Dialogic units in spoken Brazilian and Italian: a corpus based approach

In this paper we present two comparable spontaneous speech corpora and develop a cross-linguistic study on the usage of dialogic units in spoken Brazilian Portuguese and Italian. Dialogic units are information units (IU) dedicated to regulate the communication and are commonly called discourse markers. In linguistics literature, discourse markers are often defined as linguistic expressions that lose their semantic meaning and its original morphosyntactic value that do not belong to the semantic and syntactic structure of the utterance. Such expressions do not affect the truth value of the utterance [1], and are not part of the propositional content of the message conveyed, therefore not contributing to the meaning of the proposition itself [2]. They acquire different pragmatic functions, that can be textual or meta-textual. Some textual features usually attributed to discourse markers are turn-taking, silence filling, phatic function, request for attention, agreement and confirmation. Meta-textual functions can be focus, demarcation, indication of paraphrase or reformulation, modality, among others [3]. However, there is little agreement regarding the number of discourse markers, as to their functions and the criteria to define them. Discourse markers are often related to concepts such as form, attitude and emotion [4], but there are also no agreement regarding these concepts. Some authors note a strong correlation between discourse markers and some prosodic properties, such as the fact that they tend to be uttered in a dedicated prosodic unit that can be eliminated without any effect on the utterance [5].

We adopt a theoretical framework (Language Into Act Theory) [6, 7, 8] developed through empirical corpus research that identifies such expressions as dialogic information units. Dialogic units are prosodically delimited linguistic expressions that function regulating the communicative interaction [6, 9]. In Language into Act Theory [6-8], the referring unit for the analysis of the spoken language is the utterance, defined as the linguistic counterpart of a speech act. The utterance is the shortest linguistic unit that can be pragmatically interpreted and is delimited in the speech flow by prosodic boundaries that bear a conclusive value. The utterance may be organized in a single prosodic unit (simple utterance) or it can be prosodically parsed into two or more units (compound utterance), creating a prosodic pattern [10]. The units of a prosodic pattern are associated with information functions, through which information is patterned in the utterance.

Informational Patterning Hypothesis [7, 11] proposes that there is a systematic correspondence between the prosodic pattern and the information pattern of an utterance. Information Units (IU) are classified into textual and dialogic. Textual units participate to the construction of the semantic content of the utterance. Dialogic units are devoted to the successful pragmatic performance of the utterance (e.g. to regulate the relationship between speakers). Every utterance has at least one Comment unit, since it is the Comment that bears the the utterance's illocutionary force. The Comment is the only necessary and sufficient unit to form an utterance. Textual functions are: (a) Topic – identifies the domain of application for the illocution; (b) Appendix of comment – integrates the text of the comment; (c) Appendix of topic – integration of the information given in the topic; (d) Parenthesis – adds information with metalinguistic value; (e) Locutive introducer – signals a change of point of view on the subsequent locution. The dialogic functions are: (a) Incipit – opens the communicative channel while signals a contrastive value with the previous utterance; (b) Conative – pushes the listener to take part in an adequate way in the dialogue; (c) Phatic – ensures the maintenance of the communicative channel; (d) Allocutive – specifies to whom the message is directed; (e) Expressive - emotional support of the utterance; (f) Discourse Connector – signals the continuity of the discourse while establishes a relation between the previous and following units.

We present two samples of spoken corpora that received tagging at the information structure level according to the Language into Act Theory. The Italian sample comes from the C-ORAL-ROM [12] (Italian section) and the Brazilian sample comes from C-ORAL-BRASIL [13]. Our main goals are to show the distribution of information units in both languages and to discuss some interesting aspects regarding the usage of dialogic units in Brazilian and Italian.

The samples come from informal sections of oral corpora containing a broad variety of communicative situations and were selected for a strict comparison with each other. The Italian sample contains 29414 words, 5286 utterances and 11517 prosodic/information units. The Brazilian Portuguese sample has 31318 words, 5483 utterances and 9825 prosodic/information units. We extracted the data through IPIC, a theoretically-bound XML Database designed for the study of linear relation among Informative Units in spoken language corpora [14]. The data were tabulated and we analyzed the frequencies and distribution of all information units in both samples. Additionally, we investigate the specific features of dialogic units regarding its position in the utterance.

Results show a prevalence of compound utterances in Italian (30%) in comparison with Brazilian (23%) that is statistically significant (chi-square=52,848 – p<0.0001). Furthermore, in Italian information is more likely to be patterned at the textual level, with high occurrence of compound Utterances with only textual IU (44% of all compound Utterances). On the contrary, Brazilian presents a more frequent use of dialogic IU (51% of all compound Utterances), specially Expressives and Allocutives. Differences in information patterning strategies are also noted when we compare the most recurrent information patterns: Italian tends to organize information in Topic-Comment structures (5.8% of all information patterns) while Brazilian shows a relevant use of illocutive patterns (Multiple Comments, 5.2% of all information patterns).

The use of dialogic units also differs among Brazilian and Italian. In Brazilian, we have 42% of Phatics, 16% of Discurse Conectors, 13% of Expressives, 13% of Allocutives, 9% of Incipits and 7% of Conatives. The dialogic units in Italian have the following rank: 46% of Phatics, 29% of Incipits, 9% of Discurse Conectors, 8% of Conatives, 5% of Allocutives and 3% of Expressives. Comparing Brazilian and Italian with respect to all the dialogic units, we note that Brazilian uses much more Expressives and Allocutives, while Italian uses much more Incpits and Conatives. When we look at the distribution of dialogic units regarding its position inside the utterance, we notice that the Expressives are very often employed to open the utterance and/or to take the turn. In Italian, those functions are mostly performed by Incipits.

These differences suggest cultural influences in language use. Dialogic units are strongly linked to the interaction (and not the semantic content of the utterance) and therefore, sensitive to cultural nuances. Allocutives and Expressives are signs of social cohesion in discourse, while Incipits signal the speaker's opposition with respect to the previous utterance. It is likely that in Brazilian culture the Incipit is perceived as an aggressive way to take the turn or begin the utterance. For this reason, Brazilian tends to prefer Expressives to play this role.

Cross-linguistic studies are very valuable, in the sense that through the analysis of different languages we can observe which features are intrinsic to speech as a universal communicative medium and which are specific of each language. Individualizing what is specific to each language is necessary to develop and implement appropriate teaching strategies. The presence of comparable corpora and the study of the information structure in a contrastive perspective provides many useful elements for L2 teaching. It is clear that the pragmatic perspective, often invoked in education, still lacks appropriate tools of research. Corpora such as C-ORAL-ROM and the C-ORAL-BRASIL and a theoretical perspective as Language into Act Theory can provide tools to repair this deficiency.

REFERENCES

[1] Schneider, S. (1999) Il congiuntivo tra modalita e subordinazione : uno studio sull'italiano parlato. Roma: Carocci.

[2] Fraser, B. (2006) Towards a Theory of Discourse Markers. In: Fischer, K. (Ed.) Approaches to discourse particles. Amsterdam: Elsevier, p. 189-204.

[3] Fischer, K. (2006) Towards an understanding of the spectrum of approaches to discurse particles. In: Fischer, K. Approaches to discourse particles. Amsterdam: Elsevier, p. 1-20.

[4] Traugott, E. (2007) Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic. Catalan Journal of Linguistics 6, p. 139-157.

[5] Bazzanella, C.; Bosco, C.; Gili Fivela, B.; Miecznikowski, J.; Tini Brunozzi, F. (2008)

Polifunzionalità dei segnali discorsivi, sviluppo conversazionale e ruolo dei tratti fonetici e fonologici. In: Pettorino, M.; Giannini, A.; Vallone, M.; Savy, R. (Eds.) La comunicazione parlata, vol. II. Napoli: Liguori, p. 934-963.

[6] Cresti, E. (2000) Corpus di italiano parlato. Firenze: Accademia della Crusca.

[7] Cresti, E.; Moneglia, M. (2010) Informational patterning theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern. In: Moneglia, M.; Panunzi, A. (eds). Bootstrapping Information from Corpora in a Cross-Linguistic Perspective. Firenze: FUP.

[8] Cresti, E. (2011) The Definition of Focus in Language into Act Theory (LAcT). In: Mello, H.; Panunzi, A.; Raso, T. Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Firenze: FUP. p. 39-82

[9] Frosali, F. (2008) Le unità di informazione di ausilio dialogico: valori percentuali, caratteri intonativi, lessicali e morfo-sintattici in un corpus di italiano parlato (C-ORAL-ROM). In: Cresti, E. (Org.) Prospettive nello studio del lessico italiano. Firenze: Firenze University Press, p. 417-424.

[10] Hart, J't.; Collier, R; Cohen, A. (1990) A perceptual study on intonation: An experimental approach to speech melody. Cambridge: Cambridge University Press.

[11] Scarano, A. (2009) A The prosodic annotation of C-ORAL-ROM and the structure of information in spoken language. In L. Mereu (ed.), Information structures and its interfaces. Berlin and New York: Mouton de Gruyter, 51-74.

[12] Cresti, E.; Moneglia, M. (Eds.) (2005) C-ORAL-ROM. Integrated reference corpora for spoken romance languages. Amsterdam: John Benjamins.

[13] Raso, T.; Mello, H. (Orgs.) (2012) C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal. Belo Horizonte: UFMG.

[14] Panunzi, A.; Gregori, L. (2011) DB-IPIC: an XML database for the representation of information structure in spoken language. In: Mello, H.; Panunzi, A.; Raso, T. Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Firenze: FUP. p. 133-150