



Veredas atemática
Volume 19 nº 2 – 2015

MASS - Modal Annotation in Spontaneous Speech: semantic annotation scheme for modality in a spontaneous speech Brazilian Portuguese corpus¹

Luciana Beatriz Ávila² (UFV)

ABSTRACT: This paper aims at introducing a modality semantic annotation scheme for spontaneous speech data of Brazilian Portuguese. Our research is inspired by previous schemes proposed for European Portuguese and English and based on the Language into Act Theory (CRESTI, 2000). We validated our scheme on a corpus sample of the C-ORAL-BRASIL I and we annotated 781 lexical modal markers using the MMAX2 annotation tool. We found as main results that: (a) epistemic modality is the most frequent value; (b) triggers are mostly modal verbs; (c) the source of the event mention is usually the speaker; (d) the source of modality coincides with the speaker in 88,3% of the occurrences; and (e) targets are realized mostly within the informational unit.

Keywords: modality; semantic annotation; spontaneous speech; oral corpus; Brazilian Portuguese.

Introduction

This paper introduces the MASS project (Modal Annotation in Spontaneous Speech), a modality semantic annotation scheme for spontaneous speech data of Brazilian Portuguese. We describe our annotation scheme and present some results.

Linguistic annotation, in Leech's words (1993, p. 275), can be defined as “[...] the practice of adding *interpretative* (especially linguistic) information to an existing corpus of spoken and/or written language, by some kind of coding attached to, or interspersed with, the electronic *representation* of the language material itself”. Following up on this view, we propose an annotation scheme for modality in order to codify certainty degrees, as well as ability and possibility values ensued through lexical indexes in speech.

The proposal falls within the range of current interest in the distinction of real

¹ Part of this work was supported by a grant from Capes, Proc. nº BEX 9537-12-0.

² E-mail: luciana.avila@ufv.br

information from a speculative and modal one as this is a necessary task for NLP applications, such as information extraction (KARTUNNEN; ZAENNEN, 2005), uncertainty modeling in clinical texts (MOWERY et al., 2012), question answering (SAURÍ et al., 2006); hedge detection and classification (MEDLOCK; BRISCOE, 2007); and sentiment analysis (WIEBE et al., 2005).

Modality annotation includes the identification of modal indexes, their classification in a specific typology (for instance, epistemic versus non-epistemic), definition of their source and semantic scope. Many are the works developed aiming at modal expression annotation, most of them for English, dealing with modal verbs. Some relevant studies are: the relation of modality and negation (MORANTE; SPORLEDER, 2012; BAKER et al., 2012), sense-annotation of modal verbs (RUPPENHOFER; REHBEIN, 2012) and the development of a modality lexicon and the construction of automatic taggers (BAKER et al., 2010). There some annotation efforts being developed in other languages, such as the ones for Chinese (CUI; CHI, 2013) and for European Portuguese (HENDRICKX et al., 2012a, 2012b; MENDES et al., 2013). Modality annotation for Brazilian Portuguese data is an unexplored field both for written and spoken corpora.

1. Defining modality

The literature on the characterization of modality shows that there is no consensus on how to define and characterize it: modality can be taken as the expression of subjectivity, or as a distinction between *realis* and *irrealis*, or even as a quantification over possible worlds, restricted by an accessibility relation. Thus, the understanding of what this semantic category is and which elements could vehicle it is absolutely necessary.

Our research is based on the Language into Act Theory -LAcT (cf. CRESTI, 2000), d'après Austin's Speech Act Theory (1962), which associates spontaneous speech and speech acts. According to Cresti and Scarano (1998), spontaneous speech is governed by an illocutionary principle, not found in written texts, as well as specific informational articulations. In L-AcT, the analytical reference unit is the *utterance* which is pragmatically defined. An utterance carries an illocution and its locutory material does not necessarily carry a proposition. It may be simple, when comprised by just one tone unit (the Comment unit under an information structure view) or complex, when it is made up by two or more tone units (the Comment unit and any textual or dialogic units).

Modality is taken here as a conceptualizer's evaluation, nuancing which is uttered, in terms of the degree of certainty and based on the notions of possibility, necessity, ability/capacity and volition/intention, towards the locutory material in a given utterance. The scope of modality is the tone unit, as realized by information textual units, following the analysis proposed by Tucci (2007). Hence, within a given complex utterance, there might be different tone units which carry different modal values. When a tone unit carries more than one modal marker, they may not share the same modal value, in which case the dominant modality will prevail, as we can see in the examples below:

- (a) e eu **acredito** que depois que eu terminar o EDUCONLE / eu **acho** que aí eu **vou** tar mais madura ainda / **acho** que mais preparada / e eu e **vou conseguir** / sabe / vencer essas dificuldades aí //

and I believe that when I finish EDUCONLE / I think that I'll be more mature / I think more prepared / and I and I'll be able / you know / to overcome these difficulties //

In (a), there are six information units making up a complex utterance. Three of them convey epistemic modality, while the last one carries dynamic modality. Albeit belonging to the same utterance, the modalities that mark each information unit are not semantically compositional. Whereas in (b) below, two modal values within the same information unit (epistemic and deontic) will be compositional and the dominant value (epistemic) prevails over the dominated one (deontic), in a hierarchical chain:

(b) *cê acha que eu devo convidar o Guilherme //*
do you think I should invite Guilherme //

The utterance in (b) is made up of a single tone unit, and carries two modal items – the first one epistemic, the second, deontic – and the whole modal value of the unit is epistemic.

2. MASS project (Modal Annotation of Spontaneous Speech)

Our annotation project for Brazilian Portuguese spontaneous speech data closely follows the scheme proposed for European Portuguese (HENDRICKX et al., 2012a; 2012b; MENDES et al., 2013) and is also inspired by other systems previously explored for English (BAKER et al., 2010; SAURÍ et al., 2006, 2008, 2009; SZARVAS et al., 2008) and for Japanese (MATSUYOSHI et al., 2010).

As this scheme is based on the Language into Act Theory and has as referential unit the utterance and the information units (cf. CRESTI, 2000), it differs from other projects, since the analysis is not centered on the writing diamesy, and, therefore, does not provide annotation on the sentence domain. Besides, it differs from the model proposed to European Portuguese, because of the theoretical options adopted here in terms of modal values and sub-values and, mainly, in terms of conceiving the targets not as a whole proposition or a complete predicate, but rather as linguistic chunks, given that modality occurs within the scope of information unit.

The repertoire of modal meanings is by no means settled and stable across the literature, nevertheless, there is agreement among different approaches with respect to epistemic meanings; the debatable issues are related to subtleties in non-epistemic meanings. MASS takes into consideration three modal values – epistemic, deontic and dynamic – and thirteen sub-values. The distribution and definition of each one is shown in Table 1:

Values	Sub-values	Definition
Epistemic	knowledge	The conceptualizer (the speaker or another entity) expresses their knowledge or understanding about something.
	belief	The conceptualizer expresses their belief or opinion that something is the case.
	possibility	The conceptualizer expresses or points out a possibility that something is the case.
	probability	The conceptualizer expresses or points out a probability that something is the case.
	necessity	The conceptualizer, given a set of evidence, expresses or points out the necessity of something to be the case.
	verification	The conceptualizer expresses uncertainty towards a state of affairs, an event or an activity in focus.
Deontic	obligation	The conceptualizer finds themselves obligated or forced to or obliges themselves to carry on an activity for some specific reason.
	permission	The conceptualizer expresses or points out a permission to someone/something or themselves to do something, or allows that something happens.
	prohibition	The conceptualizer expresses or points out a prohibition to someone/something or themselves to do something, or forbids that something happens.
	necessity	The conceptualizer expresses or points out a restriction due to something.
Dynamic	ability	The conceptualizer expresses their own ability or someone else's ability to do or to achieve something.
	volition	The conceptualizer expresses their or someone else's wills, necessities, desires, hopes and intentions.

Table 1 – Modal values and sub-values and their definitions

In addition to modal values, the annotation scheme is made up of the following components:

- (a) **Trigger:** Triggers are the “words or string of words that express modality” (BAKER et al., 2010). We only consider lexical markers as triggers: modal verbs, epistemic or propositional attitude verbs, modal adverbs, adjective expressions, and lexical expressions that convey modality.

This component has some specific features: the modal value, its polarity (positive or negative) and the information unit which carries the modal item. These textual information units are presented in Table 2:

	IU	Informational function	Tag
Textual units	Comment	Expresses the illocutionary force of the utterance	COM
	Topic	Specifies the locus of application of the illocutionary force of the Comment	TOP
	Parenthetical	Expresses metalinguistic integration of the utterance	PAR
	Locutive introducer	Signals pragmatic suspension of the <i>hic et nunc</i> and introduces a metaillocution	INT

Table 2 – Modalized textual units (CRESTI, 2000)

- (b) **Source of the modality:** the conceptualizer, which might coincide with the speaker, the addressee, or another individual whose perspective is taken into consideration;

Values	Sub-values	Definition
Epistemic	knowledge	The conceptualizer (the speaker or another entity) that expresses their knowledge or understanding about something.
	belief	The conceptualizer that expresses their belief or opinion that something is the case.
	possibility	The conceptualizer that expresses or points out a possibility that something is the case.
	probability	The conceptualizer that expresses or points out a probability that something is the case.
	necessity	The conceptualizer, given a set of evidence, that expresses or points out the necessity of something to be the case.
	verification	The conceptualizer that expresses uncertainty towards a state of affairs, an event or an activity in focus.
Deontic	obligation	The conceptualizer that finds themselves obligated or forced to or that obliges themselves to carry on an activity for some specific reason.
	permission	The conceptualizer that expresses or points out a permission to someone/something or themselves to do something, or allows that something happens.
	prohibition	The conceptualizer that expresses or points out a prohibition to someone/something or themselves to do something, or forbids that something happens.
	necessity	The conceptualizer that expresses or points out a restriction due to something.
Dynamic	ability	The conceptualizer that expresses their own ability or someone else's ability to do or to achieve something.
	volition	The conceptualizer that expresses their or someone else's wills, necessities, desires, hopes and intentions.

Table 3 – *Sources* corresponding to each modal vales and sub-values

- (c) **Source of the event mention:** the producer, the speaker;
- (d) **Target:** the expression in the scope of the trigger within an annotation unit, that is, information units (IU) that carry modality, as described in Table 2. The *target* is marked for positive or negative polarity. It is maximally annotated, admitting discontinuity, within the scope of the information unit which comprises the modal index.

The *target-dependent* component was created to consider the cases in which the *target*, in a given utterance, is not explicit, but it is recoverable in the referential chain of the text.

These elements that participate in a modal event belong to the same modal “set”.

3. Methodology

Our scheme was applied to the Brazilian Portuguese spontaneous speech corpus C-ORAL-BRASIL I (RASO; MELLO, 2012). This corpus follows the same architecture as the European Romance spontaneous speech corpus C-ORAL-ROM (CRESTI; MONEGLIA, 2005), whereby diaphasic variation is privileged in order for a large diversity of illocutions and informational structuring to be documented. C-ORAL-BRASIL I comprises 200 texts of approximately 1,500 words each, proportionally distributed into dialogues, conversations and monologues. The corpus follows the CHILDES-CLAN transcription format to which prosodic annotation is added, marking tone unit and utterance boundaries, besides documenting several phenomena typical of speech. The entire corpus is speech to text aligned through the WinPitch software (MARTIN, 2004).

In this study an informationally annotated sample from the C-ORAL-BRASIL I was studied. It covers 20 texts, totaling 31,318 words, 5,484 utterances and 9,825 tone units. Firstly, the identification and classification of modal markers was undertaken by three annotators working independently; the codification was then qualitatively validated through group discussions involving the research group coordinator and her students. The search for modal markers was performed manually, through qualitative transcription examination, supported by the WinPitch text-to-audio aligned files and their concomitant examination through the software interface that allows speech signal listening as well as transcription and prosodic parameters visualization. The data were organized in a table containing the modal markers, the tone unit in which they occur, the type of information unit they are inserted in, the file they belong to, and any qualitative information deemed relevant.

Further, modal annotation was carried through MMAX2 (MÜLLER; STRUBE, 2006), a free software tool for linguistic annotation in multiple levels.³ MMAX2 offers a visual interface to annotate sentences by marking textual strings and creating links between the marked elements. The modal annotation was carried by a single annotator (ÁVILA, 2014a, 2014b). An example of modal annotation is shown in (c) and is summarized in Table 3:

(c) EVN_{S1}: [171] é /=PHA= a <gentes_{S1} tem que_M> <restringir também_T /=COM= **isso**>
//=APC=\$
EVN_{S1}: [171] *yeah* /=PHA= <we_{S1} have to_M> <restrict_T too /=COM= *this*>
//=APC=\$

³ <http://mmax2.net>.

Components	Counterparts
Trigger	tem que
Polarity	pos
IU	COM
Source of the event mention	*EVN
Source of the modality	A gente (1pl)
Modal value	deontic_obligation
Target	restringir também /
IU	COM

Table 4 – Annotation for the utterance [171] of *bfamcv01*

In Figures 1 and 2, example (c) is exhibited through the software visual interface and the *trigger* and its affected categories are shown:

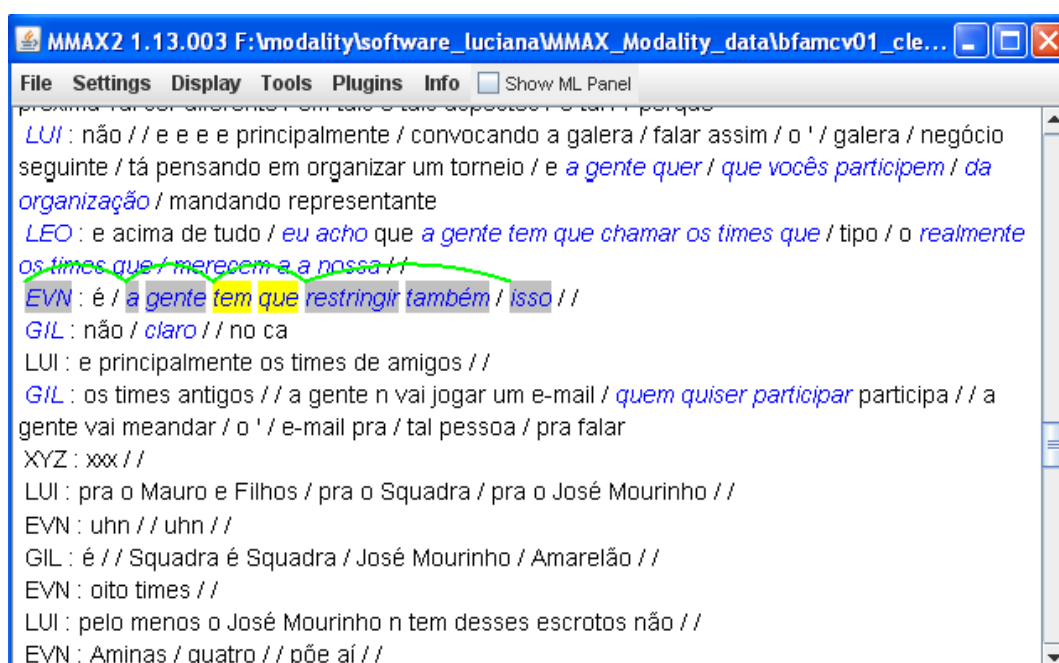


Figure 1 – Utterance (c) visualization

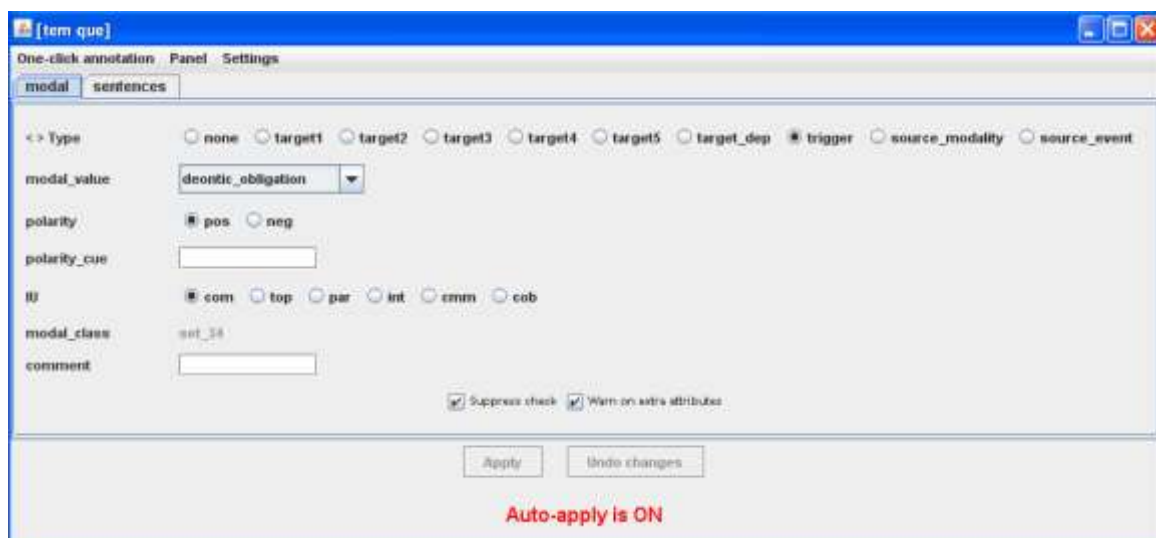


Figure 2 – The *trigger* and its affected categories in (c)

4. Results

In our sample we found 1,088 modal markers (lexical and grammatical, excluded the conditional constructions) and from these we tagged 781 triggers. Table 4 presents the distribution of modal values and sub-values in the *sub-corpus*:

Values	Sub-values	Freq	%
Epistemic		506	64,7
	knowledge	100	21,3
	belief	228	44
	possibility	120	24,1
	probability	24	4,7
	necessity	15	1,9
	verification	14	3,7
Deontic		189	24,2
	obligation	96	50,8
	permission	70	37
	prohibition	6	3,2
	necessity	17	9
Dynamic		86	11,1
	ability	17	19,8
	volition	69	80,2

Table 5 – Frequency of modal values in the sample

Epistemic modality is the most frequent value (64,75) and the *belief* sub-value, which comprises epistemic verbs and adverbs, is the most numerous, representing 44% of all epistemic indexes. In terms of deontic values, the most frequent use is “obligation”, half of all sub-values, and the deontic prohibition is the less frequent, possibly because is considered as stronger deontic permission. The occurrences of volition correspond to 80,2% of dynamic

modality, and in a low number, the cases of ability/capacity, confirming the less central character of this type.

Triggers are mostly modal verbs, 81,5% of all indexes (65%, auxiliaries and semi-auxiliaries; 31%, epistemic modals), followed by adverbs (11,8%). Less frequent are adjectives constructions (2,8%) and modal expressions (3,9%).

The source of the event mention is usually the speaker. The source of modality coincides with the speaker in 88,3% of the occurrences, but we find different levels of conceptualization involving the perspective of the addressee or another entity, besides that of the speaker. However, we must highlight that these conceptualizers perspectives' different from the utterer one are often filtered by the speaker's eyes, mainly in occurrences with belief/epistemic verbs, corroborating Wiebe and collaborators claim (2005, p. 9) that nesting is an important property of the sources.

Speaking of targets, in 79,1% of all occurrences they are realized within the informational unit. In eleven occurrences, when not realized within the informational unit, especially in Parenthetical unit, they were not annotated. In one of the cases it is unspecified, because in this particular occurrence one should have the whole scene registered both in video and audio files. Finally, we have five cases that the targets are in patterned constructions, i.e., "[...] constructions performed across TUs, with each developing a different information function [...]" (CRESTI, 2014, p. 17), three in which the Locutive Introducer unit comprises the trigger, and two in Bound Comment unit, with the targets within the Comment unit.

We found 151 occurrences of the target-dependent attribute, which corresponds to 13,9% of all occurrences. According to Cresti (2014), "[e]ach linguistic chunk, conceived to perform a certain textual function (TU) within an information pattern (IP), corresponds to a scene (BARWISE; PERRY 1981, FAUCONNIER, 1984) from a semantic point of view. As we have already said, from a syntactic point of view a TU can even correspond to a collection of fragments, but in order to allow the development of a textual function, the participating expressions must be gathered within the same scene".

The target in spoken texts is the expression affected by the modal item conveyed in the trigger within a given information unit. Therefore, it is not necessarily a complete predicate, as it is usually the case in written texts (a subordinate clause or an event with all its arguments and adjuncts). In speech, as pointed out by Cresti (2014), "a large number of spoken chunks, indeed, cannot be defined as clauses, but are rather fragments, interjections, adverbs, phrases, while nevertheless functioning properly from a communicative point of view".

4.1 Inter-annotator agreement

The annotation of the *sub-corpus* was accomplished by one single annotator so far. However, in order to validate the scheme and check its coherence and feasibility, our next step is to perform tests with at least two more annotators, undergraduate or graduate students.

The inter-annotator agreement will be computed using Kappa Statistics (COHEN, 1960), for all the annotation components.

Final remarks

MASS puts forth an annotation scheme appropriate for the treatment of modality in speech. It is important to highlight that the annotation scheme proposed can be applied to languages other than Brazilian Portuguese (taking into account their particularities), and provides “a reliable starting point” to NLP researchers (ÁVILA; MELLO, 2013) for the extraction of certainty markers and factuality, sentiment analysis, and opinion mining.

Esquema de anotação semântica para modalidade em *corpus* de fala espontânea do Português Brasileiro

RESUMO: Este artigo introduz um esquema de anotação da modalidade para dados de fala espontânea do português brasileiro. Nossa pesquisa inspira-se em outros esquemas propostos para o português europeu e o inglês e baseia-se na Teoria da Língua em Ato (CRESTI, 2000). Validamos o esquema em uma amostra do corpus C-ORAL-BRASIL I e anotamos 781 marcadores modais lexicais, usando a ferramenta MMAX2. Encontramos como resultados principais que: (a) a modalidade epistêmica é o valor mais frequente; (b) os *triggers* são em sua maioria verbos modalizadores; (c) a *source of the event mention* é normalmente o falante; (d) a *source of modality* coincide com o falante em 88,3% dos casos; e (e) os *targets*, em sua maioria, se realizam dentro da unidade informacional.

Palavras-chave: modalidade; anotação semântica; fala espontânea; corpus oral; português brasileiro.

REFERENCES

AUSTIN, J. *How to do things with words*. Oxford: Oxford University Press, 1962.

ÁVILA, L. Modalidade em perspectiva: estudo baseado em corpus de fala espontânea do português brasileiro. 253f. Thesis (PhD – Linguistics). Belo Horizonte, Universidade Federal de Minas Gerais, 2014a.

ÁVILA, L. MASS 1.0 – Manual de anotação. Belo Horizonte / Lisboa: PosLin-UFMG / Centro de Linguística da Universidade de Lisboa, 2014b.

ÁVILA, L.; MELLO, H. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.

BAKER, K., BLOODGOOD, M., DORR, B. J., FILARDO, N. W., LEVIN, L., PIATKO, C. A modality lexicon and its use in automatic tagging. In: *Proceedings of the Seventh Language Resources and Evaluation Conference (LREC'10)*, 2010.

BAKER, K.; DORR, B.; BLOODGOOD, M.; CALLISON-BURCH, C.; FILARDO, N. W.; PIATKO, C.; LEVIN, L.; MILLER, S. Use of modality and negation in semantically-informed syntactic MT. *Computational Linguistics*, 38(2): 411-438, 2012.

BARWISE, J.; PERRY, J. Situation and attitudes. *Journal of Philosophy*, 78 (11):668-69, 1981.

BICK, E. *The Parsing System “PALAVRAS”*: Automatic Grammatical Analysis of Portuguese

in a Constraint Grammar Framework. Aarhus: Aarhus University Press, 2000.

CHILDES - Child Language Data Exchange System. Disponível em: <http://childes.psy.cmu.edu/>. Último acesso: 29 nov. 2012.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, p. 37-46.

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

CRESTI, E.; SCARANO, A. Sur la notion de parlé spontané. 1998. Disponível em: <http://lablita.dit.unifi.it/preprint/preprint-98coll03.pdf>. Último acesso: 15 nov. 2012.

CRESTI, E.; MONEGLIA, M. *C-ORAL ROM*: Integrated reference corpora for spoken Romance languages. Amsterdam/Philadelphia: John Benjamins, 2005.

CRESTI, E. Syntactic properties of spontaneous speech in The Language into Act Theory: data on Italian complements and relative clauses. In: Raso, T. & Mello, H. (eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, 2014, p. 365-410.

CUI, Y.; CHI, T. Annotating Modal Expressions in the Chinese Treebank. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.

FAUCONNIER, G. *Espaces mentaux*. Paris: Éditions de Minuit, 1984.

HENDRICKX, I.; MENDES, A.; MENCARELLI, S.; SALGUEIRO, A. et al. Modality Annotation Manual, version 1.0. Centro de Linguística da Universidade de Lisboa, 2012a. (Manuscript)

HENDRICKX, I.; MENDES, A.; MENCARELLI, S. Modality in Text: a Proposal for Corpus Annotation. In: *Proceedings of the Eighth Conference on the International Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, p.1805-1812, 2012b.

KARTTUNEN, L.; ZAENEN, A. Veridicity. In: KATZ, G.; PUSTEJOVSKY, J.; SCHILDER, F. (eds.). *Dagstuhl seminar Proceedings*. Annotating, extracting and reasoning about time and events. Dagstuhl, Germany, 2005.

LEECH, G. Corpus annotation schemes. *Literary and linguistic computing*, vol. 8, nº 4, 1993, p. 275-281.

MARTIN, P. WinPitch Corpus, a Text to Speech Alignment Tool for Multimodal Corpora. *Proceedings of the Fourth Conference on the International Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, 2004.

MATSUYOSHI, S., EGUCHI, M., SAO, C., MURAKAMI, K., INUI, K., MATSUMOTO, Y. 2010. Annotating Event Mentions in Text with Modality, Focus, and Source Information. In: *Proceedings of the Seventh Conference on the International Language Resources and Evaluation (LREC'10)*, 2010.

MEDLOCK, B.; BRISCOE, T. Weakly supervised learning for hedge classification in

scientific literature. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic, Association for Computational Linguistics, 2007, p. 992-999.

MENDES, A.; HENDRIKX, I.; SALGUEIRO, A.; ÁVILA, L. Annotating the Interaction between Focus and Modality: the case of exclusive particles. In: *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, Sofia, Bulgaria, Association for Computational Linguistics, p. 228–237, 2013.

MORANTE, R.; DAELEMANS, W. Annotating modality and negation for a machine reading evaluation. In: FORNER, P.; KARLGREN, J.; WOMSER-HACKER, C. (eds.). *CLEF 2012 Conference and Labs of the Evaluation Forum - Question Answering For Machine Reading Evaluation (QA4MRE)*, Rome, Italy, 2012.

MORANTE, R.; SPORLEDER, C. Modality and Negation: An Introduction to the Special Issue, *Computational Linguistics*, 38(2), 2012.

MOWERY, D. L.; VELUPILLAI, S.; CHAPMAN, W. W. Medical diagnosis lost in translation – Analysis of uncertainty and negation expressions in English and Swedish clinical texts. In: *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (BioNLP 2012)*, Montreal, Canada, Association for Computational Linguistics, p. 56-64, 2012.

MÜLLER, C.; STRUBE, M. Multi-level annotation of linguistic data with MMAX2. In: *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. New York: Peter Lang, 2006, p. 197-214.

RASO, T.; MELLO, H. *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal e DVD multimedia*. Belo Horizonte: Editora UFMG, v. 1, 2012.

RUPPENHOFER, J., REHBEIN, I. Yes we can!/? Annotating English modal verbs. In: *Proceedings of the Eighth conference on the International Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012.

SAURÍ, R., VERHAGEN, M., PUSTEJOVSKY, J. Annotating and recognizing event modality in text. In: *Proceedings of the 19th International FLAIRS Conference*, FLAIRS 2006, 2006.

SAURÍ, R. A Factuality Profiler for Eventualities in Text. Ph.D. Thesis. Brandeis University, 2008.

SAURÍ, R.; PUSTEJOVSKY, J. FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 2009. Disponível em: <http://link.springer.com/article/10.1007/s10579-009-9089-9/fulltext.html>. Último acesso: 28 jan. 2013.

SZARVAS, G., VINCZE, V., FARKAS, R., CSIRIK, J. The BioScope corpus: annotation for negation, uncertainty, and their scope in biomedical texts. In: *Proceedings of the Workshop on*

Current Trends in Biomedical Natural Language Processing, Association for Computational Linguistics, 2008, p. 38-45.

TUCCI, I. 2007. La modalità nel parlato spontaneo e il suo dominio de pertinenza. Una ricerca corpus-based (C-ORAL-ROM italiano). In: *Actes du XXVe CILPR*. (Innsbruk 3-8 September 2007). Disponível em: <http://lablita.dit.unifi.it/preprint/preprint.2008-02-06.3658867998> Último acesso: 13 jan. 2013.

WIEBE, J., WILSON, T., CARDIE, C. Annotating expressions of opinions and emotions in language. Kluwer Academic Publishers, 2005, p. 1-54.

Data de envio: 26/05/2014

Data de aceite: 04/03/2015

Data de publicação: 30/04/2015