

MODALITY CORPUS 1: 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 2023

In this annotation manual we present the annotation scheme for Portuguese and explain how to annotate. We describe how to use the software annotation tool MMAX2. We also discuss difficult cases and the annotation rules.

1. Introduction	1
2. Annotation Scheme.....	2
2.1 General annotation steps.....	3
2.2 Modal Values.....	3
2.3 What we do not consider as modal	4
2.4 Components	5
2.4.1 TRIGGER:.....	5
2.4.1.1 Features of the modal event.....	7
2.4.2 TARGET	7
2.4.3 Source of the event mention	9
2.4.4 Source of the modality	10
2.4.5 POLARITY	12
2.4.6 Structural Ambiguity	14
3. Annotation Software MMAX2.....	15
3.1 Markables.....	15
3.2 Annotation rules.....	16
3.3 MMAX visual interface.....	16
3.4 Marking a text string with two different roles	16
4. Package.....	17
6. Citation	17
References.....	17

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

Deontic modality denotes when the speaker imposes something on the hearer. We identify two sub values: deontic obligation and deontic permission.

Inspired by the work of van der Auwera et al. (1998) we also mark participant-internal necessity, to tag personal needs of the speaker or participant, and participant-internal capacity, to tag personal capacities of the speaker or participant.

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

proposition, and the value volition for hopes and wishes. Following Baker et al. (2010), we consider the modalities effort and success: the first focuses on the commitment of the participant to make something happen, and the latter focuses on the results of the commitment of the participant. In the next part we will explain each of these values in more detail.

For the annotation itself we use a software tool call MMAX2¹ (Müller & Strube, 2006) that produces stand-off annotation in XML format.

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 «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 subject of the modality, who experiences the modality

The trigger is the most important component in the scheme and this carries the modal meaning and expresses the modal value (one of the 8 that we briefly mentioned above). The other components are optional as they are not always expressed in the sentence.

2. Annotation Scheme

Data

The modality corpus 1 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 CRPC² 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).

¹ <https://github.com/nlpAThits/MMAX2>

² <http://alfclul.clul.ul.pt/CQPweb>

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 this manual.

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

2.1 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 (Farkas et al, 2010). For the trigger we only annotate the minimum, the smallest possible unit like "desejo", "presicarem" or "tem de". For the target we annotate maximally and include all relevant parts. 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.

2.2 Modal Values

In Table 1 is a list and a description of each of the 8 different modal values that we distinguish.

Values	Subvalues	Definition
Epistemic	Knowledge	The speaker presents his or someone else's knowledge or when he expresses some degree of understanding about something
	Belief	The speaker expresses his or someone else's belief in something
	Doubt	The speaker doubts something or expresses someone else's doubts
	Interrogative	To annotate interrogative sentences
	Possibility	The speaker presents what he or someone else is saying as a possibility or probability
Deontic	Permission	The speaker or participant allows the hearer to do something
	Obligation	The speaker or participant forces the hearer to do something
Participant-internal	Capacity	The speaker expresses his own or someone else's personal capacity
	Necessity	The speaker expresses his own or someone else's personal necessity
Evaluation		The speaker presents his or someone's else evaluation of facts and propositions
Volition		The speaker expresses his or someone else's wishes, hopes or wants
Effort		The speaker expresses his or someone else's commitment and application to make something happen or to do something
Success		The speaker expresses if the results of the commitment of the speaker were successful or not

Table 1. List and description of modal values.

2.3 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 do not annotate aspect. Verbs like *passar-a*, *acabar*, *continuar* signal a change of

state, not a 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).

2.4 Components

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

2.4.1 TRIGGER:

Definition: The trigger is the word or string of words that expresses modality (Baker et al., 2010).

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.

e.g. *Mudaram a exposição do livro - aplicando técnicas de "marketing" e de "merchandising" - e tentaram chamar clientes que não iam às livrarias tradicionais.*

Trigger: tentaram

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

e.g. *Ao optar por fumar agora estou a influenciar o meu futuro **desejo** de fumar.*
'Choosing to smoke now I am influencing my future desire to smoke'.

Trigger: desejo

Target: de fumar

- 1.3 an **adverb**, we tag as trigger only the adverb and only when it expresses modality of an event, not of an entity;

e.g. *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**, we only consider it as trigger if it is part of a verbal phrase but 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). We do not annotate as triggers adjectives that are included in a noun phrase, as in the case of *organizadas* in (e.g.2), as well as we do not annotate the adverb *bem*,

since it has no modal meaning;

e.g.1 Foi **difícil** mas se tenta dar às pessoas a sensação de que as compreende, é mais fácil.

Trigger: difícil

e.g.2 Depois, e a outro nível, espero vir a ter as corridas mais bem organizadas.

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*;

e.g. Onde arranja o dinheiro não **sei**, nem interessa.

Trigger: sei

Target: onde arranja o dinheiro

3. Prepositions and conjunctions are never included in the trigger but are included in the target. The only exceptions to this rule are *ter de* and *ter que*.

e.g. **Tenho de** ir.

Trigger: tenho de

Target: ir

4. If the Trigger is a multiword expression or a semifixed MWE, we only annotate the node.

e.g. Catarina nunca confessou [NP o seu desejo pela amiga]] e [esta sempre fez de tudo para não o saber

trigger: fez

target: para não o saber

source: esta

modal value: effort

polarity: positive

5. If the trigger split into two parts by an adverb which does not carry modal information, we do not consider the adverb as part of the trigger;

e.g. Temos *amanhã* de limpar a casa.

Trigger: temos@de³

6. When in one sentence there are more triggers, we annotate them separately;

e.g.1. 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

Trigger: difícil

e.g.2. É este um vício que sempre atinge os míseros: nunca **conseguir** **crer** na felicidade!

Trigger: conseguir

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.

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 interrogative and imperative clauses. When the clause is an interrogative, we consider as trigger the question mark (?). When the clause is an imperative, we consider as trigger the exclamation mark (!). In an order like example e.g.2, 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.

e.g.1. *Qual será o mecanismo que permite transferir directamente a energia do interior da estrela para a sua coroa externa?*

Trigger: ?

e.g.2. *João, senta!*

Trigger: !

2.4.1.1 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

Definition of 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: when multiple modal values apply at the same time and none of these values is clearly more natural than the others, we annotate multiple modal values.

2.4.2 TARGET

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

The target can be a) a noun phrase; b) a subordinate clause c) verbal phrase

- a) Nominal target: tag the whole nominal phrase, including articles, adjectives or restrictive clauses;

e.g. *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;

e.g. *Ele acha [que a Igreja fica favorecida pelo facto de estar aliada ao poder*

político].

- c) the verbal phrase, the complements of the verb should be included.

e.g. *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 - @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:

e.g. *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

3. Target not lexically expressed: not annotated

e.g. *Felizmente ninguém na plateia respondeu : "Sabemos sim , senhor Little Axe ... "*

Trigger: sabemos

Target:

4. Target in interrogatives:

- direct interrogatives: in interrogatives the target is the core sentence. Therefore we decided to mark the whole interrogative as target.

e.g. *Sete vezes sete?*

Trigger: ?

Target: Sete vezes sete?

- Indirect interrogatives: the trigger is expressed by a word and the target is the elements in the scope of the trigger:

e.g. *E assim chegamos ao cerne da questão, que é saber [quem nomeou Manuel dos Santos].*

Trigger: saber

Target: quem nomeou Manuel dos Santos

5. Target in imperatives: the target is the person to whom the order or permission is directed:

e.g. *João, senta!*

Trigger: !
Target: João

6. When there is a pronoun in an element, and the sentence also contains the full name where the pronoun refers to, we decided to 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.

e.g. 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'.

7. Ambiguity in unexpressed source of modality

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 referred 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.

e.g. *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:

8. If the target is semantically linked to another part of the sentence, we only annotate the target itself avoiding the annotation of semantic relations, as the example in which we tag as target only *sensações tácteis parecidas*, and not *os livros de papel*, as they are semantically but not syntactically linked to each one to the other.

e.g. *Parecem-se com os livros de papel, querem permitir [sensações tácteis
parecidas] e dar ao leitor algo mais – a possibilidade de ter uma biblioteca e
um banco de dados ambulante, pesando pouco mais de um quilo.*
Trigger: permitir
Target: sensações tácteis parecidas

2.4.3 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 when the name of the sp/wr is present in the sentence.

e.g. *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
 Source of the modality: a Maria
 Modal value: deontic_permission
 Polarity: negative

2.4.4 Source of the modality

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

The proposition tagged as source includes a) nouns (e.g.1); b) articles (e.g.1); c) adjectives (e.g.2); d) prepositional phrases (e.g.3); e) restrictive clauses (e.g.4).

e.g.1 **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.

e.g.2 O governante português afirmou que **nenhuma força internacional** será capaz de fazer aplicar os acordos se as partes não os cumprirem.

e.g.3 Enquanto isso, **o médico do clube alemão** referiu que o jogador precisa de descansar.

e.g.4 **As diferentes abordagens que têm sido utilizadas no estudo do cérebro** precisam de um denominador comum, declarou à revista britânica *new scientist* *guy mckhann*⁴, da universidade johns hopkins.

The following is a list of what we consider as source of the modality for each modal value:

Modal value: epistemic_knowledge	Source of the modality: who has the knowledge
Modal value: epistemic_belief	Source of the modality: who has the belief
Modal value: epistemic_doubt	Source of the modality: who has the doubt
Modal value: epistemic_interrogative	Source of the modality: who asks the question
Modal value: epistemic_possibility	Source of the modality: who speaks or writes and who has the possibility to do something
Modal value: participant-internal_necessity	Source of the modality: who has the necessity
Modal value: participant-internal_capacity	Source of the modality: who has the capacity

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

Modal value: deontic_permission	Source of the modality: who/what establishes the permission
Modal value: deontic_obligation	Source of the modality: who/what establishes the obligation
Modal value: effort	Source of the modality: who / what makes the effort / works towards a goal
Modal value: success	Source of the modality: who or what was successful
Modal value: volition	Source of the modality: who wants to
Modal value: evaluation	Source of the modality: who evaluates something

Difficult cases:

1. null subject: tag the verb also in the source of the modality

e.g. *Querem permitir sensações tácteis.*

Trigger: *querem*

Source of the event mention: sp/wr

Source of the modality: *querem*

2. null subject_impersonal: tag source of the modality on the 'se' and don't tag the se in the trigger:

e.g. *Quando se pode dizer o inverso...*

Trigger: *pode*

Target: *dizer o inverso*

Source of the event mention: sp/wr

Source of the modality: *se*

3. null subject_reflexive: tag the source of the modality on the pronoun and don't tag the pronoun in the trigger;

e.g. *Apetece -me*

Trigger: *apetece*

Target:

Source of the event mention: -me

Source of the modality: -me

If the subject is present the annotation is the same, as the grammatical subject has the modal function of the target:

e.g. *Apetecem -me morangos.*

Trigger: *apetecem*

Target: *morangos*

Source of the event mention: -me

Source of the modality: -me

4. passive sentences: source of the modality = agent

e.g. *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: foram consideradas

Source of the event mention: sp/wr

Source of the modality: Tribunal de Execução de Penas

If the source of the modality is not expressed, we tag the verb as trigger:

e.g. *Os resultados esperados na prova foram conseguidos.*

Trigger: foram conseguidos

Source of the modality: foram conseguidos

5. interrogatives: source of the event mention = source of the modality = sp/wr

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

7. when the source and the target are the same, tag the same expression in both:

Falha assim o equilíbrio do actores.

Trigger: falha

Target: o equilíbrio dos actores

Source of the modality: o equilíbrio dos actores

8. if the source of the modality is a concrete object, tag it as source of the modality anyway (for certain modal values):

e.g. *A janela permitia ter muita luz na sala.*

Trigger: permitia

Source of the event mention: sp/wr

Source of the modality: a janela

9. If the source is a pronoun and there is no other referential expression, tag the pronoun in the clause. If there is a pronoun and an NP it is coreferent with inside the same clause, tag the NP. If it is coreferent to an NP outside the clause, tag the pronoun.

e.g. *Catarina nunca confessou [NP o seu desejo pela amiga]] e [esta sempre fez de tudo para não o saber].*

Catarina never confessed her desire for her friend and this friend has always done everything not to know *it*.

trigger: desejo

target: pela amiga

source: Catarina

modal value: volition

polarity: positive

2.4.5 POLARITY

We only mark the polarity of the modal value and not that of the entire clause or sentence: in the following example 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:

e.g. *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

In some cases the polarity of the modal value is not really positive but more neutral, as for the values *epistemic interrogative* and *epistemic doubt*. However, we do not annotate neutral cases, because it is easier to have only 2 values as the introduction of a 3rd value will cause new problems and difficulties to arise.

If the negative particle focuses on a modal trigger which has in its target another modal trigger, both the triggers are marked with negative polarity.

e.g. *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_capacity
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

With the verb *dever* (with value deontic obligation), a negative trigger before the verb does not imply the negation of the obligation. The sentence below establishes the obligation of not doing something: so, although the negation is before the verb, it does not have scope over *dever*, but rather over its target. In these cases, we consider the polarity as positive and we include the negative element inside the target. The sentence will be annotated in the same way as the equivalent one “A Ana deve não ler esse livro”.

e.g. A Ana não deve ler esse livro. (obligation)
Trigger: deve
Target: A Ana não@ler esse livro
Source of the event mention: sp/wr
Source of the modality: sp/wr
Modal value: deontic_obligation
Polarity: positive

The same is not true with *dever* when it has epistemic value. So, the same sentence above can have the value epistemic probability (it is not probable that Ana reads the book). In these cases, the negative element has scope over *dever* and the polarity of the value epistemic probability is negative.

e.g. A Ana não deve ler esse livro. (epistemic probability)
Trigger: deve
Target: A Ana@ler esse livro
Source of the event mention: sp/wr

Source of the modality: sp/wr
 Modal value: epistemic_possibility
 Polarity: negative

If the negative element is after the verb and has scope over the target, the modal value is positive probability that something does not happen. (However, the epistemic reading is not as natural with the negation after the verb *dever*.)

e.g. A Ana deve não ler esse livro. (epistemic probability)
 Trigger: deve
 Target: A Ana@não ler esse livro
 Source of the event mention: sp/wr
 Source of the modality: sp/wr
 Modal value: epistemic_possibility
 Polarity: positive

2.4.6 Structural Ambiguity

With the verbs *poder*, *dever* and *ter de* in some sentences there are two possible interpretations: according to the interpretation chosen 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*. By marking this explicitly, we can later on look at all these cases and see whether we can discover new regularities or a better annotation solution.

A Ana pode responder às perguntas

Trigger: pode Target: responder às perguntas Source of the modality: A Ana Modal value: participant-internal_capacity Polarity: positive	Trigger: pode Target: A Ana@responder às perguntas Source of the modality: sp/wr Modal value: deontic_permission Polarity: positive
--	---

A carne tem de ser guardada no congelador

Trigger: tem de Target: ser guardada no congelador Source of the event mention: sp/wr Source of the modality: A carne Modal value: participant-internal_necessity Polarity: positive	Trigger: tem de Target: A carne@ser guardada no congelador Source of the event mention: sp/wr Source of the modality: sp/wr Modal value: deontic_obligation Polarity: positive
---	---

A Maria deve ser examinada pelo médico

Trigger: deve Target: ser examinada pelo médico Source of the event mention: sp/wr Source of the modality: A Maria Modal value: participant-internal_necessity Polarity: positive	Trigger: deve Target: A Maria@ser examinada pelo médico Source of the event mention: sp/wr Source of the modality: sp/wr Modal value: deontic_obligation Polarity: positive
--	--

3. Annotation Software MMAX2

MMAX2 is a 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. MMAX 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.

3.1 Markables

In MMAX 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".

3.2 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.

3.3 MMAX 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.

3.4 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 same string, but click on the right mouse button, then select the source markable and click on "mark as link". Of course the link it self is not visible in the main screen, but when you use the Markable Set Browser under Tools-> Browser, you can see whether your action has succeeded.

e.g. 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

4. 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.

6. Citation

When using this version of the corpus, please cite the following paper:

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).

References

Kathrin Baker, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine 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.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL 2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. *Proceedings of CoNLL 2010, ACL*.

Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji 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 LanguageResources Association (ELRA).

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. Oliveira. 1988. *Para uma semântica 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.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39:165–210.