

Thematic panel: **Lexicon and Semantics in Speech**

Modality in perspective: corpus-based study of spontaneous speech of Brazilian Portuguese

This study deals with the description of modality in a *corpus* of spontaneous Brazilian Portuguese spoken language, more specifically, the diatopy of Minas Gerais, mainly the metropolitan zone of the capital of the state, Belo Horizonte. Modality, in a nutshell, is a semantic category applied to a conceptualizer's linguistic production, which qualifies and relativizes the uttered locutive material in terms of degree of certainty, possibility, necessity, capability, and volition. However, there is no consensus on a precise definition of such category, which can be explained by a number of reasons: (a) modality is a field of study for logics and linguistics that utilizes various methodologies; (b) the category is both differentiated and interrelated to others, such as time, aspect, mode; (c) this notion is mistaken with and overlaps those of illocution, attitude, and emotion. Based on the principles of the Speech Patterning Theory (CRESTI, 2000) and in Corpus Linguistics, this paper aims at describing the behavior of different modality markers, lexical and grammatical, such as for example: (semi-) auxiliary modal verbs, modal adverbs, verbs of propositional attitude, volitional verbs, modal adjective constructions, emerging forms, future, and conditional in the C-ORAL-BRASIL corpus. C-ORAL-BRASIL is a *corpus* of Brazilian Portuguese spoken language, the fifth branch of C-ORAL-ROM, a comparable *corpus*, representative of the four main European Romance Languages (Italian, Spanish, French, and Portuguese), prosodically segmented into utterances and their informational sub-units, and speech-transcription aligned through the software WinPitch — allowing at the same time the examination of sound, spectrogram and text. The utterance is the minimal reference unit pragmatically interpretable. To better characterize the purposes of this research, we utilized a sample of the informal part of the C-ORAL-BRASIL *corpus*, a *sub-corpus* composed of 20 texts, equally distributed in three interactional typologies, divided into private and public: 7 monologues – 6 private and 1 public –; 7 dialogues – 5 private and 2 public –; and 6 conversations – 4 private and 2 public ones. For the treatment in this paper, 1.197 modal indexes, distributed in 1.046 utterances, were utilized. Quantitatively, the types and instantiated tokens were classified according to: (a) Metadata, information on each file: interactional typology (monologue versus dialogue versus conversation, public versus private), speaker schooling level (roughly, level 1: primary school, level 2: secondary school, level 3: college and professional education), and textual typology (narrative, expositive, argumentative, and instructional. These variables are not parameters controlled either in C-ORAL-BRASIL or in C-ORAL-ROM.); (b) Lemmas and lexemes: identification of the modal item of each utterance and their respective forms; (c) Information structure: information unit which contains the modal index (Comment, Multiple Comment, Bound Comment, Topic, Topic List, Parenthetical, Parenthetical List, Locutive Introducer, Appendix); (d) Modal values and respective modal sub-values: epistemic (knowledge, belief, possibility, probability, necessity, verification), deontic (obligation, permission, necessity), dynamic (ability and volition); (e) Part of speech of the modal index; (f) Compositionality: whether that is just one modal index in each utterance; whether there are two or more indexes in the same informational unit and same utterance; whether there is an index in the same utterance and in different information units. Quantitatively, we analyzed the frequency of modalized utterances in the sample, and their distribution. Besides, we described the frequency of morpholexical indexes in relation to the information

unit which comprises the index, and the frequency of the modal types related to the markers and each information unit. Additionally, we analyzed, qualitatively and quantitatively, each modal marker, in terms of: syntactic ambiances; relation between syntactic structure and information structure, distribution in context and interactional typology. After the quantitative and qualitative analysis, we can conclude that (i) the category of modality behaves distinctively in written and spontaneous oral language; (ii) the modal indexes are widely utilized in dialogic situations, be them public or private; (iii) only a few informational units can be modalized: Comment, Parenthetical, Topic, Locutive Introducer; (iv) the Comment unit is the most modalized and there is no realization restriction for this IU; (v) the epistemic modality is the most frequent one among the types of meaning; (vi) verbs are the preferred strategy for modality marking.

Keywords: modality; oral *corpus*; spontaneous speech; Brazilian Portuguese.