An empirical account of the relation between discourse structure and pauses in Portuguese

Eva Arim*, Francisco Costa*, Tiago Freitas*

The work reported in this paper is part of a project that aims at describing the role of prosody in conveying discourse structure in European Portuguese. We conducted an empirical study to investigate whether the organization of discourse could be reflected by the occurrence of pauses. In order to achieve that, we started by asking sixteen subjects to annotate two texts using two discourse segmentation methods: Grosz and Sidner's, and Passonneau and Litman's. We then confronted the segmentation obtained with pause distribution and length. We also verified which of the two methods can produce segmentation closer to pause occurrence. The results obtained are still provisional, but they at least show that the longest pauses are associated with the strongest discourse breaks.

1. Introduction

Written texts are presented in a way that enables readers to immediately identify how they are structured and organized. It is possible to recognize different discourse units by looking at graphical markings such as indentation, punctuation, the use of capital letters, etc. Spoken discourse displays a similar organization, but other cues are used to evince it. In this sense, we can also speak of paragraphs in oral texts, where prosody is considered to be an important factor in determining these units. Research done so far in several languages has achieved considerable success in showing that a relation between discourse structure and prosody does exist, but little work has been conducted with Portuguese data.

Pauses are one of the prosodic features generally considered relevant in marking the organization of spoken discourse (Swerts, 1997; Oliveira, 2002ab). This paper describes a pilot study on the relation between discourse organization and the occurrence of pauses in European 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, to minimize the risks of circularity (Brown, 1980). Previous work on other languages has shown that there is no direct match between syntactic structure and prosodic constituency (Cutler et al., 1997; Pijper & Sanderman, 1994). Instead, prosody seems to be constrained by semantic and pragmatic aspects. Therefore, if we want to investigate the role of certain prosodic variables in conveying structure, instead of comparing their distribution to syntactic representations, it can be more informative to compare them to empirically obtained discourse structure.

In order to obtain a representation of discourse structure independent of prosody, we will be using two discourse segmentation models. We chose the models of Grosz and Sidner (Grosz & Sidner, 1986; Hirschberg & Grosz, 1992; Nakatani, Grosz, Ahn & Hirschberg, 1995) and Passonneau and Litman (Passonneau & Litman, 1995), as these have been widely used and there is extensive research on them, enabling us to compare our results with those presented by other reseachers. Both models generate discourse segmentation using speaker intention as the main criterion. The difference is that while the former generates a hierarchical structure, the latter generates only a linear type of segmentation.

The study we are reporting in this paper is part of a wider project we are working on, the purpose of which is to identify the relationship between some prosodic elements and the structure of discourse in European Portuguese. It is mainly focused on spontaneous speech. There is little research done so far focusing on spontaneous data from Portuguese. The importance of using spontaneous speech in this kind of work has to do with the fact that it can be prosodically different from prepared or read speech. One of its possible applications is to make speech technology more natural sounding and more efficient in recognizing natural speech.

In the future, we will be using one of the two aforementioned models to produce discourse segmentation against

* Instituto de Linguística Teórica e Computacional – ILTEC, Lisbon

which we will compare the distribution of several prosodic features. This preliminary study will enable us to find out which of these models produces segmentation closer to the distribution of pauses. We will take that information into account when choosing the model for subsequent research.

2. Background

Previous research on this subject has showed that several factors are responsible for the occurrence of pauses in speech. O'Connell & Kowal (1983) mention anxiety, breathing, syntactic complexity, emphasis, interruption, etc. Other studies correlate the frequency of pauses with other variables, namely gender (Kowal et al., 1975), age and educational level (Sabin et al., 1979), socio-economical level (Bernstein, 1962). Several researchers have also observed that pause tends to be shorter and less frequent in reading aloud than in spontaneous speech (Barik, 1977; Goldman-Eisler, 1968; Grosjean & Deschamps, 1973; Megyesi & Gustafson-Čapková, 2001). It has also been noted that acoustic pauses accompany disruptive utterances (false starts, repetitions, filled pauses), as well as discourse markers and conjunctions.

As far as the relationship between discourse and pause is concerned, it has been stated that the longest pauses coincide with the boundaries of linguistic units (Goldman-Eisler, 1968), the pauses following a unit similar to a prosodic phrase are perceived as being longer than those associated with hesitations (Booemer & Dittman, 1962), and pause length is proportional to the level of independence between two adjacent discourse units in spontaneous monologues (Swerts, 1997). Several other studies identified a positive correlation between discourse structure and pause (e. g., Grosz & Hirschberg, 1992). On the other hand, Megyesi & Gustafson-Čapková (2001) report a low correlation between the acoustic pauses and the discourse structure of spontaneous dialogues, but found a significant one in texts read by non-professional readers.

3. Method

For the purpose of the analysis, we have selected two excerpts from a Portuguese corpus. These consist of interviews from the radio featuring spontaneous speech and contain both male and female speakers. Both excerpts were digitally recorded and average around one and a half minutes in length.

We have asked sixteen subjects to annotate these two transcripts according to the previously mentioned models. They all received an orthographic transcription of the selected texts, but only eight of them heard the original recordings. Since we hypothesize that there is a relation between discourse structure and prosody, we expected the listening and non-listening groups to display a different behavior. The subjects were also split into two different groups according to the model they were instructed to work with. None of the participants had any experience in this sort of task.

3.1. Acoustic Pauses

The structures obtained with the two mentioned models were analyzed in order to check whether and how they relate to the distribution of pauses. To this end, we measured the pauses that occurred in the aforementioned excerpts. This was done in Praat (Boersma, 2003) using a specific script created by Mietta Lennes (Lennes, 2002). Since the pause components of the voice onset time of many consonants were identified as pauses, we manually corrected the output of this procedure.

We only considered silent pauses, defined as the acoustical correlate for the perception of a silent portion in the speech signal, produced in conjunction, or not, with an inspiration, swallowing, any laryngo-phonatory reflex, or a silent expiration. Only silent pauses longer than or equal to 100 ms were taken into account. This seemed to be the lowest value we could work with in the sense that below that level there is systematic confusion between silent intervals and voice onset times. Filled pauses (pauses containing a voiced fragment in the speech signal, such as drawls, repetitions of utterances, words, syllables, sounds, false starts, hummings, etc.) were left out of this study.

There are some authors who do not consider filled pauses to be pauses at all, and split what they do recognize as pauses into two groups: complex pauses (pauses containing physiological phenomena, like breathing or swallowing), and silent pauses (Megyesi & Gustafson-Čapková, 2001). According to Grosjean and Collins (1979), complex pauses are the only ones occurring in fast speech. We believe this is a strong argument for considering both kinds of pauses, given that our

data consist of spontaneous speech. In fact, 58% of all non-filled pauses in our data were complex pauses.

4. Results and Discussion

4.1. Discourse segmentation models

We start by reporting the results obtained in a previous experiment we conducted in order to test the reliability of the two discourse segmentation models employed[1]. In order to compare a model that produces hierarchical segmentation to another one that only produces a linear one, we discarded all hierarchical information. We found out that the model which produced the highest level of consistency among coders was Passonneau & Litman's, achieving a kappa value[2] of 0.73. Among the four groups, the one that scored best was the one that used this model and listened to the audio recordings (kappa = 0.74). Table 1 below shows the relevant values for all conditions:

	Grosz & Sidner's Model	*Passonneau & Litman's Model*
Listening	kappa = 0.59	kappa = 0.74
Non-Listening	kappa = 0.68	kappa = 0.69
Overall	kappa = 0.65	kappa = 0.73

Table 1. Observed coder agreement

The different scores between the listening and the non-listening groups corroborate the hypothesis that discourse structure is reflected in prosody. In Litman & Passonneau's model the effect of hearing the speech shows up in a positive way, suggesting that prosody can make discourse structure more explicit. In Grosz and Sidner's model, access to prosodic information might have caused people to look for prosodic means of signaling hierarchy between segments, resulting in a more disparate segmentation.

4.2. Acoustic pauses

We now turn to our results regarding acoustic pauses. We identified 66 pauses within the two texts considered, which contain a total of 504 words. This results in approximately eight words per pause. The largest pause measured 1533 ms and the shortest 120 ms. Pause length averaged 416 ms. Surprisingly, we found that 14% of the total speech time corresponded to pauses. This is a very small number when compared to the ones reported by other studies: Brotherton (1979) found that silent pauses represent 25% of the total speech time, Johns-Lewis (1986) reports that pause ratio is larger (56%) in speech involving "reflective interpretation" than in public speeches or interviews (20%), while Sabin (1976) presents a pause ratio of 35% in narratives. This may be attributed to the fact that our data consisted of radio interviews, a context in which speakers can feel constrained to avoid long periods of silence in order to keep the audience tuned in, which also explains why the largest pause we found was shorter than 2 seconds.

In order to assess the relationship between discourse structure and the distribution of pauses, we followed Swerts (1997) and computed what he calls boundary strength, which is basically the percentage of subjects assigning a discourse boundary at a given possible boundary site That is, a discourse boundary recognized by all subjects is considered stronger (it is considered to have a boundary strength of 1) than a boundary identified by only 50% of the subjects (considered to have a boundary strength of 0.5). We then checked whether stronger boundaries were signaled by the occurrence of larger pauses. Our results are presented in Table 2 below.

We decided to break down pauses into five different categories according to their duration. This partition was based on data found in the literature: Glukhov (1975) found significant differences in pause frequency in several languages (among them Portuguese), but did not find any differences among pauses larger than 150 ms (suggesting that the 150 ms threshold can set apart two very different groups of pauses); Lehiste (1982) states that the shortest pauses associated to a paragraph boundary (a discourse segment boundary, in our terms) are 520 ms long; and Brown et al. (1980) consider *topic*

[1] See Arim et al. (to appear) for a full description of the study.
[2] See Carletta (1996) for some discussion on the reliability of this statistical method.

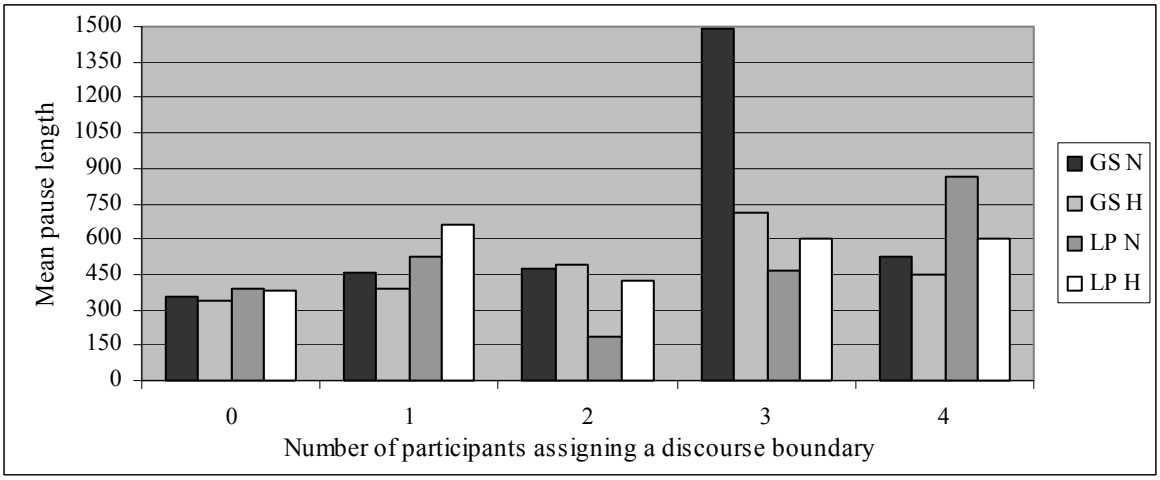
pauses to be 600-800 ms long.

	Grosz & Sidner's Model	*Passonneau & Litman's Model*
No pause	0.02	0.01
100-150 ms	0.03	0
150-520 ms	0.24	0.11
520-600 ms	0.14	0.05
600-800 ms	0.16	0.03
>800 ms	0.60	0.33

Table 2. Mean boundary strength values associated with different degrees of pause length.

What is striking about the data presented above is that the group of pauses ranging from 150 ms to 520 ms is associated with the second highest mean boundary strength. This can be explained in part by the fact that many of these pauses coincide with turn boundaries and that almost every discourse annotator placed discourse boundaries at turn shifts, and in part by the fact that these pauses are more frequent (see below for more details). Excluding this special case, pause length seems to increase together with boundary strength values for Grosz & Sidner's model, while Passonneau & Litman's scores worse in this respect.

Graph 1 below shows that pauses do not seem to have been a decisive factor in identifying discourse segments, since the results obtained for the coders that listened to the audio recordings did not differ from those of the non-hearing group[3].



Graph 1. Mean pause length (in milliseconds) associated with different degrees of boundary strength (expressed in the number of participants assigning a discourse boundary), broken down for the discourse model employed by the subjects (Grosz & Sidner's – GS – and Passonneau & Litman's – PL), as well as for the condition: hearing (H) and non-hearing (N).

As we mentioned before, pause length ranges from 120 ms to 1533 ms, spanning an interval of 1413 ms. The group of pauses ranging from 150 ms to 520 ms accounts for 26% of that interval, but includes 66% of all pauses. If we compare the same figures with the other pause categories, we will also find large differences (see Table 3 below), which shows that if we had grouped pauses in categories with uniform amplitudes, those groups would still be very heterogeneous with respect to the number of pauses falling within each of them.

[3] It should be noted that the highest column in this graph is the reflex of an outlier.

	Percentage of the amplitude of pause length	*Percentage of pauses occurring in pause group*	*Mean boundary strength across both models*
100-150 ms	2%	8%	0.01
150-520 ms	26%	66%	0.17
520-600 ms	6%	11%	0.10
600-800 ms	14%	6%	0.09
>800 ms	52%	9%	0.47

Table 3. Comparison between the amplitude of each group of pauses and the frequency of pauses belonging to that group, as well as the corresponding mean boundary strength across both models.

The results in Tables 2 and 3 do not indicate a proportional link between discourse structure and pause length, but they do point to the relevance of large pauses in signaling strong conceptual breaks in spontaneous speech, showing a clear link between pauses longer than 800 ms and the strongest boundaries between discourse segments. This may be a prosodic feature specific to conversational speech. While the strongest discourse boundaries are not always marked by large pauses (as can be seen in Graph 1), most of the largest pauses surface in places where a strong informational break can be identified. Once again, this can be ascribed to the fact that pauses did not seem to be a frequent means of echoing turn giving in our data, but turn boundaries were often also considered discourse segment boundaries by our coders.

5. Conclusions and Future Directions

We attribute the fuzziness of our results to the various factors responsible for the emergence of pauses, which we described in section 2. Looking at the places where pauses appear, we can see that they often co-occur with disfluencies, but it is difficult to determine precisely which pauses constitute disfluencies and which ones have a discursive function without risking manipulating the data (one cannot just state that every pause that does not coincide with a discourse break is a performance error). Additionally, it is often the case that a pause often follows a discourse structuring expression like *in fact*, whereas subjects annotating discourse generally place a boundary before that expression. Our data does not constitute a large body of evidence to support the relationship between a discourse marker and a following pause, but it has been claimed that this relation holds in other languages. Maybe some of the pauses that appeared not to correspond to a discourse segment boundary actually do, but they just do not surface at exactly the same spot. We cannot, of course, exclude the hypothesis that these results are specific to the type of speech we analyzed. On the other hand, these outcomes are consistent with the generalized notion that pause is not the prosodic element most indicative of discourse organization (Swerts et al., 1992; Cutler et al., 1997).

Since this study was only a pilot experiment, our corpus was not very large. One of the main purposes of this preliminary experiment was not only to investigate a few aspects of pause distribution, but also to gain some insight on the discourse segmentation model that can be more useful when we want to compare prosody and discourse. We limited our study to only two discourse theories because our end goal is not the evaluation of these theories but the factors governing discourse prosody. We will return to the study of pauses in the subsequent stages of our project, and at that time we will be using larger data samples and will not be concerned with evaluating theories, which will enable us to deepen our analysis. We also plan to process the results of a perceptual experiment on pauses we have already concluded, in order to examine the relation between acoustic pauses and perceived pauses and the one between discourse structure and perceived pauses.

References

Arim E., Costa F. & Freitas T., to appear, "Testing two discourse segmentation models with Portuguese data". XVII International Congress of Linguists, Prague, Czech Republic, July 2003.

Barik H. C., 1977, "Cross-linguistic study of temporal characteristics of different speech materials", *Language and Speech* 20, p. 116-126.

Boersma P., 2003, "Praat : doing phonetics by computer", http://www.fon.hum.uva.nl/praat/.

Boomer D. S. & Dittman A. T., 1962, "Hesitation pauses and juncture pauses in speech", *Language and Speech* 5, p. 215-220.

Brotherton P.,1979, "Speaking and not speaking; process for translating ideas into speech", in *Of Time and Speech*, Siegman A. W. & Feldstein S. (eds.), Hillsdale, New Jersey, Lawrence Erlbaum, p. 178-209.

Brown G., Currie K. & Kenworthy J., 1980, *Questions of Intonation*, London, Croom Helm.

Carletta J.,1996, "Assessing Agreement on Classification Tasks : The Kappa Statistic", *Computational Linguistics* 22 (2), p. 249-254.

Cutler A., Dahan D. & Donselaar W., 1997, "Prosody in the Comprehension of Spoken Language : A Literature Review", Language and Speech 40 (2), p. 141-201.

Glukhov A. A., 1975, "Statistical analysis of speech pauses for Romance and Germanic languages", *Soviet Physics. Acoustics* 21, p. 71-72.

Goldman-Eisler F., 1968, *Psycholinguistics: experiments in spontaneous speech*, London, New York, Academic Press.

Grosjean F. & Collins M., 1979, "Breathing, pausing and reading", *Phonetica 36 (2)*, p. 98-114.

Grosjean F. & Deschamps A., 1973, "Analyse des variables temporelles du français spontané. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radiophonique)", *Phonetica* 28, p. 191-226.

Grosjean F. & Deschamps A., 1975, "Analyse contrastive des variables temporelles de l'anglais et du français", *Phonetica 31*, p. 144-184.

Grosz B. & Hirschberg J., 1992, "Some Intentional Characteristics of Discourse Structure", *Proceeding of the International Conference on Spoken Language Processing*, p. 429-432.

Grosz B. J. & Sidner C. L., 1986, "Attention, Intention and the Structure of Discourse", Computational Linguistics 12(3), p. 175-204.

Hirschberg J. & Grosz B., 1992, "Intonational Features of Local and Global Discourse Structure", Proceedings of the Workshop on Spoken Language Systems, p. 441-446.

Johns-Lewis C. M., 1986, "Prosodic differentiation of discourse modes", in *Intonation in Discourse*, Johns-Lewis C. M. (ed.), London, Croomhelm, College-Hill, p. 199-219.

Kowal S., O'Connell D. C. & Sabin E. J., 1975, "Development of temporal patterning and vocal hesitation in spontaneous narratives", *Journal of Psycholinguistic Research 4*, p.195-207.

Lennes M., 2002, "Mietta's Praat scripts", http://www.helsinki.fi/~lennes/praat-scripts/.

Megyesi B. & Gustafson-Čapková S., 2001, "Pausing in Dialogues and Read Speech in Swedish : Speakers' Production and Listeners' Interpretation", paper presented at the Workshop on Prosody and Speech Recognition 2001.

Nakatani C. H., Grosz B. J., Ahn D. D. & Hirschberg J., 1995, "Instructions for Annotating Discouses", Technical Report Number TR-21-95, Center for Research in Computing Technology, Harvard University, Cambridge, MA.

O'Connell D. C. & Kowal S., 1983, "Pausology", *Computers in Language Research 2 (19)*, p. 221-301.

Oliveira M., 2002a, "Pausing Strategies as Means of Information Processing in Spontaneous Narratives", in *Proceedings of the 1st International Conference on Speech Prosody*, B. Bel & I. Marlien (ed.), Aix-en-Provence, France, p. 539-542.

Oliveira M., 2002b, "The Role of Pause Occurrence and Pause Duration in the Signalling of Narrative Structure", in *Advances in Natural Language Processing. Third International Conference, PorTAL*, E. Ranchhod & N. Mamede (eds.), p. 43-51.

Passonneau R. J. & Litman D. J., 1995, "Discourse Segmentation by Human and Automated Means", Computational Linguistics.

Pijper J. R. & Sanderman A. A., 1994, "On the Perceptual Strength of Prosodic Boundaries and its Relation to Suprasegmental Cues", Journal of the Acoustical Society of America 96 (4), p. 2037- 2047.

Sabin E. J., 1976, *Pause and Rate Phenomena in Adult Narratives*, Saint Louis, Saint Louis University.

Sabin E. J., Clemmer E. J., O'Connell D. C. & Kowal S., 1979, "A pausological approach to speech development", in *Of Time and Speech,* Siegman A. W. & Feldstein S. (eds.), Hillsdale, New Jersey, Lawrence Erlbaum.

Swerts M., 1997, "Prosodic Features at Discourse Boundaries of Different Strength", *Journal of the Acoustical Society of America 101 (1)* 514-521.

Swerts M., Geluykens R. & Terken J., 1992, "Prosodic correlates of discourse units in spontaneous speech", in Proceedings of the International Conference on Spoken Language Processing, Banff, p. 421-424

Zellner B., 1994, "Pauses and the temporal structure of speech", in Fundamentals of speech synthesis and speech recognition, E. Keller (ed.), p. 41-62.