

Syntactic variability in Topic-Comment utterances in spoken Brazilian Portuguese

Aims

This poster presents an overview of the syntactic variability found in spontaneous spoken Brazilian Portuguese (BP), focusing on utterances that are composed by a Topic-Comment information pattern. We discuss the role of syntax, prosody and pragmatics in the analysis of spoken language.

Methods

The linguistic variety of BP considered for this study is the one of Belo Horizonte-MG, as represented in C-ORAL-BRASIL I reference corpus for spontaneous speech. The data comes from the Brazilian Mini-corpus, which is a representative sample (20 sessions, 31,465 words) of C-ORAL-BRASIL informal section is available online at the IPIC Information Structure Database. The mini-corpus provides text-to-speech alignment, audio files and transcriptions, which incorporate the annotation of prosodic boundaries and information functions.

The prosodic annotation scheme segments the speech flow in two distinct levels. The first delimits utterances (prosodically and pragmatically autonomous linguistic expressions); the second level refers to the internal structure of the utterance, that can be constituted by one or several tone units. Each tone unit within an utterance convey a information function: Comment (COM/COB/CMM), Topic (TOP/TPL), Appendix of Comment (APC), Appendix of Topic (APT), Parenthetical (PAR/PRL) and Locutive Introducer (INT), Incipit (INP), Conative (CNT), Allocutive (ALL), Phatic (PHA), Expressive (EXP) and Discourse Connector (DCT). This annotation scheme is based on the Information Patterning hypothesis and the Language into Act Theory, according to which the speech act is the fundamental communication activity and the utterance is its linguistic realization. Therefore, the utterance is the reference unit for the study of spontaneous speech. Prosodic parsing is the basis for structuring speech: prosodic units correspond to information units. The more important units of an information pattern are Comment (COM) and Topic (TOP). The COM is the single unit needed to form an information pattern, because it conveys the illocutionary force, being sufficient to perform a speech act. The TOP has the function of defining the domain of reference in which the illocution expressed in the COM must be interpreted.

Through the IPIC online database, all utterances containing Topics (507) were exported into a spreadsheet, where they were classified. In the 507 utterances with Topics, there is a total of 600 Topic units (some utterances have more than one Topic), including concluded and interrupted utterances. The morphosyntactic content of all topic units was classified according to the type of phrasal head and the presence of a verbal form.

Results

Topics from the sample have a great variability regarding its morphosyntactic content. Overall, it was observed that utterances will present more or less complex content according to the communicative situation. When the interaction among participants is less anchored in immediate context, utterances tend to be more complex, with more occurrences of TOP. The summary of all syntactic structures of Topics in the sample are presented in the Table below.

Morphosyntactic content	Frequency	%
Total	599	100,0%
Nominal phrase	238	39,7%
Nouns	136	57,1%
Personal pronouns	36	15,1%
Demonstrative pronouns	30	12,6%
Proper names	25	10,5%
Indefinite pronouns	7	2,9%
Numerals	4	1,7%
Verbal phrase	222	37,0%
Finite verbal head	195	87,8%

Infinitive verbal head	16	7,2%
Gerund verbal head	10	5,0%
Prepositional phrase	75	12,5%
Adverbial phrase	62	10,3%
Adjectival phrase	3	0,5%

The TOP is most commonly accomplished through a noun (39.5%) or verb (37.7%) phrase. In the first case, the NP is typically filled by nouns and its determiners, but there is a high occurrence of demonstrative pronouns in dialogues, and personal pronouns in monologues. It is also interesting to notice that only around 3% of nominal topics present some kind of pronominal replication (copy pronoun) in the Comment unit. In the second case, there is a predominant use of VP to express hypothetical situations in conversations and dialogues, while in monologues it is more relevant that the TOP delimits a temporal domain of reference. Just by analyzing a relative small set of data, it is possible to notice that Topic-Comment utterances on spontaneous speech present a great variability in its syntactic content. The great frequency of Topics formed by verbal clauses indicates that the sentence is hardly an adequate reference unit for spontaneous speech.