MODALITY CORPUS: ANNOTATION MANUAL

Version 0.1 Silvia Mencarelli, Jan 2012

Version 0.2 Iris Hendrickx, Feb 2012

Version 1.0 Iris Hendrickx, Amália Mendes, Silvia Mencarelli, May 2021

Version 2.0 Iris Hendrickx, Amália Mendes, Silvia Mencarelli, November 2021

Version. 3.0 Iris Hendrickx, Amália Mendes, Silvia Mencarelli, August 2022

In this annotation manual, we present the annotation scheme for Portuguese and the raw data that was annotated. We describe how to use the software annotation tool MMAX2. We also discuss difficult cases and the annotation rules.

Version 1.0 corresponds to the version that is connected to the paper by Hendrickx et al (2012).

Version 3.0 follows the slightly modified version of the modality annotation scheme presented in version 2.0, in line with the published unified schema detailed in Ávila et al. (2015). Version 3.0 describes a new corpus annotated for modality. We will refer to this corpus as CRPC-Modal, while the corpus described in versions 1.0 and 2.0 is Modality Corpus 1.

1. Introduction

Modality is usually defined as the expression of the speaker's opinion and of his attitude towards what he is saying (Palmer,1986).

Here we introduce our annotation scheme for modality in Portuguese. We created a new scheme, because we wanted a more theoretically-oriented perspective which focuses on a detailed variety of modal values and not on components such as factuality or conceptual relations. We consider seven modal values.

Epistemic modality denotes the commitment of the speaker towards the truth of the proposition (considered as possible, probable or certain). We identify four sub-values: epistemic belief, epistemic possibility, epistemic probability and epistemic knowledge.

Deontic modality denotes when the speaker or another participant imposes something on the hearer or another entity, or when the circumstances establish a necessity. We identify three sub-values: deontic obligation, deontic permission and deontic_necessity.

Inspired by the work of van der Auwera et al. (1998) we also mark **participant-internal modality**, with three sub-values: participant-internal necessity, to tag personal needs of the speaker or participant; participant-internal ability, to tag personal capacities of the speaker or participant; participant-internal volition for hopes and wishes.

We use the value **evaluation** to annotate the speaker's or participant's evaluation of the proposition

For the annotation itself we use a software tool call MMAX2[1] that produces stand-off annotation in XML format (Müller and Strube, 2006).

In this annotation scheme, we annotate the different components that participate in the expression of modality. These are the components that can play a role in modality:
- Trigger: the word or string of words that expresses modality (Baker et al., 2010)
- Target: the expression in the scope of the trigger
- Source of the event mention: the producer of the event mention, generally the speaker or writer
- Source of the modality: the entity that is the source of the modal value

The trigger is the most important component in the scheme and this carries the modal value (one of the seven that we briefly mentioned above). The other components are optional as they are not always expressed in the sentence.

2. Data

2.1 Modality corpus 1

The modality corpus for Portuguese is a sample of 1946 sentences extracted from the written part of the Corpus de Referência do Português Contemporâneo (CRPC) (Généreux et al., 2012), a highly diverse corpus of 312 million words covering a large variety of textual genres and Portuguese varieties. The written sub-part of the corpus consists of 310 million words, sampled from texts mostly after 1970 gathered from many different genres and domains such

---

[1] https://github.com/ottiram/MMAX2

as scientific papers, technical reports, literary works, newspaper texts, parliament transcriptions and judicial documents.

We extracted the sentences on the basis of a list of 40 Portuguese verbs that can express a modal meaning. We attempted to select equal sets of verbs that are associated to each type of modality. For example, the verbs *saber* 'know', *pensar* 'think', *crer* 'believe', *perceber* 'understand' and *julgar* 'judge' are generally associated with the epistemic meaning and therefore chosen to trigger epistemic modality, while the verbs *permitir* 'allow', *obrigar* 'oblige', *exigir* 'require', *conceder* 'allow', *deixar* 'allow' are usually associated with the deontic meaning. The modal verbs are used as a selection criterium to gather sentences containing at least one modal expression. However, the annotation of modality covers all modal elements present in the sentences, including nouns, adverbs and adjectives. We used the online interface for CRPC[2] to query for each verb lemma and retrieved the first 50 sentences from a randomly ordered list. We restricted the search query to cover only European Portuguese and excluded documents from Politics and Law to avoid formal language usage.

With this method, we were sure to have at least one trigger in the sentences that were selected. If more modal triggers were found in the sentence, they were annotated. This lead to a large set of modal cues than the 40 initial verbs. During the annotation, we didn't limit ourselves to verbs, and included other part of speech triggers: verbs, adjectives, nouns, adverbs (see section 3).

This method of selection of sentences for our corpus has some implications. First, querying for modal verbs that are associated with certain modal values influences the frequencies of occurring modal values in the corpus sample. Secondly, we chose to annotate single sentences and not full texts. Therefore, we might not have enough linguistic context to disambiguate between modal meanings in some cases.

The 1946 sentences were annotated with the annotation scheme described in section 3.
The annotated sentences are divided in four files. The number of tokens of the raw corpus is given in Table 1.

files annotated	tokens
mod_sentences_1.mmax	16257
mod_sentences_2.mmax	32511
mod_sentences_3.mmax	48060
mod_sentences_4.mmax	61725
Total	158553

Table 1. Raw corpus 1: tokens per file

2.2 The CRPC-Modal Corpus

The CRPC-Modal Corpus is composed of full-text files, extracted from the written subpart of the Reference Corpus of Contemporary Portuguese (Généreux et al, 2012). Most texts are news articles, but there are also a few fiction and scientific texts, and texts from varia. Contrary to Modality Corpus 1, where single sentences were extracted and annotated, in the CRPC--Modal Corpus, the annotator had access to the full text and could consider the larger context to disambiguate the modal values. We list below the full set of files. Fiction texts are marked with (F), scientific/technical texts with (S) and varia with (V). All remaining files are news articles. An asterisk signals that no modal value was found in the file (the files were separately annotated for different levels of annotation, such as events and discourse relations).

---

aefll001 (V)	dn81313	dn86102	dn86328	dn86707
ambi9006 (S)	dn81314	dn86103	dn86402	dn86708
banco001 (V)	dn81315	dn86104	dn86406	dn86711
bo86101	dn81316*	dn86105	dn86407	dn86713
bo86102	dn81317*	dn86106	dn86409	dn86715
bo86202	dn81318	dn86107	dn86410	dn86716
dn81101	dn81319	dn86108	dn86411	dn86717
dn81102	dn81320	dn86109	dn86412	dn86718
dn81103	dn81321	dn86110	dn86413	dn86722
dn81104	dn81322	dn86111*	dn86414	dn86725
dn81105	dn81323	dn86112	dn86416	dn86726
dn81106	dn81401	dn86113	dn86417	dn86727
dn81107	dn81402	dn86114	dn86418	dn86729
dn81108	dn81403	dn86115	dn86419	dn86731
dn81109	dn81404	dn86116	dn86421	dn87101
dn81111	dn81405	dn86201	dn86422	dn87102
dn81112	dn81406	dn86202	dn86502	dn87104
dn81113	dn81407	dn86203	dn86503	dn87106
dn81114	dn81408	dn86204	dn86507	dn87107
dn81115	dn81409	dn86205*	dn86508	dn87108
dn81116	dn81410*	dn86206	dn86509	dn87109
dn81117	dn81411	dn86207	dn86510	dn87110
dn81201	dn81412	dn86208	dn86511	dn87112
dn81202	dn81413	dn86209	dn86513	dn87113
dn81203*	dn81414	dn86214	dn86514	dn87114
dn81204	dn81415	dn86217	dn86515*	dn87115
dn81205	dn81416	dn86219	dn86517	dn87201
dn81206	dn82417	dn86220	dn86518	dn87202
dn81207	dn82418	dn86221	dn86519	dn87203*
dn81208	dn82419	dn86222	dn86520	dn87205
dn81209	dn82501	dn86223	dn86522*	dn87206
dn81210	dn82502	dn86301	dn86523	dn87207*
dn81211	dn82509*	dn86302	dn86601	dn87208
dn81212	dn82517	dn86303*	dn86603	dn87209
dn81213	dn82601	dn86304	dn86604	dn87210
dn81214	dn82602	dn86305	dn86605*	dn87211
dn81215	dn82606	dn86306	dn86606*	dn87213*
dn81216*	dn82608	dn86307	dn86607	dn87215
dn81217	dn82701	dn86308	dn86610	dn87216
dn81218	dn82703	dn86309*	dn86612	dn87218
dn81219	dn82706	dn86310	dn86613	dn87219
dn81301	dn82708	dn86313	dn86614	dn87220
dn81302	dn82710	dn86314	dn86616	dn87221
dn81303	dn82711	dn86316	dn86618	dn87222
dn81304	dn82713	dn86317	dn86619	dn87223
dn81305	dn82715	dn86318	dn86623	dn87301
dn81306	dn82716	dn86319	dn86624	dn87302
dn81307	dn82717	dn86320	dn86625	dn87305
dn81308	dn82719	dn86321	dn86701	dn87307
dn81309	dn82720	dn86322	dn86702	dn87312
dn81310	dn82721	dn86324	db86703	dn87313
dn81311	dn82723	dn86325	dn86705	dn87314
dn81312	dn86101	dn86326	dn86706	dn87315

dn87316	ex88105	id92102	pu92201	pu92214
dn87318	ex88107	maced1 (S)	pu92203	pu92215
dn87319	ex88420	mcarv1_1 (F)	pu92205	pu92217
dn87323	ex88601	mcarv1_2 (F)	pu92210	ruipi1_1 (S)
ex88102	geo9_1 (S)	mcarv1_3 (F) (F	pu92211	text1art (V)
ex88104	id92101	pu92101	pu92212	

3. Annotation Scheme

3.1 General annotation steps

1. Do the annotation sentence per sentence;
2. Do not change or alter the corpus sentence, leave all spelling errors and punctuation as it is;
3. Start by identifying the trigger and the modal value that it expresses, then annotate the other elements.
4. We explicitly annotate ambiguity and polarity (see section 2.4.4.1).
5. We take a "min-max strategy", in line with the approach of Farkas et al (2010). For the trigger, we only annotate the minimum, the smallest possible unit like "desejo", "precisarem" or "tem de". For the target, we annotate maximally and include all relevant parts, such as adjuncts. For the sources, we annotate full noun phrases (or verbs, in case of null subjects).
6. In discontinuous elements, we always keep the order of the text.

3.2 The full scheme and the list of modal Values

The following table shows the components and attributes of the scheme.
Each component will be discussed in this manual.

Components	Attributes	
Trigger	Polarity	
	Polarity_cue	
	Ambiguity	
	Modal type	
	Comment	
	Modal values	*Modal sub-values*
	Epistemic	Possibility; Probability; Knowledge; Belief
	Deontic	Obligation; Permission; Necessity
	Participant-internal	Necessity; Ability; Volition
	Evaluation	
Target	Polarity	
Source of the modality		
Source of the event mention		

The following table describes each of the seven different modal values that we distinguish, and their subvalues.

Values	Subvalues	Definition
Epistemic	Knowledge	The speaker presents his or someone else's knowledge or expresses some degree of understanding about something
	Belief	The speaker expresses his or someone else's belief in something
	Possibility	The speaker presents a situation that he or someone else is viewing as possible
	Probability	The speaker presents a situation that he or someone else is viewing as probable
Deontic	Permission	The speaker or participant allows someone or himself to do something, or something to happen
	Obligation	The speaker or participant requires someone or himself to do something
	Necessity	The speaker presents a situation has necessary
Participant-internal	Ability	The speaker expresses his own or someone else's personal capacity to do or to achieve something
	Necessity	The speaker expresses his own or someone else's personal necessity
	Volition	The speaker expresses his or someone else's wishes, hopes or wants
Evaluation		The speaker presents his or someone else's evaluation of propositions

Table 2. Description of modal values.

3.3 What we do not consider as modal

⚔ We do not annotate declaratives and evidentials, since they do not express any modal value that we consider. We consider that declaratives as unmarked or neutral within the epistemic modal system.

⚔ We do not annotate aspect. Verbs like *passar a, acabar, continuar, começar* signal specific phases of the temporal structure of a situation, not modality.

⚔ We do not annotate tense. So, we don't tag the past tense (although it provides certainty about the realization of an event), nor future (possibility) nor conditional (unless there is a conjunction introducing the conditional clause that we can consider a trigger).

3.4 Components

Here we discuss the annotation of each component in detail and present some difficult cases.

3.4.1 Trigger

3.4.1.1 Definition

The trigger is the word or string of words that expresses modality (Baker et. al., 2010). In the examples provided in this manual, the trigger is underlined and the other components are rendered in italics.

1.   If the trigger is:

    1.1. a **modal verb**, we annotate the modal verb as trigger. In case of reflexive verbs, we only mark the verb and not the reflexive.

(1)   Mudaram a exposição do livro - aplicando técnicas de "marketing" e de "merchandising" - e <u>tentaram</u> chamar clientes que não iam às livrarias tradicionais.
     Trigger: tentaram
     Target: chamar clientes que não iam às livrarias tradicionais

    1.2. a **noun** and it is part of a Nominal Phrase (NP), we only annotate the head noun;

(2)   Ao optar por fumar agora estou a influenciar o meu futuro <u>desejo</u> de fumar.
     Trigger: desejo
     Target: de fumar

    1.3 an **adverb,** we tag as trigger only the adverb and only when it has scope over an event;

(3)   Dizem-me alguns que não devo ligar ao assunto, pois, <u>possivelmente</u>, as sete dezenas de linhas que o provedor do leitor dedicou ao assunto passaram despercebidas.
     Trigger: possivelmente
     Target: as sete dezenas de linhas que o provedor do leitor dedicou ao assunto passaram despercebidas

    If the adverb is preceded by another adverb, as for example *muito provavelmente*, we consider as trigger only the modal adverb *provavelmente*.

    1.4 an **adjective**, when part of a verbal phrase we do not include the auxiliary since it does not carry any modal information. Adjectival triggers are annotated only when they are part of a predicate (an event).

(4)   Foi <u>difícil</u> mas se tenta dar às pessoas a sensação de que as compreende, é mais fácil.
     Trigger: difícil

    1.5 a **conjunction** that allows an hypothetical or counterfactual reading of the target, we annotate the conjunction;

(5)   <u>Se</u> o fogo se atear no templo destes deuses de madeira dourada ou prateada , os seus sacerdotes procuram fugir pondo –se a salvo
     Trigger: se
     Target: o fogo se atear no templo destes deuses de madeira dourada ou prateada

1.6 an **auxiliary verb in the future compound tense**, with the modal value of epistemic_probability or epistemic_possibility, we annotate the auxiliary as the trigger;

(6) A culpa terá sido da publicidade que se não fez.
Trigger: terá
Target: A culpa@sido da publicidade que se não fez.

2. The trigger does not include **negative particles**. We view negation as a separate phenomenon that interacts with modality. We deal with negation in the component *Polarity*.
(7) Onde arranja o dinheiro não <u>sei</u>, nem interessa.
Trigger: sei
Target: onde arranja o dinheiro

3. When the trigger is a modal semi-auxiliary, **prepositions and conjunctions** are never included in the trigger but are included in the target. The only exceptions to this rule are *ter de, ter que* and *haver de.* In (8), the trigger is *ter de* and the preposition *de* is included in the trigger, while in (9) the trigger is *capaz* and the preposition is included in the target.

(8) <u>Tenho de</u> ir.
Trigger: tenho de
Target: ir

(9) nenhuma força internacional será *capaz* de fazer cumprir os acordos
Trigger: capaz
Target: de fazer cumprir os acordos

4. When the verb is in the **passive** construction, we don't include the verb "ser" in the trigger.

5. When the verb is in a **compound form**, we only annotate the past participle form:

(10) Ora , qualquer mediano observador tinha <u>percebido</u> há muito que havia uma guerra surda contra o comissário Cardoso e Cunha
Trigger: percebido
Source of the modality: qualquer mediano observador
Target: há muito que havia uma guerra surda contra o comissário Cardoso e Cunha

6. If the trigger is a **multiword expression** or a semifixed multi-word expression with non-compositional meaning, we include all the elements in the trigger, even when discontinuous, except the final preposition. For instance, in (11) the trigger is "correu-se … o risco". Another example is "reconhece … o direito", in (12).

(11) <u>correu</u> se no entanto <u>o risco</u> de chegar a uma situação análoga à que hoje caracteriza os países recentemente industrializados de outros continentes
trigger: correu@o risco[3]
target: de chegar a uma situação análoga à que hoje caracteriza os países recentemente industrializados de outros continentes

(12) A Constituição <u>reconhece</u> às comissões de trabalhadores e às associações sindicais <u>o direito</u> de participar na elaboração da legislação do trabalho

---

3　　　Note that we use @ to signal non-continuous strings in the manual, in section 3 we show how to handle this in the annotation software.

Trigger: reconhece o direito
Modal value: deontic_permission
Polarity: positive
Target: às comissões de trabalhadores e às associações sindicais @ de participar na elaboração da legislação do trabalho

7. If the predicate is a complex unit of **support verb** (or light verb) followed by a noun or adjective, the verb is included in the trigger because both contribute to the meaning of the predicate.

    (13)   Dar a possibilidade / ter possibilidade
                Ter tendência

In case of ellipsis of the support verb, the noun is annotated. In (14), *dar capacidade* 'to give capacity' implies that some entity will have capacity (*ter capacidade*). We annotate the noun *capacidade*.

    (14)   Dão às Câmaras grande capacidade de intervenção neste domínio.
                Trigger: capacidade
                Modal value: participant-internal-capacity
                Source: as Câmaras
                Target: de intervenção neste domínio

8. If the trigger is split into two parts by, for example, an adverb, we do not consider the adverb as part of the trigger.

    (15)   Temos amanhã de limpar a casa.
                Trigger: temos@de

9. When one sentence contains two or more triggers, we annotate them separately.

    (16)   Por mais forte que uma pessoa tente ser, é muito difícil continuar com o mesmo estado de espírito com que se estava anteriormente.
                Trigger: tente
                Target: ser

                Trigger: difícil
                Target: continuar com o mesmo estado de espírito com que se estava anteriormente

    (17)   É este um vício que sempre atinge os míseros: nunca conseguir crer na felicidade!
                Trigger: conseguir
                Target: crer na felicidade

                Trigger: crer
                Target: na felicidade

10. Punctuation that marks a clause ending is not annotated as part of the trigger, except for imperative clauses. When the clause is an imperative, we consider as trigger the exclamation mark (!). In an order like (18) (modal value deontic obligation), the verb also has the imperative inflection, and we could have included the verb in the trigger. However, in some imperatives the verbs do not carry the imperative inflection. We chose to have one method of annotation: always the exclamation mark as single trigger and the target is the verb in the imperative form and its complements.

    (18)   João, senta!

Trigger: !
Target: *João, senta!*

11. The trigger may be an expression that weakens the modal value of the speech act:

    (19)   agradecemos que nos devolva
                Trigger: agradecemos
                Modal value: deontic_obligation
                Target: que nos devolva

3.4.1.2 Features of the modal event

As the trigger is the central component in the modal expression, we attribute the specifications of the type of the modality to the trigger. We specify for each trigger the following features:
- ⚔ Modal value
- ⚔ Polarity
- ⚔ Ambiguity

3.4.1.2.1 Modal value

The following table illustrates lemmas and constructions that express the modal values. In the last column, the source of the modality is expressed as X and the target as Y, to illustrate different types of constructions of the lemma (not exhaustively).

modal value	example of lemmas	exemples of constructions
epistemic_knowledge	saber	X <u>sabe</u>, <u>sabe</u>-se
epistemic_belief	considerar evidente levar a crer defender	X <u>considera</u> é <u>evidente</u> que algo leva (X) a crer Y X <u>defende</u> Y (defender um ponto de vista)
epistemic_possibility	possível possivelmente possibilidade poder possibilitar permitir eventual eventualidade atribuível parecer a se quando	é <u>possível</u> que Y <u>possivelmente</u> Y <u>possibilidade de</u> Y <u>pode acontecer Y</u> algo <u>possibilita</u> Y algo <u>permite</u> Y <u>situação eventual</u> <u>eventualidade</u> de Y atribuível Y (pode ser atribuído) parece Y (isso parece derivar disso) a serem aí descarregados = se forem aí descarregados) se Y (conditional) quando Y (quando houver...)
epistemic_probability	provável probabilidade dever suscetível prever fazer prever previsto	é <u>provável que Y</u> <u>probabilidade de Y</u> <u>deve acontecer Y</u> algo <u>suscetível</u> de acontecer X prevê Y (alguém prevê algo; o plano prevê que) algo <u>faz prever</u> Y (causative) previsto Y; Y está previsto
participant-internal_necessity	carecer	X <u>carece</u> de Y

	precisar necessidade	X precisa de Y X tem <u>necessidade</u> de Y a <u>necessidade</u> de X por Y
participant-internal_ability	<u>capaz</u> <u>capacidade</u> <u>apto</u> <u>aptidão</u> <u>ter competência</u> <u>conseguir</u>	<u>X é capaz de Y</u> <u>capacidade de X para Y</u> X apto a Y aptidão de X para Y X tem competência para Y X conseguiu Y
participant-internal_volition	<u>desejar</u> <u>desejo</u> <u>vontade</u> <u>querer</u> ânsia	<u>X deseja Y</u> <u>desejo de X por Y</u> <u>X tem vontade de Y; a vontade de Y</u> <u>X quer Y</u> <u>X tem ânsia de Y; X, na ânsia de Y</u>
deontic_permission	poder permitir proibir (neg) impedir (neg) permitido aprovação dar a possibilidade	alguém <u>pode</u> Y X <u>permite</u> Y X <u>proíbe</u> Y (negative polarity) X impede Y (negative polarity) é <u>permitido</u> Y <u>aprovação</u> de Y (por X) X dá a possibilidade de Y
deontic_obligation	dever ter de obrigar exigir exigido impor imposição caber agradecer a	alguém <u>deve</u> Y alguém <u>tem de</u> Y, X obriga a Y X <u>exige</u> Y é <u>exigido</u> Y X <u>impõe</u> Y imposição de Y (por X) <u>cabe</u> Y (cabe a alguém fazer algo) agradecemos que Y a ter em conta (que se deve ter em conta)
deontic_necessity	<u>necessidade</u> <u>necessário</u> <u>necessariamente</u> <u>preciso</u> <u>fundamental</u> levar importar dispensável	<u>necessidade de Y</u> <u>necessário Y</u> <u>necessariamente Y</u> é <u>preciso</u> Y é <u>fundamental</u> Y X levou a Y (o sismo levou milhares a fugir) <u>importa Y (importa lembrar que)</u> <u>dispensável Y (negative polarity = not necessary)</u>
evaluation		

Table 3. Examples of lemmas annotated for modal values

3.4.1.2.2 Polarity

Polarity is the component used to mark if there is negation scoping on the modal value. The polarity values are positive and negative. We do not annotate the polarity of the full clause, we only look at the trigger and its modal value.

Negative polarity may be expressed by an isolated element, as the negative adverb; or a

trigger may have intrinsically negative polarity, e.g. impossível (negative polarity of epistemic_possibility), xemple (negative polarity of deontic_permission or epistemic_possibility).

1. We only mark the polarity of the modal value and not that of the entire clause or sentence: in (20), the polarity of the modal value expressed by the trigger *obrigou* is positive since the negation is contained in the target of the modality trigger:

(20) O João <u>obrigou</u> *a Maria a **não** entrar na casa da amiga com aquela raiva* para evitar que a discussão acabasse em luta.
Trigger: obrigou
Target: a Maria a não entrar na casa da amiga com aquela raiva
Source of the event mention: sp/wr
Source of the modality: O João
Modal value: deontic_obligation
Polarity: positive

2. When there are two triggers, the negative particle affects the first trigger.

3. If both are affected by the negative polarity, both the triggers are marked with negative polarity.

(21) É este um vício que sempre atinge os míseros: **nunca** <u>conseguir</u> <u>crer</u> *na felicidade*!
Trigger: conseguir
Target: crer na felicidade
Source of the event mention: sp/wr
Source of the modality: conseguir
Modal value: participant-internal_ability
Polarity: negative

Trigger: crer
Target: na felicidade
Source of the event mention: sp/wr
Source of the modality: crer
Modal value: epistemic_belief
Polarity: negative

4. Although we mark the negative polarity over the trigger, in some cases there are other negative elements in the clause and the combination of the two negative cues creates a positive polarity of the modal value. In those cases, although there is a negative cue over the trigger, we annotate as positive polarity, as in (22) and (23).

(22) E *não há nenhum combate* em que os intervenientes *não* <u>procurem</u> perceber e tornar claro o que é que os satisfará para estabelecerem a paz.
Trigger: procurem
Modal value: Participant-internal volition
Polarity: positive

(23) Dois peritos médicos para os quais a esquizofrenia não oferecia qualquer dúvida
Trigger: dúvida
Modal value: Epistemic_belief
Polarity: positive

5. Two negative elements related to the source of the modality have a negative polarity interpretation in cases such as:

(24) Não existe nenhuma lei que permita à PT cortar o serviço telefónico"
Trigger: permita
Polarity: NEG (nada permite à PT...)

6. Two negative cues may change the modal value of a trigger: in (25) the trigger "poderia" would be annotated as epistemic_possibility but, due to the negative polarity of the target (deixar de), the global value is deontic_necessity with positive polarity.

(25) O Banco não poderia deixar de o convidar
Trigger: poderia
Modal value: deontic_necessity
Polarity: positive
Target: deixar de o convidar

7. A word with a negative suffix that indicates negative possibility may be annotated with positive polarity if the value in the context is not epistemic_possibility. For instance, *indispensável* or *imprescindível*.

(26) Achamos indispensável
Trigger: indispensável
Modal value: deontic_necessity
Polarity: positive

8. When the polarity is negative, the cue for the polarity is manually written down in the attribute "negative-cue". It is usually another lexical element in the context. In the examples below the trigger is underlined and we identify the lexical element with negative polarity as the cue for the polarity:
não <u>devem</u> fazer isso: negative polarity, polarity_cue = não
incapaz de reconhecer: negative polarity, polarity_cue = incapaz

In some cases, the negative polarity is due to a prefix of the trigger itself, and in that case we identify the prefix as the cue:
<u>incapaz</u> de: negative polarity, polarity_cue = in-

9. The negative cue may have scope over a second modality trigger. In the example below, "parecia" is considered as epistemic belief with polarity positive and the negative cue "ninguém" has scope over the second modal trigger "poder". It could be paraphrased as "parecia que ninguém podia prever"

Ninguém parecia poder prever
Parecia: trigger, epistemic_belief, positive polarity (it expresses a belief of the sp/wr)
Poder: trigger, participant-internal capacity, negative polarity

3.4.1.2.3 Ambiguity

Ambiguity is marked when multiple modal values apply at the same time and none of these values is clearly more natural than the others. We may annotate more than two modal values.

Example: Ambiguity of "parecer": epistemic_possibility or epistemic_belief

(27) A incomodidade atribuível a um estímulo sonoro que se identifica em relação ao ruído de fundo <u>parece</u> derivar diretamente do caracter intrusivo daquele estímulo…"

13

"é possível que derive diretamente do caracter intrusivo…"

"acredito com base em provas que derive diretamente do caracter intrusivo…".

3.4.2 Target

Definition: The target is the expression in the scope of the modal trigger.

We annotate the target maximally and we try to keep one unit if possible to avoid dis-continuity.

The target can be a) a noun phrase; b) a subordinate clause c) a verbal phrase
   a) Noun phrase: tag the whole nominal phrase, including articles, adjectives or restrictive clauses;

   (28) Comerciantes da Guarda <u>exigem</u> *a reabertura da Rua do Comércio.*

   b) Subordinate clause: tag the whole subordinate clause, including the conjunction or preposition introducing it;
   (29) Ele <u>acha</u> *que a Igreja fica favorecida pelo facto de estar aliada ao poder político.*

   c) Verbal phrase: the complements of the verb should be included.
   (30) No terreno das indústrias da cultura – cinema , livro , televisão - , <u>arriscamo</u> - *nos a ser dominados pelo mercado americano* .
   Trigger: arriscamo
   Target: No terreno das indústrias da cultura – cinema , livro , televisão - @-nos a ser dominados pelo mercado americano
   Modal value: epistemic_possibility

However, we do not include the separation commas, adverbial phrases, partial phrases or other types of clauses.

Difficult cases:
   1. Prepositions and conjunctions following the trigger are included in the target, except for the verb *ter de* and *ter que*;

   2. Discontinuous target with syntactic relation:

   (31) Logo a seguir, ensaiou o primeiro «drible», quando *lhe* <u>perguntaram</u> *se condenava as relações extra-conjugais*.
   Trigger: perguntaram
   Target: lhe@se condenava as relações extra-coniugais

Adjuncts between complements are considered as part of the target to avoid other cases of discontinuity.

   (32) <u>Autorizei</u> *o meu filho ontem a ir ao parque*.
   Trigger: autorizei
   Target: o meu filho ontem a ir ao parque

   3. When the target is not lexically expressed, it is not annotated:

   (33) Felizmente ninguém na plateia respondeu : "<u>Sabemos</u> sim , senhor Little Axe …"
   Trigger: sabemos

Target: Null

4. Target in imperatives: the target is verb in the imperative and its complements:

(34) *Olha a casa !*
      Trigger: !
      Target: Olha a casa !

5. When there is a pronoun in an element, and the sentence also contains the full name where the pronoun refers to, we keep only the pronoun inside the element and not mark up the referent. Co-reference resolution is a task in itself and should not be part of our annotation. In relative clauses, we tag the element "que" as part of the target (or other component) instead of the NP it is coreferent with.

(35) Espero por essa reunião e <u>espero</u> *que ela seja suficientemente abrangente, alargando -se a vários assuntos* ...
      Trigger: espero
      Target: que ela seja suficientemente abrangente, alargando -se a vários assuntos …
      Comment: we do not include 'essa reunião' in the target with the pronoun.

6. If the complement of an element of the target is not expressed inside the target but can be recovered from another segment of the sentence, we only annotate the target itself, as in the example in which we tag as target only *sensações tácteis parecidas*, and not *com os livros de papel* (which can be inferred to be the complement of *parecidas*).

(36) Parecem-se com os livros de papel, querem <u>permitir</u> *sensações tácteis parecidas*
      Trigger: permitir
      Target: sensações tácteis parecidas

7. A target is never overlapping with its trigger.

8. The target may include the subject in cases of epistemic possibility, epistemic probability, deontic permission and deontic obligation, deontic necessity.

(37) O João pode chegar tarde
      Trigger: pode
      Modal value: epistemic possibility
      Target: o João@chegar tarde
      Source of the modality: sp/wr
      Source of the event: sp/wr

(38) O João tem de acabar o trabalho
      Trigger: tem de
      Modal value: deontic obligation
      Target: o João@acabar o trabalho
      Source of the modality: sp/wr
      Source of the event: sp/wr

(39) "Esta prática necessita de ser realizada com cuidados especiais"
      Trigger: necessita
      Modal value: deontic_necessity

15

Target: esta prática @ser realizada com cuidados especiais.
Source of modality: sp/wr

3.4.3 Source of the event mention

Definition: The source of the event mention is the producer of the event mention.

It is, in most cases, the speaker or writer (sp/wr). We only mark this source when there is a reference to the entity in the sentence. We mark a noun phrase (40) or, in cases of null subject, we mark the verb that refers to the source of the event mention (41). In (40), there is a source of an event mention (O João) but the speaker or writer is also a source of enunciation, that reports the content of what João said. The speaker or writer is assumed by default and not marked. When another source besides the speaker or writer is explicitly mentioned in the text, we annotate as source of the event mention.

(40) *O João* contou que quando foram à praia *a Maria* não <u>deixou</u> *que os putos metessem os pés na água*.
Trigger: deixou
Target: que os putos metessem os pés na água
Source of the event mention: o João
Modal value: deontic_permission

(41) "Não *podes* sair!", disse.
Trigger: podes
Target: sair
Source of the event mention: disse
Modal value: deontic_permission

3.4.4 Source of the modality

Definition: The source of the modality is the holder of the modality.

The element tagged as source are noun phrases, and we include adjectives, prepositional phrases, and also restrictive clauses as in (44).

(42) *O ministério*, ao anunciar que retirará a classificação de superior aos nove institutos, oferecendo-lhes a possibilidade de concederem licenciaturas pelo politécnico, <u>procura</u> repor a lei, minimizando os danos aos alunos que optaram por esses institutos privados.

(43) O governante português afirmara que *nenhuma força internacional* será <u>capaz</u> de fazer aplicar os acordos se as partes não os cumprirem.

(44) *As diferentes abordagens que têm sido utilizadas no estudo do cérebro* <u>precisam</u> de um denominador comum, declarou à revista britânica new scientist guy mckhann[4], da universidade johns hopkins.

Table 4 presents individual definitions of what we consider as the source of the modality for each modal value. This kind of explicitness has proved to be of great help for the annotators, since the source of the modality can be defined in different ways depending on the modal value and on the context.

---

4   In the corpus the first name Guy Mckhann is written with small letters and not with capital letters.

Epistemic_knowledge	Source of the modality: who has the knowledge
epistemic_belief	Source of the modality: who has the belief
epistemic_possibility	Source of the modality: who considers that something is possible / what makes something possible
epistemic_probability	Source of the modality: who considers that something is probable
participant-internal_necessity	Source of the modality: who has the necessity
participant-internal_ability	Source of the modality: who has the capacity or ability
participant-internal_volition	Source of the modality: who wants to
deontic_permission	Source of the modality: who/what establishes the permission
deontic_obligation	Source of the modality: who/what establishes the obligation
deontic_necessity	Source of the modality: no overt source of what makes something necessary
evaluation	Source of the modality: who evaluates something

Table 4. Definition of the source of modality for each modal value

Difficult cases:

1. If the source of the modality is a pronoun and there is no other referential expression, tag the pronoun in the clause. If there is a pronoun and an NP that is co-referent with this pronoun inside the same clause, tag the NP. If it is co-referent to an NP outside the clause, tag the pronoun. For instance, in (45), the possessive pronoun *seu* refers to the NP *Catarina*, and we tag *Catarina* as the source of the modality because it occurs in the same clause as the pronoun and it is more informative.

   (45) *Catarina* nunca confessou o seu <u>desejo</u> *pela amiga* e esta sempre fez de tudo para não o saber.
   Trigger: desejo
   Target: pela amiga
   Source of the modality: Catarina

2. When there is no referential expression to the source of the modality in the sentence and there is no pronoun that recovers the source in the sentence, annotate the source outside the sentence if there is an expression.

3. If there is no referential expression in the text, inside or outside the sentence, and there is a null subject: tag the verb as the source of the modality, when the subject of the verb is the source of the modality.

   (46) *Querem permitir sensações tácteis.*
   Trigger: querem
   Source of the event mention: sp/wr
   Source of the modality: querem

4. Null subject: ambiguity
   If the subject is null and the verb carrying the modality has no markup of number and person, we tag as source of the modality another element that refers to the source which carries those marks. However, sometimes there is too much ambiguity or uncertainty to determine a suitable referent. In the next example, the trigger *obrigando* does not have information about who is obliging. We could consider that the null subject of the verb *ameaçaram* is also the source of the modality, but it could also be another factor that is the subject of *obrigando*. As we are not sure, we do not tag a source of modality in this case.

   (47) Descontentes com a forma como fora organizado o cerimonial , ameaçaram mesmo faltar à assinatura de uma nova cópia da acta fundadora - - semelhante a uma que já figura na parede da pequena sala onde foi fundado o OS - - , <u>obrigando</u> os responsáveis do SPD e da Fundação Friedrich Ebert a

alterarem o que estava previsto.
Trigger: obrigando
Source of the modality:

5. null subject_impersonal: tag source of the modality on *se* and don't include *se* in the trigger (in cases where the subject of the verb is the source of the modality):

(48) <u>Sabe</u> *-se apenas que não será a Globo*
Trigger: Sabe
Target: apenas que não será a Globo
Source of the event mention: sp/wr
Source of the modality: -se

In certain contexts, the element *-se* is related to the main verb and not to the modal verb. In (45b) "se" is related to the verb "entender" and not to the verb *dever*, and there is no source of the modality of *dever*.

(49) o certo é que a determinação do que como tal se <u>devia</u> entender prestava -se inevitavelmente a dúvidas.

6. null subject, and source is pronoun: tag the source of the modality on the pronoun and don't tag the pronoun in the trigger;

(50) <u>Apetece</u> *-me*
Trigger: apetece
Target:
Source of the event mention: -me
Source of the modality: -me

In this case, if the subject is present, the source is the same, because the grammatical subject (*morangos*) is the target:

(51) <u>Apetecem</u> *-me morangos.*
Trigger: apetecem
Target: morangos
Source of the event mention: -me
Source of the modality: -me

7. passive sentences: the source of the modality is the agent (in cases where the subject of the active sentence is the source of the modality)

(52) Lembra o acórdão que "não podia no despacho recorrido, subscrito por Ricardo Cardoso, nem pode agora, dar-se qualquer relevância às circunstâncias que, noutro processo em que o arguido já foi condenado por decisão transitada, foram <u>consideradas</u> existentes pelo *Tribunal de Execução de Penas* para conceder ao arguido a liberdade condicional.
Trigger: consideradas
Source of the event mention: sp/wr
Source of the modality: Tribunal de Execução de Penas

8. imperatives: the source of the event mention is the same as the source of the modality = sp/wr

9. With certain modal values (e.g., epistemic_possibility) and lemmas (e.g. *permitir*), the source of the modality may be a concrete object (X makes Y possible):

(53) *A janela* <u>permitia</u> *ter muita luz na sala.*
Trigger: permitia

Source of the event mention: sp/wr
Source of the modality: a janela

10. In contexts of deontic necessity, there is typically no specification of the source of the modality.

    (54)   É necessário que todos colaborem
               Trigger: necessário
               Modal sense: deontic necessity
               Target: todos colaborem
               Source of the modality: -
               Source of the event: sp/wr

11. However, in contexts of deontic necessity, a source may be annotated, when it refers to what makes something necessary:

    (55)   Processo de adaptação necessário à correção de carências
               Trigger: necessário
               Modal value: deontic_necessity
               Target: processo de adaptação
               Source: correção de carências

    (56)   Resíduos que requerem equipamento
               Trigger: requerem
               Modal value: deontic_necessity
               Target: equipamento (= comprar/usar equipamento)
               Source: que (=resíduos)

12. In contexts of epistemic possibility with "é possível que", the source of the modality is the speaker/writer.

We illustrate different types of contexts, with and without an explicit source of the modality for each modal value, In the following tables. In the examples, we underline the trigger, mark the source of the modality in italics, and mark the target in bold.

Deontic_obligation	
Source animate	(57) Os estabelecimentos de 1a classe , conforme tabela , são aqueles *que* por_ a natureza de fabricação , ou importância , <u>impõem</u> **que seja prestada particular atenção a_ os aspectos técnico-funcionais e de localização .**
Source non-animate	(58) *A dificuldade de detectar quais os que estão atingidos por_ a poluição* <u>exige</u>[5] **que todos os moluscos sejam depurados antes de consumidos .**
no source	(59) é a_ as autarquias **que** <u>cabe</u> **o papel fundamental em_ a resolução de_ estes problemas**

---

[5] We consider that "exigir" has strong modal force and is annotated as deontic_obligation, even with non-animate sources.

	(60) **Os Parques Industriais** <u>devem</u> **ser definidos tendo em conta os ventos dominantes , cursos de água , características de_ os solos , etc .** (61) **a qualidade mínima** <u>exigida</u> **a_ as águas balneares** (source = indeterminado, todos) (62) **A autarquia** (...) , <u>devendo</u> **exigir de_ as indústrias uma separação clara de_ os diferentes tipos de resíduos existentes** . (source: wr/sp)

Deontic_permission	
Source	(63) *Regulamentação a emitir por_ os Municípios* , <u>proibindo</u> **os despejos de resíduos em_ as águas e de lixos em_ locais não controlados.** [negative polarity of trigger] (64) criação de legislação adequada *que* <u>permitisse</u> **actuar**
no source	(65) a <u>aprovação</u> **de_ as condições de salubridade , higiéne , segurança , comodidade e técnico-funcionais** . (66) A correcta gestão de_ os resíduos sólidos urbanos , em particular evitando as descargas selvagens , <u>impedindo</u> **as descargas de resíduos industriais em_ os aterros municipais**

Deontic_necessity	
Source	(67) **Processo de adaptação** <u>necessário</u> *à correção de carências* (68) **uma aproximação a_ a natureza** <u>necessária</u> *a_ o equilíbrio físico e intelectual* (69) **as medidas** <u>necessárias</u> *para minimizar os prejuízos de_ a frequência turística*
no source	(70) **Dois grandes grupos de orientações** são essencialmente <u>necessários</u> : **reforçar a observação de_ o meio marinho e prevenir e combater as poluições , sendo este último ponto centrado em_ a poluição de_ as águas continentais e em_ a eliminação de_ a que é gerada por_ as actividades industrial e agrícola** . (71) **A prevenção começa** <u>necessariamente</u> **por um bom conhecimento e uma vigilância permanente de_ o meio marinho** . (72) é <u>preciso</u> **manter uma vigilância geral sobre a água , sedimentos e seres vivos (moluscos , crustáceos ,**

	peixes) , tendo especialmente em atenção a qualidade de_ as águas de_ as praias , de_ os portos de abrigo e de_ as zonas mais sensíveis de_ o ponto de vista ecológico (estuários , rias , etc .) .

epistemic_knowledge	
Source	(73) Em St . Louis , no Missouri , por exemplo , *os habitantes* <u>sabem</u> de antemão **qual vai ser a qualidade do ar no dia seguinte** .
no source	(74) Está , em_ o entanto , <u>comprovado</u> **que o controlo de_ a poluição atmosférica é possível** .

epistemic_belief	
Source	(75) **os répteis** , quase sempre <u>considerados</u> **perigosos** por_ *os turistas* (76) <u>*Consideramos*</u> **urgente analisar todos os contratos de concessão (Bares , Fotocópias , Livraria , Papelaria) , colocando todas estas informações a_ o dispor de todos os estudantes de_ esta Faculdade** .
no source	(77) torna -se <u>evidente</u> **a necessidade de dispôr de um local ou aterro**

epistemic_possibility	
Source	(78) *A janela* <u>permite</u> **passar a luz** (79) Para *Félix e Silva* «**tudo** <u>pode</u> **acontecer** "Félix e Silva": source of modality and source of event
no source	(80) **A utilização cada vez maior de adubos** <u>pode</u> **originar um excesso de nitratos e fosfatos em_ a água , que favorecem o crescimento descontrolado de algas e plantas aquáticas , e asfixiar a vida marinha em_ algumas zonas mais sensíveis como estuários e rios**. (81) onde **o cidadão** <u>possa</u> **entregar líquidos usados altamente poluentes , como óleos , solventes , tintas etc.** (82) <u>eventual</u> **melhoria de_ as condições de_ os aterros sanitários camarários , com particular vigilância sobre as águas de lixiviação** .

epistemic_probability	
Source	(83) *A entrada em vigor de um novo quadro legislativo de_ a qualidade de_ a água e de um novo modelo institucional de gestão de_ os recursos hídricos* , <u>faz prever</u> **estarem reunidas as condições para que a situação actual seja a curto prazo invertida** . (84) Não *se* <u>prevendo</u> **que caia chuva em_ os próximos dias** *(*85) *A Comissão* tinha anteriormente <u>previsto</u> **um crescimento da força interna bruta**
no source	(86) **Os pontos onde os acidentes sejam mais** <u>susceptíveis</u> **de ocorrer** devem ser particularmente vigiados

participant-internal_ability	
Source	(87) Grande <u>capacidade</u> **de dispersão** de_ *as marés* . (88) *redes de vigilância meteorológica* <u>capazes</u> **de detectarem com antecedência furacões , tornados ou tempestades.**

participant-internal_necessity	
Source	(89) *Os estabelecimentos de 2a e 3a classes* não <u>carecem</u> **de aprovação de projectos** .

participant-internal_volition	
Source	(90) *qualquer visitante* espera e <u>procura</u> **encontrar a harmonia de_ as tradições e costumes que o ponha em contacto com a população local , a sua história e o seu modo de vida**
no source	(91) feita com a <u>intenção</u> **de separar materiais ferrosos para os encaminhar para a sucata** .

evaluation	
Source	(92) **Um " diálogo " constante** afirma *-se* assim como

	fundamental
no source	(93) **O controlo de poeiras em_ a armazenagem e transporte de carvão** é também <u>importante</u> **.**

Example where both source of the modality and source of the event are annotated:

> (94)   Mota Amaral adiantou que o seu Gabinete não assumiu , ainda , idêntico comportamento , por acreditar na possibilidade de se resolver o problema , mediante o diálogo com o Governo central .
> Source of the event: Mota Amaral
> Source of the modality: o seu Gabinete

3.4.5 Structural Ambiguity

With the verbs *dever* and *ter de* in some sentences there are two possible interpretations: according to the interpretation, a different modal value is expressed and different elements are tagged in the trigger. In the examples following, two annotations are possible. Between these interpretations, we usually annotate the sentence according to the interpretation shown in the second column, and we add in the comment field "SA" which stands for *Structural ambiguity* in order to keep this information.

(95) O João tem de acabar o trabalho

Trigger: tem de	Trigger: tem de
Target: acabar o trabalho	Target: o João@acabar o trabalho
Source of the event mention: sp/wr	Source of the event mention: sp/wr
Source of the modality: O João	Source of the modality: sp/wr
Modal value: participant-internal_necessity	Modal value: deontic_obligation
Polarity: positive	Polarity: positive

The annotation in the second column includes the subject in the target with value possibility, permission and obligation. If there is a 'se' impersonal in the sentence, it will be tagged as part of the target.

> (96)   em negócios não *se* <u>deve</u> *falar de futebol*
> Trigger: deve
> Target: em negócios@se@falar de futebol
> Source of the event mention: sp/wr
> Source of the modality: sp/wr
> Modal value: deontic_obligation
> Polarity: positive

4. Changes in modal annotation scheme v. 2.0 and 3.0

In the modal scheme 2.0 and 3.0 we have made changes to the list of modal values that were used in version 1.0 of our scheme. The changes are described in Ávila et al. (2015) and the changes are discussed below.

(i) We eliminate the value Epistemic_doubt, which translates into an Epistemic_possibility value with negative polarity (I doubt that this will happen ≈ maybe it is not possible that this will happen).

(ii) Direct interrogative sentences are syntactically marked as such and their annotation as modal instances in the EP scheme involved marking the entire sentence as trigger and target, what seems unnecessary. We removed the epistemic_interrogative value. Indirect interrogative sentences express a possibility value that is captured as such in the scheme.

(iii) Necessity is a concept that required further revision in our scheme. In spite of the difficulty in establishing whether a necessity is external or instead is an obligation established by the entities involved in the state of affairs, it is by no doubt important to be able to distinguish the clear-cut cases. With this in mind, we add the value Deontic_necessity (*é necessário que* 'it is necessary that').

(iv) The Participant-internal-capacity value is now Participant-internal-ability. We enlarge the sub-values of the Participant-internal category, so as to include Volition, which is best captured as a subcategory.

(v) Since Effort and Success are types associated to the Participant-internal_ability sub-value, we decided to leave them out.

(vi) Finally, we keep the category Evaluation, because it is interesting for studies of belief and opinion, although for this value we need more input and it should be revised in the future.

	Modal values	*Modal sub-values*
	Epistemic	Possibility; Probability; Knowledge; Belief
	Deontic	Obligation; Permission; Necessity
	Participant-internal	Necessity; Ability; Volition
	Evaluation	

5. Annotation Software MMAX2

MMAX2 is an annotation software tool developed at EML-research that can freely be downloaded from https://github.com/ottiram/MMAX2. The software is platform-independent and is written in Java. MMAX2 offers a visual interface to annotate sentences by marking textual strings and creating links between the marked elements. The annotations are stored as stand-off XML.

In this annotation scheme, we consider modality as an event that has several marked elements ("markables") that participate in the modal event like trigger, target and source of modality and event. We say that all these markables belong to the same modal event, which we call here a "set". The trigger markable has some specific features to be filled in: the modal value and its polarity. It also offers a text box for specifying ambiguity and additional comments.

5.1 Markables

In MMAX2 the basic element is called a "markable", it is just a piece of text that you marked. A markable is usually a textual string, but it can also consist of two discontinuous text parts. Markables can theoretically stretch over different sentences, they can overlap with each other and it can even happen that two markables can cover exactly the same piece of text.

We currently define 4 possible types of markables that each can only occur one or zero times in one modal event:

⌇ Trigger

⌇ Target

⌇ Source of the modality

⌇ Source of the event mention

The element trigger is the element that carries most of the information about the modal event and therefor this markable has the following features: modal_value, polarity, ambiguity and comment. In the field "ambiguity" you should write down the ambiguous modal values. The field comment is a free field to denote doubts or difficulties.

How to create a modal event?

- First create the markables. Drag with the mouse to mark an element and chose "Create Markable at level modal". Select in the box the appropriate features for the markable.

- The essential part here is to link the markables that play a role in the same modal event. This works as follows: Click on the element that you consider the trigger, it will become yellow. Then move your mouse to the target markable and click on the right mouse button, a pop-up "mark as link" will occur, click on this to establish the green link between trigger and target. Do the same for the event and modality sources. You can also now see in the box left that the feature "modal_class" now is no longer "empty" but says something like "set_2". Check that all markables of the same event also carry the same value, in this example all markables should have "set_2".

5.2 Annotation rules

In general, if a selected string is member of more than one markable, a pop-up window will appear from which you can select the correct markable.

A relation can be **removed** by first selecting the one element with the left-button, then the element with which it is linked with the right button (as above), and then selecting the "Unmark as link".

A new markable can be **created** by selecting a string of text with the left mouse button. After creating the markable supply the features. Make sure no other material was already selected (yellow), otherwise you will be adding to an existing markable.

A markable can be **expanded** by first selecting the markable, and then selecting additional material with the left mouse button. The new material will be added. Note that there is no requirement that the new markable must be a continuous string. A markable can be made smaller by selecting it, and than selecting the part to be removed with the left mouse button.

A markable can be **deleted** completely by right clicking on it (when it is not yellow). Make sure no other material is already selected.

To make a **discontinuous markable**, first create a markable. Click on the markable to make it yellow, then select the second part that you want to add with your mouse; a pop-up "add this to markable" appears and click on it to confirm.

5.3 MMAX2 visual interface

In the right top there is markable level control panel where you can see two levels: modal and sentences. As we do not actively do anything with the sentences, you must change the sentence level from "active" to "visible".

In the main screen, top left, under the menu " File" you can find a button "Auto-save", it is good to set this to 5 or 10 minutes. And also press "Save all" regularly during your work to avoid losing your annotations.

Use the Auto-apply option to avoid clicking apply all the time.

Use the option "Font" -> Line spacing to make the distance between the lines bigger, this can help in seeing the annotations better.

If it seems that your latest changes do not appear in the screen, use the button "refresh" or "reapply style sheet".

Use the Markable Set Browser under Tools-> Browsers to check and delete the markables in a set.

5.4 Marking a text string with two different roles

In the example below the trigger and the source of modality are the same. You can create 2 different markables, one for "Querem" as trigger, and one for "Querem" as source. Next you select the trigger markable with the left mouse button, making it yellow, and then point to the same string, but click on the right mouse button, then select the source markable and click on "mark as link". The trigger and the component will be linked by a green line on the screen. When you select a markable, you see if it belongs to a set and you can check if the elements that you linked belong to the same set id.)

(43) Querem *permitir sensações tácteis*.
       Trigger: Querem
       Target: permitir sensações tácteis
       Source of the event mention: sp/wr -->not present in text, not annotated.
       Source of the modality: Querem

5.5 Searching the annotated data

On the main MMAX2 screen, open Tools> QueryConsole.
A new window will open.
Write a search query (see examples in the table below) and press enter.
The results will appear listed in that window, below the search query.
Double click on one of the results to see the hit in the annotated file window.

display modal (type={trigger});	all triggers
display modal (modal_value={epistemic_knowledge});	all triggers with value epistemic_knowledge
display modal (*markable_text={permit.*});	all modal elements beginning with "permit"
display modal (*markable_text={procur.*} and modal_value={volition});	all modal element beginning with "procur" and with value volition
display modal (type={trigger} and *markable_text={pode} and modal_value={epistemic_possibility});	
display modal (*markable_text={permit.*} and modal_value={deontic_permission});	

26

6. Package

The annotated corpus is distributed as a set of files to use with MMAX2.
It includes the following directories:

/MMAX_Modality_data
    /Basedata (contains raw XML files)
    /Markables (contains annotated XML files)
    /MMAX2 (software – to be copied here)
    /Schemes (contains XML modal scheme)

Lauch MMAX2 application in the MMAX2 folder.

The software MMAX2 can be downloaded here:
https://github.com/ottiram/MMAX2
To run MMAX2, Java is required.

6. Citation

When using this corpus, please cite one of these two papers:

Hendrickx, I., Mendes, A., & Mencarelli, S. (2012). **Modality in Text: a proposal for corpus annotation**. *In Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012, May 21-27 2012, Istanbul* (pp. 1805-1812).

Mendes, A., Hendrickx, I., Ávila, L., Quaresma, P., Gonçalves, T., & Sequeira, J. (2016). **Modality Annotation for Portuguese: from manual annotation to automatic labeling**. *Lilt - Language Issues Language Technology*, 14:5, Special volume on Modality: Modes of Modality in NLP.

References

L. B. Ávila, A. Mendes & I. Hendrickx. 2015. Towards a Unified Approach to Modality Annotation in Portuguese. In Proceedings of the IWCS Workshop on Models for Modality Annotation, MOMA 2015 (pp. 1-8). ACL - Association for Computational Linguistics.

K. Baker, M. Bloodgood, B. Dorr, N. W. Filardo, L. Levin, and C. Piatko. 2010. A modality lexicon and its use in automatic tagging. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

J. L. Bybee, R. Perkins, and W. Pagliuca. 1994. The evolution of grammar: Tense, aspect and modality in the languages of the world. University of Chicago Press, Chicago.

R. Farkas, V. Vincze, G. Móra, J. Csirik and G. Szarvas. 2010. The CoNLL 2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. Proceedings of CoNLL 2010, ACL.

I. Hendrickx, , A. Mendes & S. Mencarelli. 2012. Modality in Text: a proposal for corpus annotation. In Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012, May 21-27 2012, Istanbul. pp. 1805-1812.

S. Matsuyoshi, M. Eguchi, C. Sao, K.Murakami, K. Inui, and Y. Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

A. Mendes, I. Hendrickx, L. Ávila, P. Quaresma, T. Gonçalves & J. Sequeira. 2016. Modality Annotation for Portuguese: from manual annotation to automatic labeling. Lilt - Language Issues Language Technology, 14:5, Special volume on Modality: Modes of Modality in NLP.

C. Müller and M. Strube. 2006. Multi-Level Annotation of Linguistic Data with MMAX2. In: Sabine Braun, Kurt Kohn, Joybrato Mukherjee (Eds.): Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods. Frankfurt: Peter Lang, pp. 197-214. (English Corpus Linguistics, Vol.3).

F. Oliveira. 1988. Para uma semántica e pragmática de DEVER e PODER. Ph.D. thesis, Universidade do Porto.

F. R. Palmer. 1986. Mood and Modality. Cambridge textbooks in linguistics. Cambridge University Press.

R. Saurí, M. Verhagen, and J Pustejovsky. 2006. Annotating and recognizing event modality in text. In Proceedings of the 19th International FLAIRS Conference.

J. Van der Auwera and V. Plungian. 1998. Modality's semantic map. Linguistic Typology, pages 79–124.

J. Wiebe, T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language.Language Resources and Evaluation, 39:165–210.