

# 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

*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 2.0 presents a slightly modified version of the modality annotation schema in line with the published unified schema detailed in Ávila et al. (2015).*

1. Introduction .....	1
2. Data .....	2
3. Annotation Scheme .....	3
3.1 General annotation steps .....	3
3.2 Modal Values .....	3
3.3 What we do not consider as modal .....	4
3.4 Components .....	4
3.4.1 Trigger .....	5
3.4.1.1 Definition .....	5
3.4.1.2 Features of the modal event .....	7
3.4.2 Target .....	7
3.4.3 Source of the event mention .....	9
3.4.4 Source of the modality .....	9
3.4.5 Polarity .....	12
3.4.6 Structural Ambiguity .....	12
4. Annotation Software MMAX2 .....	13
4.1 Markables .....	13
4.2 Annotation rules .....	14
4.3 MMAX2 visual interface .....	14
4.4 Marking a text string with two different roles .....	15
5. Package .....	15
6. Citation .....	19
References .....	19

## 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<sup>1</sup> that produces stand-off annotation in XML format (Müller & 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

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 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<sup>2</sup> 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.

---

<sup>1</sup> <http://mmax2.net>

<sup>2</sup> <http://alfclul.clul.ul.pt/CQPweb>

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 led 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 parts of speech triggers: verbs, adjectives, nouns, adverbs (see section 3).

This method of selection of sentences for our corpus has some implications. First of all, 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: tokens per file

## 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 Modal Values

Table 2 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 tentaram 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 desejo 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, possivelmente, 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 difícil 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;

- (5) Se 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*

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*;

- (6) Onde arranja o dinheiro não sei, 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 (7), the trigger is *ter de* and the preposition *de* is included in the trigger, while in (8) the trigger is *capaz* and the preposition is included in the target.
  - (7) Tenho de ir.  
Trigger: tenho de  
Target: ir
  - (8) nenhuma força internacional será *capaz* de fazer cumprir os acordos  
Trigger: capaz  
Target: de fazer cumprir os acordos
4. 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 (8) the trigger is “correu-se ... o risco-
  - (9) correu se no entanto o risco de chegar a uma situação análoga à que hoje caracteriza os países recentemente industrializados de outros continentes  
trigger: correu -se@o risco<sup>3</sup>  
target: de chegar a uma situação análoga à que hoje caracteriza os países recentemente industrializados de outros continentes
5. If the trigger is split into two parts by, for example, an adverb, we do not consider the adverb as part of the trigger;
  - (10) Temos amanhã de limpar a casa.  
Trigger: temos@de
6. When one sentence contains two or more triggers, we annotate them separately;
  - (11) 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
  - (12) 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
7. 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 (12) (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

---

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.

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.

- (13) João, senta!  
 Trigger: !  
 Target: *João, senta!*

#### 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

**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.

**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.

#### 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 discontinuity.

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;

(14) Comerciantes da Guarda exigem a reabertura da Rua do Comércio.

- b) Subordinate clause: tag the whole subordinate clause, including the conjunction or preposition introducing it;

(15) Ele acha que a Igreja fica favorecida pelo facto de estar aliada ao poder político.

- c) Verbal phrase: the complements of the verb should be included.

(16) No terreno das indústrias da cultura - cinema , livro , televisão - , arriscamo - 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:
 

(17) Logo a seguir, ensaiou o primeiro «drible», quando *lhe* perguntaram 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.

- (18) Autorizei o meu filho ontem a ir ao parque.  
Trigger: autorizei  
Target: o meu filho ontem a ir ao parque
3. Target not lexically expressed: not annotated
- (19) Felizmente ninguém na plateia respondeu : "Sabemos sim , senhor Little Axe ..."  
Trigger: sabemos  
Target:
4. Target in imperatives: the target is verb in the imperative and its complements:
- (20) *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.
- (21) Espero por essa reunião e espero 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*).
- (22) Parecem-se com os livros de papel, querem permitir 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.
- (23) O João pode chegar tarde  
Trigger: pode  
Modal sense: epistemic possibility  
Target: o João@chegar tarde  
Source of the modality: sp/wr



Source of the event: sp/wr

- (24) O João tem de acabar o trabalho  
Trigger: tem de  
Modal sense: deontic obligation  
Target: o João@acabar o trabalho  
Source of the modality: sp/wr  
Source of the event: 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 (25) or, in cases of null subject, we mark the verb that refers to the source of the event mention (26).

- (25) *O João contou que quando foram à praia a Maria não deixou 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
- (26) *“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 restrictive clauses (29).

- (27) *O ministério, ao anunciar que retirará a classificação de superior aos nove institutos, oferecendo-lhes a possibilidade de concederem licenciaturas pelo politécnico, procura repor a lei, minimizando os danos aos alunos que optaram por esses institutos privados.*
- (28) *O governante português afirmara que nenhuma força internacional será capaz de fazer aplicar os acordos se as partes não os cumprirem.*
- (29) *As diferentes abordagens que têm sido utilizadas no estudo do cérebro precisam de um denominador comum, declarou à revista britânica new scientist guy mckhann<sup>4</sup>, da universidade johns hopkins.*

Table 3 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
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 the modality
evaluation	Source of the modality: who evaluates something

Table 3. Definition of the source of modality for each modal value

Difficult cases:

1. null subject: tag the verb as the source of the modality, when the subject of the verb is the source of the modality  
(30) *Querem permitir sensações tácteis.*  
Trigger: *querem*  
Source of the event mention: *sp/wr*  
Source of the modality: *querem*
  
2. 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.  
  
(31) *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 PS - - , obrigando os responsáveis do SPD e da Fundação Friedrich Ebert a alterarem o que estava previsto.*  
Trigger: *obrigando*  
Source of the modality:
  
3. 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):  
(32) *Sabe -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*
  
4. null subject and source is pronoun: tag the source of the modality on the pronoun and don't tag the pronoun in the trigger;  
(33) *Apetece -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:

(34) Apetecem -me morangos.

Trigger: apetecem

Target: morangos

Source of the event mention: -me

Source of the modality: -me

5. 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)

(35) 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 consideradas 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

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

7. With certain modal values (e.g., epistemic\_possibility) and lemmas (e.g. *permitir*), the source of the modality may be a concrete object:

(36) *A janela* permitia ter muita luz na sala.

Trigger: permitia

Source of the event mention: sp/wr

Source of the modality: a janela

8. 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 is its co-referent with inside the same clause, tag the NP. If it is coreferent to an NP outside the clause, tag the pronoun. For instance, in (37), 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.

(37) *Catarina* nunca confessou o seu desejo pela amiga e esta sempre fez de tudo para não o saber.

Trigger: desejo

Target: pela amiga

Source of the modality: Catarina

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

(38) É necessário que todos colaborem

Trigger: necessário

Modal sense: deontic necessity

Target: todos colaborem

Source of the modality: -

Source of the event: sp/wr

10. In contexts of epistemic possibility with “é possível que”, the source of the modality is the speaker/writer.

### 3.4.5 Polarity

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), impedir (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 (38), 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:

(39) O João obrigou 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.

(40) É este um vício que sempre atinge os míseros: **nunca** conseguir crer 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 (40).

(41) E **não** há *nenhum* combate em que os intervenientes **não** procurem perceber e tornar claro o que é que os satisfará para estabelecerem a paz.

Trigger: procurem

Polarity: positive

### 3.4.6 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.

*O João tem de acabar o trabalho*

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

- (42) em negócios não se deve 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. 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.

##### 4.1 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 therefore 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 choose "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`".

## 4.2 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 then 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.

## 4.3 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.

#### **4.4 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. 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)

The software MMAX2 can be downloaded here:

<https://github.com/ottiram/MMAX2>

To run MMAX2, Java is required.

# 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.....	2
2. Data.....	2
2.1 Modality corpus 1.....	2
2.2 The CRPC-Modal Corpus .....	3
3. Annotation Scheme.....	5
3.1 General annotation steps.....	5
3.2 The full scheme and the list of modal Values.....	5
3.3 What we do not consider as modal.....	6
3.4 Components.....	6
3.4.1 Trigger.....	7
3.4.1.1 Definition .....	7
3.4.1.2 Features of the modal event.....	10
3.4.1.2.1 Modal value .....	10
3.4.1.2.2 Polarity.....	11
3.4.1.2.3 Ambiguity.....	13
3.4.2 Target .....	14
3.4.3 Source of the event mention .....	15
3.4.4 Source of the modality.....	16
3.4.5 Structural Ambiguity.....	22
4. Changes in modal annotation scheme v. 2.0 and 3.0 .....	23
5. Annotation Software MMAX2 .....	24
5.1 Markables .....	24
5.2 Annotation rules .....	25
5.3 MMAX2 visual interface .....	25
5.4 Marking a text string with two different roles.....	25
5.5 Searching the annotated data.....	26
6. Package.....	26
6. Citation .....	26
References.....	27



# 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.....	2
2. Data.....	2
2.1 Modality corpus 1.....	2
2.2 The CRPC-Modal Corpus .....	3
3. Annotation Scheme.....	5
3.1 General annotation steps.....	5
3.2 The full scheme and the list of modal Values.....	5
3.3 What we do not consider as modal.....	6
3.4 Components.....	6
3.4.1 Trigger.....	7
3.4.1.1 Definition .....	7
3.4.1.2 Features of the modal event.....	10
3.4.1.2.1 Modal value .....	10
3.4.1.2.2 Polarity.....	11
3.4.1.2.3 Ambiguity.....	13
3.4.2 Target .....	14
3.4.3 Source of the event mention .....	15
3.4.4 Source of the modality.....	16
3.4.5 Structural Ambiguity.....	22
4. Changes in modal annotation scheme v. 2.0 and 3.0 .....	23
5. Annotation Software MMAX2 .....	24
5.1 Markables .....	24
5.2 Annotation rules .....	25
5.3 MMAX2 visual interface .....	25
5.4 Marking a text string with two different roles.....	25
5.5 Searching the annotated data.....	26
6. Package.....	26
6. Citation .....	26
References.....	27

# 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.....	2
2. Data.....	2
2.1 Modality corpus 1.....	2
2.2 The CRPC-Modal Corpus .....	3
3. Annotation Scheme.....	5
3.1 General annotation steps.....	5
3.2 The full scheme and the list of modal Values.....	5
3.3 What we do not consider as modal.....	6
3.4 Components.....	6
3.4.1 Trigger.....	7
3.4.1.1 Definition .....	7
3.4.1.2 Features of the modal event.....	10
3.4.1.2.1 Modal value .....	10
3.4.1.2.2 Polarity.....	11
3.4.1.2.3 Ambiguity.....	13
3.4.2 Target .....	14
3.4.3 Source of the event mention .....	15
3.4.4 Source of the modality.....	16
3.4.5 Structural Ambiguity.....	22
4. Changes in modal annotation scheme v. 2.0 and 3.0 .....	23
5. Annotation Software MMAX2 .....	24
5.1 Markables .....	24
5.2 Annotation rules .....	25
5.3 MMAX2 visual interface .....	25
5.4 Marking a text string with two different roles.....	25
5.5 Searching the annotated data.....	26
6. Package.....	26
6. Citation .....	26
References.....	27

## 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 semântica e pragmática de DEVER e PODER*. Ph.D. thesis, Universidade do Porto.

Christoph Müller and Michael Strube. 2006. Multi-Level Annotation of Linguistic Data with

MMA2. In Sabine Braun, Kurt Kohn, and 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. 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.