On the Communicative Function of Prosody: Some Experiments. *IPO Annual Progress Report* 28: 67-75.
- [4] Cutler, Anne, Delphine Dahan & Wilma van Donselaar (1997) Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech* 40 (2): 141-201.
- [5] Flammia, Giovanni (1998) Discourse Segmentation of Spoken Dialogue: An Empirical Approach. Ph. D. thesis, MIT.
- [6] Grosz, Barbara & Julia Hirschberg (1992) Some Intentional Characteristics of Discourse Structure. *Proceeding of the International Conference on Spoken Language Processing*: 429-432.
- [7] Grosz, Barbara J. & Candace L. Sidner (1986) Attention, Intention and the Structure of Discourse. *Computational Linguistics* 12(3): 175-204.
- [8] Hirschberg, Julia & Barbara Grosz (1992) Intonational Features of Local and Global Discourse Structure. *Proceedings of the Workshop on Spoken Language Systems*: 441-446.
- [9] Hirschberg, Julia, Christine H. Nakatani & Barbara J. Grosz (1995) Conveying Discourse Structure through Intonation Variation. *Proceeding of the ESCA Workshop on Spoken Dialogue Systems: Theories and Applications*, Virgo, Denmark, ESCA.

- [10] Litman, Diane J. & Rebecca Passonneau (1993) Empirical Evidence for Intention-Based Discourse Segmentation. *Proc. of the ACL Workshop on Intentionality and Structure in Discourse Relations*.
- [11] Litman, Diane J. & Rebecca Passonneau (1995) Combining Multiple Knowledge Sources for Discourse Segmentation. *Proc. of 33rd ACL*: 108-115.
- [12] Nakatani, Christine H., Barbara J. Grosz & Julia Hirschberg (1995) Discourse Structure in Spoken Language: Studies on Speech Corpora. *Proceeding of the AAAI Symposium Series: Empirical Methods in Discourse Interpretation and Generation*.
- [13] Nakatani, Christine H., Barbara J. Grosz, David D. Ahn & Julia Hirschberg (1995) Instructions for Annotating Discourses. *Technical Report Number TR-21-95*. Center for Research in Computing Technology, Harvard University: Cambridge, MA.
- [14] Oliveira, Miguel (2002) Pausing Strategies as Means of Information Processing in Spontaneous Narratives. In: B. Bel & I. Marlien, *Proceedings of the 1st International Conference on Speech Prosody*. Ain-en-Provence, France, 539-542.
- [15] Oliveira, Miguel (2000) Prosodic Features in Spontaneous Narratives. Ph.D. thesis, Simon Fraser University.
- [16] Oliveira, Miguel (2002) The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. In: E. Ranchhod & N. Mamede (Eds.) *Advances in Natural Language Processing*. Third International Conference, PorTAL 2002. Faro, Portugal: Springer, 43-51.
- [17] Passonneau, Rebecca J. & Diane J. Litman (1993) Intention-Based Segmentation: Human Reliability and Correlation with Linguistic Cues. *Proc. of the ACL*.
- [18] Passonneau, Rebecca J. & Diane J. Litman (1995) Discourse Segmentation by Human and Automated Means. *Computational Linguistics*.
- [19] Pijper, Jan Roelof & Angelien A. Sanderman (1994) On the Perceptual Strength of Prosodic Boundaries and its Relation to Suprasegmental Cues. *Journal of the Acoustical Society of America* 96 (4): 2037-2047.
- [20] Ramilo, Maria Celeste & Tiago Freitas (2002) A Linguística e a Linguagem dos Média em Portugal: descrição do Projecto REDIP. Comunicação apresentada no XIII Congresso Internacional da ALFAL, San José, Costa Rica.
- [21] Swerts, M. & R. Collier (1992) On the Controlled Elicitation of Spontaneous Speech. *Speech Communication* 11 (4-5): 463-468.
- [22] Swerts, Marc & Ronald Geluykens (1993) The Prosody of Information Units in Spontaneous Monologue. *Phonetica* 50: 189-196.
- [23] Swerts, Marc & Ronald Geluykens (1994) Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech* 37 (1): 21-43.
- [24] Swerts, Marc (1997) Prosodic Features at Discourse Boundaries of Different Strength. *Journal of the Acoustical Society of America* 101 (1): 514-521.
- [25] Swerts, Marc, René Collier & Jacques Terken (1994) Prosodic Predictors of Discourse Finality in Spontaneous Monologues. *Speech Communication* 15: 79-90.