

Information patterning strategies in spontaneous speech: a cross-linguistic study

Elena NUZZO (Università di Verona)
Greta ZANONI (Università di Bologna)

In this paper we develop a cross-linguistic study on information patterning strategies in two romance languages: Italian and Brazilian Portuguese. The language sample comes from two comparable corpora of spontaneous speech: C-ORAL-ROM (Italian section) and C-ORAL-BRASIL. We aim to investigate the frequency of occurrence of information units in Italian (IT) and Brazilian Portuguese (BP) and to determine what are the most frequent information patterns in both languages, in order to identify the differences/similarities in the way each language organizes information.

Both speech samples are annotated at the informational level according to the units established by Language into Act Theory (Cresti, 2000). In this framework, 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 breaks that bear a conclusive value. Mostly, a prosodically terminated sequence corresponds to the performing of a single speech act. The utterance may be prosodically parsed into two or more units, creating a prosodic pattern. The units of the prosodic pattern are associated with informational functions, through which information is patterned in the utterance. Informational Patterning Hypothesis proposes that there is a systematic correspondence between the prosodic pattern and the information pattern of an utterance (Scarano, 2009; Cresti & Moneglia, 2010).

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). The set of textual information units is the follow: (a) Comment – accomplishes the utterance's illocutionary force; (b) Topic – identifies the domain of application for the illocution; (c) Appendix of comment – integrates the text of the comment; (d) Appendix of topic – integration of the information given in the topic; (e) Parenthesis – adds information with metalinguistic value; (f) 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. The most important information unit (IU) is the Comment and it is the only necessary and sufficient unit to form an utterance.

There are two cases when one terminated sequence does not correspond to a single illocutionary value: Multiple Comments and "Stanzas". Multiple Comments are a chain of Comments forming an illocutionary pattern. It is an actional model that patterns two or more illocutionary acts for the performance of one conventional rhetoric effect. A "Stanza" (Cresti, 2009) is a terminated sequence that does not correspond to only one speech act, but to a global linguistic activity, as a result of the intention of performing an oral text, such as narratives and argumentations. It corresponds to a sequence of Bound Comments (COB), with homogeneous illocutionary forces. A "Stanza" may contain other information units forming sub-patterns.

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 is derived from the C-ORAL-ROM Italian (Cresti; Moneglia, 2005) and contains 29414 words, 5286 terminated sequences and 11517 prosodic/information units. The Brazilian Portuguese sample is derived from C-ORAL-BRASIL (Raso; Mello, 2010) and has 31318 words, 5483 terminated sequences 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 (Panunzi; Gregori, in press).

Results show a prevalence of compound utterances in Italian (30%) in comparison with Brazilian (23%) that is statistically significant ($\chi^2=52,848 - p<0.0001$). 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).

These differences could be interpreted as a result of cultural influences in language use, since dialogic IU like Allocutives and Expressives are signs of social cohesion in discourse. However, a more qualitative look into the data is needed, in order to assure that such differences do not derive from sampling incompatibilities or problems in the information annotation.

References

Cresti, E. 2000. *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

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

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.

Panunzi, A.; Gregori, L. in press. DB-IPIC: an XML database for the representation of information structure in spoken language. To appear in: Mello, H.; Panunzi, A.; Raso, T. (eds.) *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*.

Raso, T.; Mello, H. 2010. The C-ORAL-BRASIL corpus. In: Moneglia, M.; Panunzi, A. (eds) *Bootstrapping Information from Corpora in a Cross Linguistic Perspective*. Firenze: FUP.

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.