



Prosody As Marker of Discourse Segmentation in Suyá

Miguel Oliveira, Jr.

School of Languages, Linguistics and Cultures
University of Manchester, United Kingdom
mfussmo2@manchester.ac.uk

Abstract

The present study investigates whether – as in several well-documented languages –, prosody plays a role in the signaling of discourse segmentation in Suyá, an Amazonian language of the Jê group. Inspired by the literature, the following prosodic variables were selected for analysis: pause, pitch reset and boundary tones.

1. Introduction

The degree of syntactic, semantic, and/or pragmatic cohesiveness between words in an utterance determines whether they belong together to a larger linguistic constituent or not. To the same extent, utterances bear different sorts of relations with other utterances in an even larger linguistic constituent that, when grouped together, form what is generally referred to as a “discourse”. In this view, discourse is considered to be a structure composed by hierarchically arranged entities that preserve a similar orientation. In written language, these entities are called “paragraphs.” They are often signaled by typographic means, such as an indent line at the beginning and an incomplete line at the end (which may be absent in cases where the end of the paragraph coincide with the end of the line). Spoken discourse also presents such macro-structures, which are referred to as “discourse segments” ([32]), “topics” ([37]), “information units” ([15]), and even “paragraphs” ([26]). These units are marked in speech by the use of different linguistic phenomena, such as anaphora, cue phrases, discourse markers, reference and tense.

One of the most important structuring, or demarcative devices in spoken discourse is prosody. Variation in pitch range ([6], [23], [37]), pausal duration ([15], [16], [27], among others), speech rate ([13], [35]), and amplitude ([6], [16], [23]) have all been studied, with some success, as potential correlates of discourse structure in speech.

Independent of any prosodic evidence, some discourse types (or genres) are considered to have an internal structure that can be observed solely by taking into account the content of their constituents. Procedural texts, for example, are thought to be composed of semantically independent segments (sections or units) that can be easily recognized because of their causal chain (those that form a continuous flow of cause and effect in the text). [1] and [2], for example, propose a grammar of procedural texts, which according to them are in general highly structured and modular, exhibiting a particularly rich micro-rhetorical structure integrated into the syntactic-semantic structures of instructions and preparatory actions. Since prosody is regularly used to segment “chunks of information” in discourse, as previous studies have demonstrated, it is expected that the semantically independent segments in procedural texts are bounded by

prosodic features too, making its structure transparent to the listener.

Suyá, a member of the Jê linguistic family, itself a member of the Macro-Jê family, is spoken by approximately 370 people in and near the Xingu Park of Central Brazil, being thus a highly endangered language. There are two subgroups of Jê, the Central Jê and the Northwest Jê. Suyá belongs to the Northwest Jê group, which also includes Apinayé, Kayapó, Kreen-Akarore, and Timbira. Suyá is one of the least-studied surviving Jê languages ([34])¹. This study is part of a large documentation project ([11]) that aims, among other things, at producing reference and pedagogical grammars, a dictionary, a community-approved orthography, and a large text collection².

2. Methods

The material used in this study consists of four elicited procedural texts, uttered by two native female speakers of Suyá. Texts PC_050326_08³ and PC_050326_09 were produced by a 55 years old woman and were about the making of *beiju* (a manioc bread) and of *perereba* (a drink made out of sweet manioc). Texts PC_050403_05 and PC_050403_06 were produced by a 27 years old woman and were about the making of *esteira de talo* (a kind of sieve used for, among other things, straining juice from manioc) and *khi* (cooked potatoes). The texts ranged from 50 sec. to 1 min. 54 sec., totaling 5 min. 21 sec. of recording. The recordings were made with a unidirectional, Shure headset microphone, connected to a TASCAM DAT recorder. A native speaker of the language who is a teacher in the main village of the Suyá people translated the texts later.

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 *risk of circularity* ([6], [37]). Previous work on other languages has shown that there is no direct match between syntactic structure and prosodic constituency - see [10] and [33]. 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.

¹ None of the languages in the Macro-Jê family is tonal.

² The Documentation and Description of Suyá project will also include studies of the ethnography of communication, prosody, and information structure in the language, providing a very extensive and intensive documentation of the language. See <http://ling.man.ac.uk/Info/staff/DE/DEHome.html> for more details.

³ The file name here corresponds to the original name of the files in the corpus. These files – as many others – will be made available in some major archiving databases soon.

In order to have some sort of information structure against which prosody can be confronted, some authors rely on discourse segmentations resulting from discourse analysis ([16], [19], [23] [27], [30], [32]). The problem with using the discourse analysis approach is that a priori we do not know whether it will yield more than an individual's intuition of discourse structure. If we are to depend on a discourse segmentation method, we must assure that we are employing one that is reproducible, because the more replicable a discourse segmentation model is, the stronger the evidence that discourse structure does exist.

For the purpose of the analysis presented here, we chose the model of [32], which produces linear intention based segmentation. This model has been widely used, which allows us to eventually compare our results with those obtained in work done for other languages. Besides, recent studies have demonstrated that this model displays a fair level of inter-coder consensus, which makes it more reliable and reproducible (see [3] for details)⁴.

All four procedural texts were firstly divided into intonation units⁵ and subsequently segmented in sequential chunks⁶, each representing a single *intention*, as proposed by the model of [32]. Discourse boundary placing was restricted thus to the end of an intonation unit. A total of 215 intonation units were devised, 38 of which were marked as ending a discourse unit. For each intonation unit a nucleus was designated as well as the boundary tone (low or non-low)⁷.

Three prosodic features were chosen to be analyzed: pause duration, boundary tone and pitch reset. Following the works by [21] and [22], pause here was regarded as a period of silence equal or greater than 100 ms. Pitch reset was calculated as the difference between the pitch range values of two adjacent intonation units⁸.

⁴ Following [32], we have asked eight naïve coders to annotate the transcripts using the speaker's intention as criterion. A high inter-coder agreement value was found ($kappa > 0.7$), what validated the use of the model.

⁵ Following [8] and [9], an intonation unit was regarded as a unit composed of at least one prominent syllable with a major pitch movement (the nucleus). The assignment of intonation-group boundaries was made thus *a priori* by taking into account phonetic cues. In some difficult cases, semantic and grammatical cues were also considered.

⁶ Based on the segmentation made by the naïve labelers.

⁷ This classification was inspired by the problems reported in [6], [15], [37], and [38]) regarding the reliability in the distinction of "high" from "mid" tones. "Non-low" tones in the present study covers both "high" and "mid" tones, thus.

⁸ Pitch range here was measured as the value of the fundamental frequency maximum for the intonation unit. This value was extracted from within the vowel of the syllable containing the fundamental frequency peak of the intonation unit ([16], [23], [28], [30]). Pitch values in the signals were extracted automatically, using Praat's ([5]) default fundamental frequency extraction algorithm. The original pitch contours were then stylized by hand, in a semi-automatic process that used both visual and auditive cues. This was done to avoid the interference of octave jumps and to smooth the contours ([31], [38]). Fundamental frequency peak values in the signal contours could be taken automatically from the program's information window. In the measurement, fundamental frequency values, expressed in semitones, were rounded up. The phenomenon of intrinsic pitch was taken into

3. Hypothesis

It is hypothesized that discourse boundaries are prosodically different from boundaries occurring within them. This can be verified, specifically, through:

- Pause: longer pauses tend to occur more often at the end of discourse boundaries than elsewhere;
- Pitch reset: pitch reset is higher at the end of discourse boundaries than elsewhere; and
- Boundary tones: low tones tend to occur at the end of discourse boundaries, whereas non-low tones are generally found inside discourse units.

This hypothesis is aimed at relating the three prosodic variables to different types of boundary in a procedural text: boundaries that coincide with the end of sections and boundaries that occur within those sections. Support for this hypothesis comes from studies of prosodic correlates of discourse structure (see references in Section 1 above).

4. Results

4.1. Pauses

It is commonly accepted in any intonation analysis that pause functions primarily to encode intonation unit boundaries ([6], [7]). The role of pauses in discourse has more far reaching effects: on a global level, pauses assume a more particular function in the organization of procedural texts.

A number of studies confirm the general hypothesis concerning pause duration as an indicator of boundaries between larger discourse units: the longer a pause is, the greater the chances that the place where it occurs coincides with a major discursive break (see, for example, [14], [16], [24], [32], among others)⁹. In Suyá, this also seems to be the case, as table 1 below demonstrates:

Table 1: *Mean pause duration at the end of a major discourse unit (YES) and elsewhere (NO), expressed in milliseconds.*

Discourse boundary	Pause duration
YES	871 msec.
NO	641 msec.

The mean duration of pauses occurring at the end of discourse units differs significantly from those occurring elsewhere. T-test result ($t=-2.909$, $df=113$, $p<0.005$) confirms this finding by showing a significant correlation between discourse boundary and longer pause duration.

consideration at this time. Furthermore, pitch values that were considered to be unreliable due to the presence of laryngealization or minimal intensity were disregarded from the final computation.

⁹ Although pause duration may be associated with an assortment of uses, such as breathing, marking emphatic stress, hesitating, reflecting situational and dispositional anxiety, planning, etc., any attempt to separate out these operations has always proved to be difficult (see, for example, [18]). As [17] points out, silent pauses are the product of a number of simultaneous operations.

4.2. Boundary tone

Boundary-marking pitch movements have often been associated with discourse segmentation. In general, melodic contours are said to distinguish between continuation and finality, with rise and level tones indicating the former, and low tones signaling the later ([4], [6], [15], [38], [39]). The results in the present investigation suggest that the same strategy is employed in Suyá. Table 2 below displays the difference (in percentages) in the distribution of low and non-low boundary tones between discourse and non-discourse boundaries. From these results, it may be deduced that discourse boundaries are most of the time (66% of all cases) accompanied by low boundary tones; non-discourse boundaries, on the other hand, systematically end in non-low boundary tone (75% of all cases). This difference is significant ($\chi^2=21.810$, $df=213$, $p<0.001$).

Table 2: *Distribution of boundary tone (low & non-low) as a function of boundary types (end of a major discourse unit, YES & elsewhere, NO) Values are relative to the total amount of occurrences.*

Discourse boundary	Boundary tone	
	Low	Non-low
YES	66%	34%
NO	25%	75%

4.3. Pitch reset

It has been suggested that the melodic discontinuity that occurs between information units - a consequence of the natural declination of pitch in the course of an utterance - is an important cue for discourse segmentation ([16], [20], [23], [29], [33], [37]). Table 3 below indicates a very clear association of pitch reset value and boundary type in Suyá:

Table 3: *Mean pitch reset values at the end of a major discourse unit (YES) and elsewhere (NO) Values are expressed in semitones.*

Discourse boundary	Pitch reset (semitones)
YES	3.26
NO	1.98

Higher pitch reset values are clearly associated with narrative boundaries, as illustrated in the chart above. Statistical analyses yielded significant results ($t=-2.975$, $df=213$, $p<0.0032$). Therefore, pitch reset appears to be a very accurate indicator of discourse section boundaries in procedural texts in Suyá.

5. Discussion

The present paper investigated whether Suyá, an Amazonian language of the Jê group, makes use of prosody as a structural device in discourse. An empirical model based on the speaker's intention was used to verify whether pause, boundary tone, and pitch reset is systematically employed in procedural discourse as cues to its internal structure.

The resulting analysis indicated that pausing is consistently used in Suyá as a means of segmenting "chunks

of information" in procedural texts. Specifically, pauses are consistently longer at the end of discourse units than within these units, suggesting thus that it is employed to make the segmentation of procedural texts into "sections" characterized by a semantic consistency transparent to the audience.

It was also found that the difference of pitch range values of adjacent intonation units (pitch reset) is a very significant cue to discourse segmentation in the language — discourse units tend to be separated from one another by means of a higher pitch reset value.

Furthermore, boundary-marking pitch movements were also found to be associated with discourse structure, because they often indicate whether a given unit has come to an end or not. Low boundary tones are regularly used to mark the end of a unit.

These are not new findings when it comes to the study of prosodic features in discourse, but they are nevertheless of interest for at least an extremely important reason: very little is still known about the organization of discourse in Amazonian languages, especially on whether prosody is employed as a structuring device in any of them — as it happens in many other (European) languages. It would be obviously naïve to assume that Suyá would behave like the European languages for which discourse characteristics in general and prosodic features in particular have been studied.¹⁰ In that sense, this preliminary study stands as an important contribution to the debate. After all, only with a large body of research on different aspects of prosody and discourse carried out in different communities can one test the validity of the existing findings. This would consequently provide a deeper understanding of the various mechanisms of discourse in different languages and cultures.

6. Conclusions

The present experimental investigation is part of a major project that aims at documenting and describing a dying Amazonian language, Suyá. It demonstrated that speakers of this Northwestern Jê language systematically use prosody as a demarcative device in discourse, segmenting larger chunks of information with longer pauses, low boundary tone and higher pitch reset. This suggests that they seem to be aware of the existence of an underlying discourse structure, and that this awareness is evidenced through the regular use of prosodic features. Other prosodic features, such as intensity, final vowel lengthening, and speech rate need also to be accounted for in the future. Prosody, however, is not the only cue that can be used in the assignment of discourse structure. The interplay of an array of other linguistic and non-linguistic cues — such as syntactic, pragmatic, gestural, and visual cues — is what causes the audience to infer the intended global structure of a given text. It would be thus interesting to investigate the interaction of prosody with some other contextualizing cues in the process of demarcating discourse structure in Suyá, as for example, the relation between prosodic and gestural cues — something that they seem to exploit a lot.

¹⁰ See the works by Everett ([12], for example), which deals with the idiosyncrasies of some Amazonian languages.

7. References

- [1] Aouladomar, F., Saint-Dizier, P., 2005. Towards generating procedural texts: an exploration of their rhetorical and argumentative structure, *EWNLG*, Aberdeen.
- [2] Aouladomar, F., 2005. A semantic analysis of instructional texts. *IWCS05*. Tilburg.
- [3] Arim, E., Costa, F. & Freitas, T. 2003. A study on the reliability of two discourse segmentation models. *Actas do VI Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada* (PROPOR 2003), Faro, Portugal.
- [4] Blaauw, E. 1995. On the perceptual classification of spontaneous and read speech. *Research Institute for Language and Speech*, Utrecht University.
- [5] Boersma, P. 2005. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.
- [6] Brown, G., Currie, K. & Kenworthy, J. 1980. *Questions of Intonation*. London, Croom Helm.
- [7] Chafe, W. 1989. Some things that narratives tell us about the mind. In B. B. A. Pellegrini. *Narrative Thought and Narrative Language*. Hillsdale, Lawrence Erlbaum.
- [8] Cruttenden, A. 1997. *Intonation*. Cambridge, Cambridge University Press.
- [9] Crystal, D. 1969. *Prosodic Systems and Intonation in English*. London, Cambridge University Press.
- [10] Cutler, A., D. Dahan and W. Donselaar. 1994. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40 (2):141-201
- [11] Everett, D. 2004. *Documentation and description of Suyá (Jê)*, project funded by the Economics and Social Research Council and the Arts and Humanities Research Board.
- [12] Everett, D. 2005. Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human. *Current Anthropology* 46 (4): 621-646.
- [13] Fon, J. 1999. Speech rate as a reflection of variance and invariance in conceptual planning in storytelling. *Proceeding of the ICPHS*.
- [14] Gee, J. P. & Grosjean, F. 1984. Empirical evidence for narrative structure. *Cognitive Science* 8: 59-85.
- [15] Geluykens, R. & Swerts, M. 1994. Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication* 15: 69-77.
- [16] Grosz, B. & Hirschberg, J. 1992. Some intonational characteristics of discourse structure. *Proceeding of the International Conference on Spoken Language Processing*, Banff.
- [17] Grosjean, F. (1980). Temporal variables within and between languages. In H. Dechert & M. Raupach. *Towards a Cross-Linguistic Assessment of Speech Production*. Frankfurt, Lang.
- [18] Grosjean, F. & Collins, M. (1979). Breathing, Pausing and Reading. *Phonetica* 36: 98-114.
- [19] Grosz, B. J. and C. L. Sidner. 1986. Attention, intention and the structure of discourse. *Computational Linguistics* 12(3): 175-204.
- [20] Hakoda, K. & Sato, H. 1980. *Prosodic rules in connected speech synthesis*, Translation of the Institute of Electronics and Communication Engineers 63-D.
- [21] Henderson, A., Goldman-Eisler, F. & Skarbeck, A. 1965. Temporal patterns of cognitive activity and breath control in speech. *Language and Speech* 8: 236-242.
- [22] Hieke, A. E., Kowal, S. & O'Connell, D. C. 1983. The trouble with "articulatory pauses". *Language and Speech* 26: 203-214.
- [23] Hirschberg, J. & Grosz, B. 1992. Intonation features of local and global discourse structure. *Proceeding of the DARPA Workshop on Spoken Language Systems*, Arden House.
- [24] Kowal, S. & O'Connell, D. C. 1987. Some temporal aspects of stories told while or after watching a film. *Bulletin of the Psychonomic Society* 25(5): 364-366.
- [25] Kowal, S., Wiese, S. & O'Connell, D. 1983. The use of time in storytelling. *Language and Speech* 26(4): 377-392.
- [26] Lehiste, I. 1975. The phonetic structure of paragraphs. In A. Cohen & S. Nootboom. *Structure and Process in Speech Perception*. Berlin, Springer-Verlag: 195-206.
- [27] Litman, D. J. & Passonneau, R. J. 1995. Combining Multiple Knowledge Sources for Discourse Segmentation. *Proceeding of the Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95)*, Cambridge, MA.
- [28] Menn, L. & Boyce, S. 1982. Fundamental frequency and discourse structure. *Language and Speech* 25(4): 341-379.
- [29] Nakajima, S. & Allen, J. F. 1992. Prosody as a cue for discourse structure. *Proceeding of the International Conference on Spoken Language Processing*, Banff, Canada.
- [30] Nakatani, C. H. & Hirschberg, J. 1995. Discourse structure in spoken language: Studies on speech corpora. *Proceeding of the AAAI Symposium Series: Empirical Methods in Discourse Interpretation and Generation*.
- [31] Nootboom, S. G. & Kruyt, J. G. 1987. Accents, focus distribution, and the perceived distribution of given and new information: an experiment. *Journal of the Acoustical Society of America* 82(5): 1512-1524.
- [32] Passonneau, R. J. & Litman, D. J. 1993. Intention-based segmentation: Human reliability and correlation with linguistic cues. *Proceeding of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-93)*, Columbus, Ohio.
- [33] Pijper, J. R. and A. 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.
- [34] Rodrigues, A. 1999. "Macro-Jê". In Dixon, R.M.W.; Aikhenvald, A.Y. eds. *The Amazonian Languages*. Cambridge: Cambridge University Press.
- [35] Selting, M. 1992. Prosody in conversational questions. *Journal of Pragmatics* 17: 315-345.
- [36] Siegman, A. & Feldestein, S. (eds.). 1979. *Of Time and Speech*. Lawrence Erlbaum, Hillsdale, N. J.
- [37] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America* 101(1): 514-521.
- [38] Van Donzel, M. 1999. *Prosodic Aspects of Information Structure in Discourse*. Faculteit der Geesteswetenschappen. Amsterdam, University van Amsterdam: 195.
- [39] Wichman, A., House, J. & Rietveld, T. 1997. Peak displacement and topic structure. *Proceeding of the ESCA Tutorial and Research Workshop on Intonation*, Athens.