Prosody As Marker of Direct Reported Speech Boundary

Miguel Oliveira, Jr. & Dóris A. C. Cunha

Departmento de Letras
Universidade Federal de Pernambuco, Brasil
mjaoj@ig.com.br & dorisarruda@terra.com.br

Abstract

The present paper aims at analyzing the role of prosody as a marker of direct reported speech boundaries in discourse. The beginning of a citation in speech is often linguistically marked, generally by means of a verb of saying. However, it is not always a straightforward task to determine at what point exactly a citation ends. Through the analysis of a series of excerpts extracted from spontaneous interviews, we investigate to what extent prosody functions as a cue for the delimitation of a direct citation in speech.

1. Introduction

According to [1], "the transmission and assessment of the speech of others, the discourse of another, is one of the most widespread and fundamental topics of human speech. In all areas of life and ideological activity, our speech is filled with overflowing with other people's words, which are transmitted with highly varied degrees of accuracy and impartiality." He goes as far as to propose that "in real life people talk most of all about what others talk about," implying thus that we are often replicating the discourse of others.

Notwithstanding, people frequently indicate whether what they are saying constitutes their own speech or the discourse of another. The most realistic representation of the presence of other voices in one's speech is achieved by means of what is generally referred to as "direct reported speech", defined by [11] as "the recorded broadcast of utterances previously pronounced by identified enunciators".

In direct reported speech, speakers attempt to reproduce utterances in a way that the original co- and context are brought to the conversational setting ([2]). In doing so, however, there occur transformations and functionalizations, in relation to the original context, that typically, meant by the speakers as he/she positions him/herself in the conversational context. Graphically, direct reported speech is represented by means of quotation marks. In speech, prosody and voice quality are used to designate this function ([9], [14]).

Traditionally, studies on reported speech have focused on grammatical aspects of this type of linguistic representation. However, grammatical approaches are not enough to deal with several questions regarding the structure of citation, especially when it comes to non-literary texts ([10]). So, for instance, as it is sometimes quite straightforward to specify the beginning of a citation in speech, signaled by the use of a verb of saying, or any other manifest linguistic indication, it is not always clearly identifiable where exactly the quotation ends. Consider the following example:

Example
so I always say to them... "if you want to be psychologists... be sure to become good liberal workers" because in Recife... wherever you go... there is a psychologist... a drug store... and a bank... these are the things we often see in Recife[1]

Although the transcription clearly indicates the end of the reported speech sequence, nothing in the excerpt suggests that what comes next does not belong to the citation itself. The decision of the boundary delimiting the reported speech and the discourse of the enunciator was made based on elements outside the textual level. We hypothesize that prosodic information plays a very important role on such decision-making.

The literature points out that one of the most important demarcative devices in spoken discourse is prosody. Variation in pitch range ([6], [16], [30], [32], among others), pausal duration ([34], [13], [8], [21], among others), speech rate ([20], [19], [12], [29]), and amplitude ([6], [17], [16], [13]) have all been studied, with some success, as potential correlates of discourse structure in speech.

The purpose of the present study is thus to investigate whether any correlation between the indication of the end of a reported speech sequence, as marked in the transcription of some interviews recorded by Project NURC[2], and a number of prosodic features - such as pause, difference in intensity, boundary tone and pitch reset - exist, which would substantiate the hypothesis that prosodic information is often essential in clarifying ambiguous structures in speech.

The importance of a study like this is undeniable - besides providing crucial information with respect to the way citation is represented in speech and on the specific cues that listeners take into account as a demarcative device for this particular case, it may contribute to the improvement of a number of speech technology systems, such as Automatic Speech Recognition and Automatic Transcription and Dialogue Understanding systems.

---

[1] Excerpt from Interview 078, Project NURC/RE, as translated from the original, as published in [28]. For detailed information regarding Project NURC and its material, see [28].

[2] The transcriptions of the material recorded by Project NURC is generally made by several people, and is systematically reviewed before publication. To use a data that is already transcribed and, therefore, provides an independent segmentation of its own, is very important methodologically, because then we are avoiding the so-called risk of circularity. It has been repeatedly claimed that one of the major problems in the study of prosodic correlates of discourse structure is the risk of circularity that investigations of this type often incur ([6], [32], [34]). This is due to the fact that the segmentation of a discourse is not generally uncontroversial, and most of the time prosody is used as a criterion for establishing its structure, which makes the reason for this investigation its own end.

Although a number of studies have already proposed a correlation of prosodic phenomena and reported speech ([3], [9], [14], [18]), no attempt has been made so far, to the best of our knowledge, to analyze prosody as a marker of reported speech boundaries in spoken discourse.

2. Methods

The material used in this study is drawn/collected from dialogues recorded by Project NURC[3]. They consist of nine excerpts, totalizing 5.7 minutes of recording, in which the problem, as exemplified above, is present. These excerpts were given to three experts in Brazilian Portuguese prosody, who had access to both the transcriptions and the digital audio files of all the excerpts. The experts were instructed to divide the fragments into intonation units and to indicate the type of boundary tone (low or non-low)[4] at the end of the last intonation unit in the excerpts[5]. A total of 232 intonation units were devised, 110 of which were labeled in the original transcriptions as reported speech.

In order to conduct acoustical analyses, the speech files were digitized at a rate of 22.05 KHz with 16-bit resolution using the speech-editing software Sound Studio, version 2.1 (Felt Tip Software). The data was subsequently analyzed in the speech-editing program Praat, version 4.1.5 ([5]).

Pitch values in the signals were extracted automatically, using the default fundamental frequency extraction algorithm in the program. The original pitch contours were then stylized by hand, in a semi-automatic process that used both visual and auditive cues. This process was intended to avoid the interference of octave jumps and to smooth the contours ([25], [31], [35]). Fundamental frequency peak values in the signal contours could be taken automatically from the program's information window.

Pitch reset was calculated as the difference between the pitch range values of two adjacent intonation units. Pitch range was considered to be the value of the fundamental frequency maximum for the intonation unit. This value is extracted from the vowel of the syllable containing the fundamental frequency peak of the intonation unit ([22], [24], [13], [16], [17]).

Pauses were measured directly in the speech signals. A cut-off point of 100 ms was adopted for the present study. Filled pauses were disregarded.

Intensity was measured in each intonation unit using Praat's native algorithm. Difference in intensity among intonation units, expressed in decibels, was calculated afterwards.

[3] Interviews numbers 078 (female speaker), 145, 216 and 266 (male speakers), NURC/RE. All these interviews, except for the last one, are published in [28].

[4] This classification was inspired by the problems reported in [6], [32], [34] and [35] 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. In order for a boundary to be considered as "low" or "non-low" in the present work, two out of the three experts had to agree in their judgment. In general, the judgments were very consistent from experts to expert. In fact, most boundary tones were classified as either "low" or "non-low" unanimously.

[5] All the excerpts are composed of more than one intonation unit. In this study, only the last intonation unit of the excerpts were taken into account for both the acoustical and the perceptual analyses.

3. Results

3.1. Pause

Pauses that occur at the end of a reported speech sequence are, in general, longer than those occurring elsewhere, as Table 1 below demonstrates.

Table 1: *Mean pause duration at the end of reported speech sequences (End of RS) and elsewhere, expressed in milliseconds.*

Boundary Type	Pause (msec)
End of RS	325
Elsewhere	244

Although a trend indicating that pause duration varies according to the type of boundary in which it occurs, statistical analysis do not show a significant effect for this variable.

3.2. Boundary tone

Of the 232 intonation units that make up the corpus of the present study, only 45 (i.e., 30% of the total) were labeled as ending in a low intonation boundary. As shown in Table 2 below, their distribution in the fragments suggests that, although the type of boundary doesn't seem to be a recurrent prosodic feature used for the classification of a boundary type, low boundary tones are much more often used at the end of a reported speech segment than elsewhere (38% versus 18%).

Table 2: *Distribution of boundary tone types (low & non-low) as a function of boundary types (end of a reported speech sequence & elsewhere) Values are relative to the total amount of occurrences.*

Boundary Type	Boundary Tone Type	
	Low	Non-Low
End of RS	38%	62%
Elsewhere	18%	82%

3.3. Pitch reset

Table 3 below indicates a very clear association of pitch reset value and boundary type: higher values correspond to the end of reported speech segments. Statistical analyses yield significant results ($t=3{,}317$, $df=215$, $p<0.001$).

Table 3: *Mean pitch reset values at the end of reported speech sequences (End of RS) and elsewhere, expressed in semitones.*

Boundary Type	Pitch Reset (semitones)
End of RS	9,03
Elsewhere	4,94

3.4. Intensity

The analysis of difference in intensity among intonation units reveals that there is a correlation of this measure and the type of boundary: higher values correspond to the one ending reported speech segments. Statistical analyses yield a slightly significant effect (t=1,779, df=220, p<0.07).

Table 4: *Mean values of difference in intensity at the end of reported speech sequences (End of RS) and elsewhere, expressed in decibels.*

Boundary Type	Intensity (decibels)
End of RS	2,43
Elsewhere	1,86

4. Discussion

A correlation of some prosodic features with the delimitation of direct reported speech sequences is unquestionable, as the numbers reported above clearly indicate. Although in some cases, statistical tests do not show significant results, a trend in the expected direction is verified.

Pauses that occur at places identified as being a reported speech boundary are almost 100ms longer than pauses that are found elsewhere in the discourse fragments that were analyzed. This is a significant duration, if one considers that 100ms is the cut-off point for the classification of a period of silence as a pause. However, pause duration *per se* doesn't seem to be enough as a marker of direct reported speech boundary.

Empirical research on the role of pause as a demarcative device has shown that pause duration is a much stronger boundary marker in larger clusters of information, such as "paragraphs" ([20]), and "narrative sections" ([26]). Since many of the citations that were analyzed here appear as part of larger-scale information units, the statistically non-significant effect may be thus justified.

In their studies on the acoustic-prosodic characteristics of discourse structure, [13], [16] and [17] found a significant correlation between quoted phrases and low intensity, if compared with other phrases. Specifically, they found that quote-final phrases were produced with a pronounced drop in intensity compared with other utterance-final phrases. This finding suggests that the boundary between quoted phrases and other phrases is characterized by a substantial difference in intensity. As the results reported above clearly indicates, a trend in this direction does exist, even though the statistical analysis do not show a significant result.

When it comes to the intonational parameters that are taken into consideration here, it is shown that a direct correlation between boundary tone and the signaling of the end of a reported speech sequence could not be found. The role boundary tones plays in discourse as a signal of topic continuity or finality has been the focus of investigation in several studies dealing with the prosodic means of indicating discourse boundary ([4], [33], [35]). [6], for example, observe that low boundary tones are often associated with the end of a topic, while non-low boundary tones regularly suggest that there is more to come on the same topic. This intonational feature seems to be a much more effective marker of topic continuation, which would justify the non-significant results here.

On the other hand, results from a statistical test demonstrate that among all prosodic features available to the listener as possible markers of direct reported speech boundary, pitch reset is the most significant. The boundaries that transcribers indicated as the end of a citation present higher pitch reset values in the fragments analyzed, suggesting that this prosodic feature plays an important role in the classification of a direct reported speech boundary where such an indication would be otherwise problematic.

A number of studies suggest 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 ([27], [32], [23], [15], [16], [13]). [18] demonstrated that overall pitch range distinguishes direct quotes from indirect quotes. The former present a greater pitch range than the latter. [3] found in their corpus a statistically significant correlation of F0 parameters in the characterization of three types of discourse: direct speech, direct reported speech containing self-quotations, and direct reported speech containing other virtual enunciators than the source speakers. Direct reported speech containing self-quotations presents, in general, a wider pitch range, if compared to direct speech. According to [3], this can be interpreted as a means used by the speaker in order to make his/her own reported speech more salient. The results in both studies therefore support the hypothesis that more expanded pitch reset characterizes direct quotation boundaries. In fact, [18] finds considerable difference in the amount of pitch reset relative to a preceding phrase. They do not consider the difference at the end of a reported speech sequence, but since pitch range values do differ according to the different discourse functions, it would be expected that the same would hold true for ending sequences as well. Of course what makes the findings reported here distinct from that reported in [18] is not the fact that pitch range reset characterizes direct quotation boundaries, but that hearers apparently perceive this as a significant cue.

5. Conclusion

Our experiment corroborates the findings reported elsewhere, which claim that reported speech is markedintonationally: we find statistically significant correlation between the end of direct reported speech sequences, as indicated by transcribers, and greater pitch reset values, suggesting that quotations are uttered in a different pitch range.

The results derived from this study are useful for applications in Automatic Speech Recognition and Automatic Transcription and Dialogue Understanding systems, as noted above. In order for such systems to produce more natural sounding speech output, conveying at the same time essential discourse information to the users and to interpret reported speech appropriately, information on how discourse contexts and dialogues acts are prosodically encoded is essential ([18]).

The present paper reports only a particular aspect of the general study on the prosody of reported speech that we are currently conducting at UFPE. More data will be added, which may corroborate the results presented here and may eventually lend legitimacy to the trends that noted at this time, but could not be statistically validated.

6. References

[1] Bakhtin, M. M., 1981. Discourse in the novel. In: M. Holquist, ed. *The dialogic imagination*, 259-422. Austin, TX: University of Texas Press.

[2] Bauman, R. & Briggs, C. L., 1990. Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19, 59-88.

[3] Bertrand, R. & Espesser, R. 2002. Voice Diversity in Conversation: a Case Study. In: B. Bel & I. Marlien, *Proceedings of the 1st International Conference on Speech Prosody*. Ain-en-Provence, France.

[4] Blaauw, E., 1995. *On the perceptual classification of spontaneous and read speech*. Research Institute for Language and Speech, Utrecht University.

[5] Boersma, P., 2003. *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] Bryant, G. & Fox Tree, J. E., 2002. Recognizing Verbal Irony in Spontaneous Speech. *Metaphor and Symbol* 17(2), 99-117.

[8] Collier, R., Piyper, J. R. d. & Sanderman, A., 1993. Perceived prosodic boundaries and their phonetic correlates. *Proceeding of the ARPA Workshop on Human Language Technology*, Plainsboro, New Jersey, USA, Morgan Kaufman Publishers.

[9] Couper-Kuhlen, E., 1998. Coherent Voicing. On Prosody in Conversational Reported Speech. *InLiSt* 1, 1-28.

[10] Cunha, D. A. C., 2001. Atividades sobre os usos ou exercícios gramaticais formais? O tratamento do discurso reportado". In Dionisio, A. e Bezerra, M. A (Ed.) *O livro didático de língua portuguesa: múltiplos olhares*. Rio de Janeiro, Lucerna.

[11] De Gaulmyn, M. M., 1992. Grammaire du français parlé. Quelques remarques autour du discours rapporté, in Actes du Congrès de l'ANEFLE *Grammaire et français langue étrangère*. Joussaud & Petrissans (dir.), Grenoble, ANEFLE, 22-23.

[12] Fon, J., 1999. Speech rate as a reflection of variance and invariance in conceptual planning in storytelling. *Proceeding of the ICPhS.*

[13] Grosz, B. & Hirschberg, J., 1992. Some intonational characteristics of discourse structure. Proceeding of the International *Conference on Spoken Language Processing*, Banff.

[14] Güntner, S., 1999. Poliphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics* 31, 685-708.

[15] Hakoda, K. & Sato, H., 1980. Prosodic rules in connected speech synthesis, Translation of the Institute of *Electronics and Communication Engineers* 63-D.

[16] Hirschberg, J. & Grosz, B., 1992. Intonation features of local and global discourse structure. *Proceeding of the DARPA Workshop on Spoken Language Systems*, Arden House.

[17] Hirschberg, J., Nakatani, C. H. & Grosz, B. J., 1995. Conveying discourse structure through intonation variation. *Proceeding of the ESCA Workshop on Spoken Dialogue Systems: Theories and Applications*, Visgo, Denmark, ESCA.

[18] Jansen,W., Gregory, M. L. & Brenier, J. M., 2001. Prosodic correlates of directly reported speech: Evidence from conversational speech. *Prosody in Speech Recognition and Understanding.* Molly Pitcher Inn, Red Bank, NJ, USA.

[19] Koopmans-van Beinum, F. J. & Van Donzel, M. E., 1996. Discourse structure and its influence on local speech rate. *Proceeding of the International Conference on Spoken Language Processing*, Philadelphia.

[20] Lehiste, I., 1982. Some phonetic characteristics of discourse. *Studia Linguistica* 36(2), 117-130.

[21] Litman, D. J. & Passonneau, R. J., 1995. Combining Multiple Knowledge Sources for Discourse Segmentation. *Proceeding of the 33rd Annual Meeting of the Association for Computational Linguistics* (ACL-95), Cambridge, MA.

[22] Menn, L. & Boyce, S., 1982. Fundamental frequency and discourse structure. *Language and Speech* 25(4), 341-379.

[23] Nakajima, S. & Allen, J. F., 1992. Prosody as a cue for discourse structure. *Proceeding of the International Conference on Spoken Language Processing*, Banff, Canada.

[24] 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.*

[25] Nooteboom, 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.

[26] Oliveira, M., 2000. *Prosodic features in spontaneous narratives*. Ph.D. Thesis, Simon Fraser University, Vancouver, Canada.

[27] Pijper, J. R. d. & Sanderman, A. A., 1994. On the perceptual strength of prosodic boundaries and its relation to supresagmental cues. *J. Acoust. Soc. Am.* 96(4), 2037-2047.

[28] Sá, P., Cunha, D., Lima, A. & Oliveira, M., Eds., 1997. *A Linguagem Falada Culta na Cidade do Recife: Diálogo Entre Informante e Documentador*. Recife, Editora Universitária.

[29] Selting, M., 1992. Prosody in conversational questions. *Journal of Pragmatics* 17: 315-345.

[30] Silverman, K., 1987. *Natural prosody for synthetic speech*. Cambridge, Cambridge University.

[31] Sluijter, A. M. C. & Terken, J. M. B., 1993. Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica* 50. 180-188.

[32] Swerts, M., 1997. Prosodic features at discourse boundaries of different strength.*Journal of the Acoustical Society of America* 101(1), 514-521.

[33] Swerts, M., Collier, R. & Terken, J., 1994. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication* 15, 79-90.

[34] Swerts, M. & Geluykens, R., 1994. Prosody as a marker of information flow in spoken discourse. *Language and Speech* 37, 21-43.

[35] Van Donzel, M., 1999. *Prosodic Aspects of Information Structure in Discourse.* Faculteit der Geesteswetenschappen. Amsterdam, University van Amsterdam.