Table of Contents

1. Introduction

Language is the main tool human beings use to establish relationships and to communicate. People use language to give information, to share ideas, to express personal opinions and feelings, to interact with other people. Through language, we can express our personality, our mood, our emotions and our ideas, our needs or our personal desires (1), and mark our presence in the conversation in a specific way.

(1) ***Queria*** *experimentar algo de diferente.*
'I wanted to experience something different'.

For example, one of the most common things teachers say to students is not to speak during classes. They generally use expressions that convey some order as shown in example (2)[1].

(2) *Não* ***fales*** *com os teus colegas durante a aula.*
'Do not speak to your classmates during the lesson'.

Through these expressions, the teacher is establishing a specific relation with the student, in which the teacher imposes himself on the student.

In Portuguese Linguistics, there is a special category called *Modality* which regards the study of the way how the speaker linguistically expresses his personality and his attitude towards facts, towards other people or towards what he or other people are saying.

In Portuguese, till now, modality has been explored from a theoretical point of view by Fátima Oliveira and Maria Henriqueta Costa Campos in their PhD dissertations, *Para uma semântica e pragmática de DEVER e PODER* (1988) and *Abordagem Enunciativa de um subsistema do sistema modal do português: os verbos DEVER e PODER* (1989). However, in these two theses, the observation of modality is restricted to the observation of the modal meaning of the verbs *poder* and *dever* and of their use.

Our investigation on modality in Portuguese is, instead, centred on the observation of the linguistic expression of modality, since we want to observe which are the linguistic devices used by Portuguese people to express their commitment towards facts and propositions. In particular, our interest for modality has emerged especially when we understood that through the analysis of the linguistic expression of modality we would explore two main aspects of Portuguese. First, we would observe how speakers express their

[1] Sentence (2) is a product o four introspection, as we wanted to give an example of a teacher giving an order to a student.

subjectivity in language. The following examples may clarify what we mean.

(3) *Um país cuja língua não pode prescindir de diminutivos, um país cuja língua está tão carregada de afectos que parece apenas ter sido concebida para as crianças,* ***penso*** *que um tal país não pode crescer.*
'A country whose language cannot neglect diminutives, a country whose language is so full of emotion to seem that it was created only for children, I think such a country cannot grow'.

(4) *Quando João Miguel deixou o bloco operatório, para onde tinha entrado com um sorriso nos lábios, tinha à sua espera o melhor amigo.*
'When John Michael left the emergency room, where he had entered with a smile on his face, his best friend was waiting for him'.

If in example (3), the speaker uses the verb *pensar* 'think' to mark that he is not stating what he is saying as a certainty but that he is expressing a personal opinion, in sentence (4), the speaker is presenting a fact, without expressing any kind of personal involvement.

Second, we would discover the different meanings a Portuguese linguistic expression can have according to the context in which it occurs. Sentences (5) and (6), for example, show how the verb *crer* 'believe' can have two different meanings according to the context in which it appears: it can express a belief in something, as in sentence (5), or it can mark some doubt, as in sentence (6).

(5) *Não* ***creio*** *em Deus.*
'I don't believe in God'.

(6) *Desta vez,* ***creio*** *que a ETA quer negociar.*
'This time, I think ETA wants to negotiate'.

In order to study the expression of modality in Portuguese, we have used a methodology that till now had not yet been applied for this purpose in Portuguese Linguistics: to observe how modality is conveyed in real language, we adopt a corpus-based approach (McEnery, T., Wilson, A., 1996), and create a specific system for the annotation of modality in corpora.

A Corpus is a collection of texts that can be used by linguists to investigate some

issues of a language. There are certain important characteristics that a collection of texts must have to be defined as 'corpus': a) it has to be a sample capable of representing the use of a variety of a language; b) it generally must have a finite size; c) it must have a machine-readable form, in order to be quickly searched; d) it has to be considered as a standard reference for the language variety it represents. Corpora are very important tools to be used for linguistic investigation since, as we have seen, they are built with the main objective of representing the real use of a language variety.

As we can see, we study modality, an aspect which mainly belongs to Semantics (the area studying the meaning of words), using a corpus-based methodology, the Linguistic Annotation of corpora, a system through which we can apply descriptive and analytic notations to row language data.

Our research aims at analysing how speakers convey modality in the current use of Portuguese. We, in fact, want to observe which can be the most expressed modalities in Portuguese, which are the most used linguistic expressions to convey them, and if there are some recurrent relations between modalities and between the modality and the linguistic expression conveying it. Moreover, we also want to observe all the elements involved in the expression of modality and affected by the modal value conveyed in the sentence, studying the relations between them. For example, we can observe which modal values a specific verb generally conveys; or if there are some recurrent combinations between a modal expression and the expression in its scope, or between the modal expression and the holder of the modality. In the end, through the annotation of the polarity of the modal value we can also observe how negation interacts with modality in Portuguese.

The modality annotation scheme that we have developed is applied to annotate a corpus sample of 1935 sentences with modal information. This corpus has been extracted from the Corpus de Referência do Português Contemporâneo[2] (CRPC), a Portuguese corpus on which the Centro de Linguística da Universidade de Lisboa (CLUL) is working since 1988, which collects various types of written and oral texts belonging to different Portuguese speaking countries, of the period from 1970 onwards, for a total amount of 311 millions of words. The sentences for our corpus were selected by querying for the lemmas of specific modal verbs. Our annotation was, then, done at the sentence level, which means that we did not observe co-referential relations between linguistic elements, so that if there was an ellipsis we could not recover the omitted element. Moreover, deciding to extract sentences for a list of verbs and not in non-sampled text is a process which has some implications: first, modality is mostly carried by verbs, and, second, the selection of the verbs clouds the frequency of the modal values and of the other components. We then annotate our corpus sentence per

[2] http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_crpc.php

sentence, observing which linguistic elements have to be tagged under each label.

The construction of our annotation scheme is based on two main steps: the first is the creation of a detailed system of modal values and the second is the creation of labels for the other elements that may be involved in the expression of modality in Portuguese. However, the scheme is the result of the process of annotation of the corpus, since it was adapted during the annotation. Our annotation scheme is composed of six main components. With the term *component* we mean information fields that we fill in with the details of the modality expression. The six components are: a) the *Modal value*, used as a place to mark which is the modal value expressed; b) the *Trigger*, used as the place to identify the linguistic expression conveying the modal value; c) the *Target*, the component in which we tag the linguistic expression in the scope of the trigger; d) the *Source of the event mention*, the place where we tag who is expressing the sentence containing the modal expression; e) the *Source of the modality*, the place where we tag who is the holder of the modality; and f) the *Polarity*, the place where we tag if the modal value conveyed is positive or negative. All these components are inspired in the tangible annotation schemes presented in the literature on the annotation of modality in English (Baker et al., 2010; Matsuyoshi et al., 2010; Saurí et al., 2006; Wiebe et al., 2005), while our list of modal values is inspired both in the modal values considered in the literature on the annotation of modality and in the ones proposed in the literature on modality in traditional Linguistics (Palmer, 1986; Oliveira, 1988; van der Auwera, 1998).

Along the annotation of our corpus, we have observed that in some cases an expression is ambiguous as it can be associated to more than one modal value. In sentence (7), for example, the expression *tenho que* 'must' is ambiguous and can be interpreted as expressing an obligation imposed by someone or by the situation or as expressing a personal necessity of the speaker or writer.

(7) *Para já **tenho que** perceber o conteúdo da crítica: se é maldosa, musculada ou se tem um fundo de verdade e se tiver, aprender com ela.*
'I must understand the content of the criticism: if it is mean, structured or if it is based on some truth, and if it is, I must learn with it'.

In order to report this ambiguity, we also created the component *Ambiguity*. Its annotation allows us to identify all those expressions that can be associated to more than one modal value, to immediately know which are the most recurrent ambiguities between modal values and if some ambiguity is always conveyed by the same linguistic elements.

In contrast to the annotation schemes presented in the literature on the annotation of modality in Computational Linguistics (Baker et al., 2010; Matsuyoshi et al., 2010), by explicitly annotating *Ambiguity* in modality, we study a phenomenon that has not been studied

systematically before. We are aware that the corpus sample choice of isolated sentences does increase the chances of having ambiguous sentences that within a larger context might not have been ambiguous. Still we believe that studying possible ambiguities between different modal values is a valuable contribution.

The thesis is divided in three main chapters: the first presents a panorama of the literature on modality and on its annotation; in the second, we introduce the annotation scheme we created; and in the third chapter we outline the results of the annotation of the corpus with our scheme.

The first chapter is composed of three main sections. In the first section, we present the existing studies on modality in traditional Linguistics. As concerns the research done on modality in Linguistics, we especially refer to the work *Mood and modality* of Palmer (1986) and to the thesis *Para uma semântica e pragmática de DEVER e PODER* of Oliveira (1988), since the two authors can be considered as the pioneers of the study of modality in English and Portuguese. However, we will also report on the typologies of modality presented by other authors in English (van der Auwera, 1998; Bybee, 1984; Huddlestone and Pullum, 2004), since it is interesting to compare the different systems of modal values these authors create.

The second section illustrates the existing works on the annotation of modality in the field of Computational Linguistics, describing the systems created by different authors for the annotation of modality in English (Baker et al., 2010; Matsuyoshi et al., 2010; Saurí et al., 2006 and 2007; Szarvas et al., 2008; Wiebe et al., 2005).

In the end of the chapter, we also show a comparison between the modal systems created in Linguistics and the schemes created for the annotation of modality in Computational Linguistics.

The second chapter of the thesis presents the scheme we built for the annotation of modality in Portuguese and the methodology used for the annotation of our corpus. We, first, describe how we built our corpus, indicating the corpus from which it was extracted and describing the methodology of extraction of the sentences. Second, we present the modal values we use for the annotation, defining each of them. Third, we present the other components of the annotation scheme and describe the criteria we used to annotate them, especially focusing on the most simple cases faced in the annotation. In this chapter, we also introduce the concept of ambiguity and explain how we deal with it in the annotation of each component.

The third chapter focuses on the results of the annotation of our corpus of sentences. Here, we show some statistics done on the data annotated at the modal level and present the most standard and easiest cases found in the annotation of each component, in order to, then,

explore the most difficult situations faced, the problems found, and the ambiguities encountered.

In the conclusion, we present an overview of our work, focusing on the positive issues of our annotation system and of our research, and we present the possible future applications of our annotation scheme, underlining the topics that need further investigation and improvement.

2. Modality

This work presents a scheme built for the annotation of modality in a Portuguese corpus. The scheme aims at analysing modality in sentences and includes components inspired in the ones presented in the literature on the annotation of modality in Computational Linguistics and modal values taken both from the literature on modality in Linguistics and from the one on the annotation of modality.

In this chapter, we resume the existing literature on modality and on its annotation, showing the systems of modality and the schemes for its annotation created by different authors. In the first part, we will explain what modality is and how it is analysed in the literature in English (Palmer, 1986; Bybee, 1994; van der Auwera, 1998; Huddlestone and Pullum, 2002) and Portuguese (Oliveira, 1988), showing which modal values the various authors identify. In the second part, we will show the schemes built for the annotation of modality in English by some of the authors who explored this theme in Computational Linguistics (Baker et al., 2010; Matsuyoshi et al., 2010; Saurí et al., 2007 and 2008; Szarvas et al., 2008; Wiebe et al., 2005). In the end, we will make a comparison between the approaches to the study of modality in Linguistics and in Computational Linguistics.

In the reading of the existing works on the annotation of modality, one of the main problems encountered has been that of choosing the terminology to use especially to refer to the investigation field studying the annotation of modality. As various possible terms (Computational Linguistics, Corpus Linguistics, Automatic Tagging, etc.) can be used to identify it, but none really suits the field we are referring to, we choose, for convenience, to use the term ‘Annotation’ to refer to the proposals done in Computational Linguistics for the annotation of modality in corpora, and the term ‘Linguistics’ to refer to the theories on modality in traditional Linguistics. We will also use the term ‘Restricted modality’ to refer to the notion of modality in the approaches presented in Linguistics and the term ‘Extended modality’ to refer to the notion of modality in the approaches presented in Annotation.

2.1 Restricted Modality

2.1.1 Modality: definition

Defining what modality is is not an easy task since, as Palmer (1986) underlines, «the notion of modality is vague and leaves open a number of possible definitions» (Palmer, 1986: 2). The concept of modality, in fact, can be explored according to two different scientific areas, the logical one (based on philosophy or mathematics) and the linguistic one. In the logical area, modality has been explored since Aristoteles and the main modalities identified are generally related to the notion of truth. In the linguistic area, which is the one which best fits our work, modality can be considered as an extra-propositional component of meaning (Baker et al., 2010), as it is related to the way meaning is presented and interpreted, and is usually defined as the expression of the speaker's opinion and of his attitude towards what he or other people are saying (Palmer, 1986; Oliveira, 2003).

The concept of 'modality' is associated to that of 'mood'. Semantically, both the terms express the speaker's attitude or opinion regarding the content of the sentence. However, traditionally the term 'mood' refers to a category expressed in the verbal morphology, whose semantic function is extended to the whole sentence, while 'modality' is a more general term, used to include all the linguistic elements that express the speaker's attitude toward what is being said (Palmer, 1986: 21 - 23). The difference can be captured by saying that mood is the expression of modality in the verb morphology.

Linguistically, modality can be realized at the morpho-syntactic, syntactic and lexical level. In morphology, modality is expressed by the mood of the verb in the sentence, becoming really a grammatical category. In syntax, modality is expressed through the syntactic form of the sentence: interrogatives can express doubt or uncertainty, imperatives express commands, declaratives express facts. Finally, when the modality values are expressed by specific words, such as adverbs and main verbs, it can be considered a lexical category (Palmer, 1986: 14 - 50).

2.1.2 Related work

In the previous section, we have introduced the concept of modality and have given a brief definition of the terms 'mood' and 'modality', defining modality as the linguistic expression of the speaker's attitude towards what he is saying or towards what others are saying. According to the attitude of the speaker, different modalities can be identified.

As we have seen, in Logics, Aristoteles associated the concept of modality to the notion of truth (Oliveira, 1988). In fact, the main modalities he identified were alethic modalities, since in Greek, *alethe* means 'truth'. They are linguistically expressed by sentences such as 'it is necessarily true/false that', 'it is really true/false that', 'it is possibly

true/false that', which in Portuguese are *é necessariamente verdadeiro/falso que*, *é realmente verdadeiro/falso que*, *é possivelmente verdadeiro/falso que*.

In Linguistics, various other types of modality can be identified. Observing the literature on modality in Linguistics, we come across different classifications. In this part, we present some of them: we will especially refer to the modal schemes presented by Oliveira (1988) for Portuguese, and by Palmer (1986), Bybee (1994), van der Auwera (1998), and Huddlestone and Pullum (2002) for English, observing how most of these modal schemes are created on the base of the same conceptual system which expresses the contrast between possibility and necessity. In all the sections related to the authors we have decided to present in most cases Portuguese constructed examples or examples taken from our corpus; only in few cases we report the original examples with their reference.

Oliveira (1988)

In the introductory part of her thesis, Oliveira (1988) briefly presents six types of modality, which according to her, are expressed in Portuguese. The author, however, does not show in detail each modality as the objective of her thesis is more that of showing the use of the verbs *dever* 'must' and *poder* 'can' than that of showing how modality is expressed in Portuguese; here, in fact, the examples we give are not taken by Oliveira, since she does not report any examples, but are constructed in order to better explain each modality. The modalities Oliveira (1988) identifies are:

- epistemic modality (*modalidades epistémicas*), which is related to the notions of knowledge and belief, and therefore is conveyed by linguistic elements expressing a different degree of truth of the proposition, such as, for example, the verbs *saber* (8) and *conhecer* 'know' or *crer* 'believe' (9) and *pensar* 'think';

(8) *Já **sei** que já não há bilhetes para o concerto*.
'I already know there are no more tickets for the concert'.

(9) *O João **crê** que foste tu a ligar-lhe*.
'John believes it was you who called him'.

- deontic modality (*modalidades deônticas*), which is related to the notions of obligation and duty, and is expressed by the verbs *ter de* (10) and *dever* 'must', *obrigar* 'oblige' (11), *impor* 'impose', *permitir* 'allow' (12) and synonyms;

(10) ***Temos de** pagar a conta do gás*.
'We have to pay the gas bill'.

(11) *Os desempregados são **obrigados** a apresentar-se nas Juntas de Freguesia para não perderem o subsídio de desemprego.*
'Unemployed people must go to the municipality in order not to lose the unemployment benefit'.

(12) *Não é **permitido** aos estudantes comer pastilhas nas aulas.*
'Students are not allowed to chew chewing gums during lessons'.

- volitive modality (*modalidades erotéticas*), which is related to the notions of will, hope and wish but also of fear and complaint, being expressed by the verbs *querer* 'want' (13), *esperar* 'hope', *desejar* 'wish', *temer* 'fear', *lamentar* 'complain' (14);

(13) *Estou muito cansada e **quero** ir para casa já.*
'I am very tired and I want to go home now'.

(14) *O João **lamenta** não poder ir ao concerto.*
'John is sorry for not being able to go to the concert'.

- evaluative modality (*modalidades avaliativas*), which deals with the speaker's evaluation of facts and situations, being conveyed by such expressions as *é bom que* 'it is a good that', *é uma pena que* 'it is a pity that' (15) and similars;

(15) ***É uma pena** teres de ir embora!*
'It is a pity that you have to leave!'

- causal modality (*modalidades causais*), which focuses the attention on the causes of occurrence of the event told in the proposition (*X causa Y* 'X causes Y', *X impedirá causar Y* 'X will prevent Y from happening') (16);

(16) ***O furacão destruiu** as casas.*
'The hurricane destroyed the houses'.

- temporal modality (*modalidades temporais*), which refers to the time of occurrence of the event told in the proposition, and is expressed by linguistic elements related especially to the frequency of occurrence of the event (*X acontece algumas*

vezes/sempre/a maior parte das vezes 'X happens sometimes/always/most of the times') (17).

(17) ***Às vezes*** *vou para casa a pé.*
'Sometimes I walk home'.

Through the reading of the next parts of this chapter, we will see that Oliveira (1988) is the only author, among the ones we analyse, who also presents 'causal and temporal modalities', which are generally not considered as belonging to modality.

Palmer (1986)

As Oliveira (1988), Palmer (1986) presents epistemic and deontic modality as well as volitive and evaluative modality.

As concerns epistemic modality, Palmer explains that the use he does of the term *epistemic* is wider than the usual one. In fact, the term *epistemic* is generally used to define something that is related to knowledge. On the other hand, Palmer uses it to define every context in which the speaker expresses his degree of commitment to what he says: the speaker can confirm (18), negate (19) or doubt (20) the content of the proposition (all the examples in this section on Palmer, except for the ones with reference, are the result of our introspection). In the case of a confirmation or of a negation, the speaker expresses strong epistemic modality, related to the domain of necessity; in the case of a doubt, the speaker expresses weak epistemic modality, related to the domain of possibility. It, then, turns out to be a subjective modality: in fact, the speaker has linguistically the possibility of giving the information with or without marking any type of involvement, any type of judgment or the kind of warrant he has for what he says. It is related to the idea that language can be used to convey information from different perspectives: the speaker can be just informing, he can be reporting something he heard or saw, or he can be expressing an opinion or a belief (Palmer, 1986).

(18) ***Estou certo*** *de que o João vem à minha festa de anos.*
'I am sure that John will come to my birthday party'.

(19) ***Estou certo*** *de que o João não vem à minha festa de anos.*
'I am sure that John will not come to my birthday party'.

(20) ***Duvido*** *que o João venha à minha festa de anos.*
'I doubt that John comes to my birthday party'.

Still within epistemic modality, he, then, classifies the propositions the speaker uses on the basis of the modal meaning they have. Four kinds of proposition are identified:

- *judgments*, propositions used to convey doubts and hypothesis that can be challenged or that may need evidentiary substantiation (Palmer, 1986). As they show the speaker's involvement, they are mainly subjective. Examples are "opinions" (21) and "conclusions" (22). The contrast between the two can be stated in terms of force of the statement: opinions are 'weak' epistemic judgments, while conclusions are 'strong' epistemic judgments. In English, the difference is very well represented by the verbs 'may' and 'must': the first is used for weak judgments, while the second is used for strong judgments. 'May' is the modal of epistemic possibility and is used to express the speaker's lack of confidence in the proposition (Palmer, 1986). On the other side, 'must' conveys the speaker's confidence in the truth of what he is saying, based on a deduction from facts known to him (Palmer, 1986).

(21) ***Pode*** *ser que ele venha.*
'He may come'.

(22) *Ele* ***deve*** *estar aí.*
'He must be there'.

- evidentials, propositions used by the speaker to support what he says with some kind of evidence. They are asserted with more confidence than judgments but require, or admit, evidentiary justifications by the speaker. They are then quite objective, but however involve also some degree of subjectivity since the kind of evidence is based on the speaker's commitment to what he says. The author explains that the type of evidence can be measured on the base of a scale. At the top of it, there is "visual evidence": a speaker can be reporting an event because he has personally participated in it or he has seen it (23); at the bottom, there is "assumed evidence": a speaker reports an event because he assumes it, but has no evidence of its realization (24).

(23) *Eles acabaram de sair da minha casa.*
'They have just left my house'.

(24) ***Diz-se*** *que no final do ano 2012 o mundo acabará.*
'It is said that in the end of the year 2012 the world will end'.

- interrogatives express doubt and uncertainty, and indicate that the speaker does not know whether or not the sentence is true (Palmer, 1986). Palmer (1986) reports that Matthews (1965: 99 – 100) divides them in two categories: the Indefinite, which is used by the speaker as a device to express that he thinks the listener too does not know if the sentence is true (25), and the Question, which is used by the speaker to express that he thinks the listener does know (26).

(25) ***Sabes aonde é que o João está?***
'Do you know where John is?'

(26) ***O João vem?***
'Is John coming?'

Palmer (1986), however, does not accept this classification because it is based on the author's own interpretation and rationalization and not on evidence from the language.

- declaratives are propositions which are stated as certain and which cannot be questioned by the hearer. Semantically, declaratives are linked with what the speaker knows (27) or believes (28), but we can argue that it is more probable that the declarative expresses belief instead of knowledge; this is because, generally, what the speaker says is taken as more subjective, like a belief, while knowledge is expected to be more neutral, more objective. Being generally unmarked morphologically, not all authors include declaratives within modality: Palmer (1986) includes them within it, saying that they are the unmarked member of the modal system; on the other side, Oliveira (1988) does not accept this idea, saying that if declaratives are modal, all language is, then, modal.

(27) *O João foi para casa.*
'John went home'.

(28) ***Penso*** *que o João foi para casa porque tinha que trabalhar muito.*
'I think John went home because he had to work a lot'.

If epistemic modality is related to belief and knowledge, deontic modality is related to action, as it regards the use of language to make action start, such as in orders and

permissions, and concerns all those types of modality containing an element of will (Jespersen, 1949, in Palmer, 1986). It can have two interpretations, one of "permission" and another of "obligation". The first interpretation is linked to the notion of possibility, as permission is one of the conditions that make possible the realization of the action (29), while the second interpretation, that of obligation, belongs to the domain of necessity (30).

(29) *A professora **permitiu** aos estudantes fazer uma pausa.*
'The teacher allowed the students to have a break'.

(30) *A professora **obrigou** os estudantes a trabalhar durante a hora do almoço.*
'The teacher forced the students to work during the break time'.

However, in some cases the same deontic proposition can convey both the values of permission or obligation according to how we interpret the sentence (31).

(31) ***Entra**.*
'Come in'.

The most prototypical expressions of deontic modality are directives, defined by Searle (1983: 166) apud Palmer (1986: 97) as those expressions used when we try to get our hearers to do things (32 – 33).

(32) ***Deves** ir para casa agora.*
'You must go home now'.

(33) ***Podes** ir para casa agora.*
'You may go home now'.

Within directives we can find imperatives, which differ from directives because they use the main verb in the imperative form. An imperative is a deontic proposition whose force has to be judged by the hearer: it can be taken as an order, to which the speaker necessarily has to obey (34 a and b), or as an expression of permission (34 c). Imperatives are the "purest" directives, in fact, and are considered to be the unmarked member of deontic modality (Palmer, 1986).

(34) *a. **Levanta**-te!*
'Stand up!'

b. ***Senta****-te!*

'Sit down!'

c. ***Entra****.*

'Come in'.

As we can see from the examples, deontic modality is strictly linked to the future as the action can only be realized in the period following the speech act, which is the only period of time which can be changed or affected from what has been said.

In a small section of his work, Palmer (1986) reports on the use of 'may' and 'must' in English, which are typically associated the first to epistemic modality, since it generally expresses possibility, and the second to deontic modality, since it generally expresses necessity. However, the author shows the possibility of interpreting 'must' in an epistemic way. In (35), in fact, 'must' is ambiguous, as it can be interpreted as expressing both possibility and necessity: it has an epistemic meaning and expresses possibility if we interpret it as uncertain (36); it has a deontic meaning and expresses necessity if we interpret it as an obligation (37).

(35) *John* ***must*** *come.*

(36) *John* ***must*** *come but we don't know if he will.*

(37) *John* ***must*** *come since his presence is required for legal reasons.*

The author also introduces dynamic modality: this modality is related to the notions of 'ability' and 'disposition', and marks when the action is possible (dynamic possibility) or prevented (dynamic necessity) by some physical (38) or mental condition (39). The author affirms that this modality takes into account the level of involvement of the speaker in the action, who may be totally involved, not involved at all or involved as a member of the society or body that instigates the action.

(38) *O João ainda não* ***pode*** *participar no jogo de amanhã por causa do joelho.*

'John cannot yet play in the match tomorrow because of his knee'.

(39) *O João* ***consegue*** *falar chinês porque viveu na China durante algum tempo.*

'John can speak Chinese as he lived in China for a while'.

Moreover, Palmer (1986) presents commissives (40), which are used when the speaker commits himself to do something. Based on the work of Searle (1979: 14), the author

initially compares them to directives: both are, in fact, subjective and very much performative as they initiate action by others (directives) or by the speaker himself (commissives). He, in fact, presents them in the chapter on deontic modality, as they both express some kind of obligation, but do not include them under the same label since through directives the speaker expresses an obligation on someone else, while through commissives he expresses an obligation on himself.

(40) *John **shall** have the book tomorrow.* (Palmer, 1986: 115)

Some examples are promises and threats (41 – 42).

(41) *You **shall** go to the circus.* (Palmer, 1986: 115)

(42) *Se tiveres bons resultados na escola, comprarei uma bicicleta para ti.*
'If you have good results at school, I will buy you a bike'.

In sentence (40), the speaker commits himself to give the book to the hearer, while in sentence (41), he commits himself to make the hearer go to the circus; and in sentence (42), he commits himself to buying a bicycle to the hearer. In the three examples, the person who has to make the action is not the hearer but the speaker himself.

The definition of commissives is very similar to that of epistemic modality, as both regard the involvement of the speaker towards something. The difference is that epistemic modality focuses on the commitment of the speaker towards the truth of the proposition, while commissives focus on the commitment of the speaker towards action.

Also volitives and evaluatives are presented within the chapter on deontic modality but the author explains that, in reality, they are neither deontic nor epistemic, as they do not express commands, but neither express the commitment of the speaker towards what he is saying.

In contrast with Oliveira (1988), in the work of Palmer (1986), volitives do not include wants, fears and complaints, but only include hopes and wishes: hopes are desires that can be realized (43), while wishes are desires that cannot be realized (44). The main reason to consider them as modal is that they involve non-factuality being concerned more with possible action than with the truth of the proposition (Palmer, 1986). Volitives are, in fact, linked to the notion of possibility: hopes express possibility, while wishes impossibility. However, they can also express necessity if the speaker presents the desire as a necessity, as in (45), where the same sentence has, simultaneously, both the value of possibility and of necessity.

(43) ***Espero*** *que o Mico não sinta muito a minha falta quando estiver em Itália.*
'I hope Mico will not suffer from my absence when I am in Italy'.

(44) *Gostava que o Miguel fosse comigo para o festival na Turquia mas ele não pode porque nessa altura estará a trabalhar.*
'I would have liked Miguel to come with me to the festival in Turkey but he can't since he will be working at the time'.

(45) *O Miguel* ***tem de*** *comprar o carro para podermos ir para Espanha no verão.*
'Mike has to buy the car so that we can go to Spain in the summer'.

Fears and complaints, on the other hand, are included by Palmer (1986) within evaluatives, together with other feelings, such as surprise, disappointment, approval or disapproval, as they all express some kind of evaluation the speaker gives about facts and situations (46) or about someone's psychological state (47). Evaluatives are, in fact, defined as attitudes towards known facts. Palmer (1986) underlines that they are not always considered as modal, but that sometimes they can be related to modal systems because they express the speaker's attitude towards what he already accepts as true.

(46) *As condições de vida nas favelas do Rio de Janeiro são mesmo* ***desumanas***.
'Living conditions in Rio's slums are really inhuman'.

(47) *O João tem* ***medo*** *do escuro.*
'John is afraid of the dark'.

Bybee (1994)

Bybee (1994) chooses to oppose epistemic modality to agent-oriented and speaker-oriented modality. As we have seen in the previous two sections, epistemic modality is related to how the speaker presents his commitment to the truth of the proposition: the speaker can present something as certain (48), as probable (49) or as possible (50) (all the examples in this section are constructed since Bybee does not give any example in the original text).

(48) *Hoje de manhã falei com a minha mãe.*
'This morning I talked to my mom'.

(49) *O Paulo já **deve** estar no trabalho agora.*
'Paul should already be at work now'.

(50) ***Pode** ser que entregue a tese em Dezembro.*
'I may hand in the thesis in December'.

Agent-oriented modality and speaker-oriented modality, on the other hand, include all those cases in which action is involved. Agent-oriented modality takes into account the internal and external conditions that influence the agent with respect to the completion of the action expressed in the main predicate. Within it, he identifies four different sub-values, the first two related more to necessity and the second two related to possibility: obligation, which considers the external conditions compelling an agent to complete the action (51), necessity which presents the physical conditions influencing the accomplishment of the action (52), ability, which focuses on the capacities of the agent influencing the completion of the action (53), and corresponds, then, to Palmer's dynamic modality, and desire (54). Speaker-oriented modality, on the other side, includes all the speech acts aiming at getting something done. Within it, we find some speech acts which are linked to the notion of necessity, such as imperatives, which express orders (55); permissives, used when the speaker allows the hearer to do something (56); prohibitives, used when the speaker forbids the hearer to do something (57); admonitives, used by the speaker to warn the hearer about something (58); other speech acts are instead linked to the notion of possibility, such as volitives, which he calls 'optatives' and express hopes, wishes and desires of the speaker or participant (59); and hortatives, which are used to prompt someone to do something (60) (Bybee, 1994).

(51) *O pai **obrigou** a filha a estudar matemática uma hora cada dia para ver se ela melhorava as notas.*
'The father obliged his daughter to study Mathematics one hour each day for her marks to improve'.

(52) *O João não **pode** correr porque torceu o joelho direito.*
'John cannot run as he twisted the right knee'.

(53) *Os meus pais não **sabem** andar de bicicleta.*
'My parents do not know how to ride a bicycle'.

(54) *O João só **queria** ter mais tempo.*
'John just wished he had more time'.

(55) ***Come*** *e* ***cala****-te!*
'Shut up and eat!'

(56) ***Podes*** *ir dormir à casa dos teus primos no sábado.*
'You can sleep at your cousins' house on Saturday'.

(57) *É* ***proibido*** *entrar na piscina sem toca.*
'It is forbidden to enter the swimming pool without bathing cap'.

(58) ***Não te apoies*** *a esse pau senão cai tudo!*
'Don't lean on that stick, otherwise everything falls down!'

(59) ***Espero*** *entregar a tese até Dezembro.*
'I hope I will hand in the thesis till December'.

(60) ***Despacha****-te a arrumar, que é para ires embora mais cedo!*
'Hurry up to clear up, so that you leave sooner!'

Van der Auwera (1998)

Van der Auwera (1998) opposes epistemic modality to participant-internal modality, participant-external modality and deontic modality, which is considered as a sub-value of participant-external modality. If epistemic modality is related to the level of commitment of the speaker to the truth of the proposition, participant-internal modality defines a kind of modality internal to the participant engaged in the state of affairs (van der Auwera and Plungian, 1998). It refers either to a capacity, in which case it corresponds to dynamic modality in the work of Palmer (1986), or to a need internal to the participant that makes possible the realization of the action. Both epistemic modality and participant-internal modality take into account the subjectivity of the participant, speaker or subject of the sentence, but epistemic modality is more linked to the notions of knowledge and belief, while participant-internal modality is more linked to action, as it focuses on the notions of capacity and necessity which lead the participant to act.

Within participant-internal modality, we can, then, distinguish between necessity and possibility: participant-internal necessity is expressed when the action depends on a personal need of the participant (61), while participant-internal possibility is expressed when the action depends on the participant's capacity (62). According to van der Auwera and Plungian (1998),

this capacity can be learnt (63) or inherent (64)[3].

(61) *Boris **needs** to sleep ten hours every night to function properly.* (van der Auwera and Plungian, 1998: 80: 1b)

(62) *Boris **can** get by with sleeping five hours a night.* (van der Auwera and Plungian, 1998: 80: 1a)

(63) *Aos quatro anos, o João já **sabia** andar de bicicleta.*
'When he was four years old, John already knew how to ride a bicycle'.

(64) *Sem óculos, o João não **consegue** ler.*
'Without glasses, John cannot read'.

Also participant-external modality considers the conditions influencing the action, but, in this case, the conditions are not given by the participant but by factors that the participant cannot control. It, therefore, includes deontic modality, as even deontic modality identifies the external circumstances that make the participant engage in the state of affairs.

Also within participant-external modality, we identify the values of possibility (65) and necessity (66).

(65) *To get to the station, you **can** take bus 66.* (van der Auwera and Plungian, 1998: 80: 2a)

(66) *To get to the station, you **have to** take bus 66.* (van der Auwera and Plungian, 1998: 80: 2b)

In sentence (65), bus 66 is one of the ways to get to the station: the participant has the 'possibility' to choose. In sentence (66), bus 66 is the only way to get to the station, taking it is, then, a 'necessity'.

As we can see, van der Auwera's organization of the modal values is different from Bybee's one: Bybee (1994) joins together participant-internal and participant-external under the label 'agent-oriented' and opposes it to speaker-oriented modality; deontic modality, in Bybee's system, would be considered both within agent-oriented modality and speaker-oriented modality, depending on who establishes the obligation, if it is the agent or the speaker; van der Auwera (1998), on the contrary, does not consider speaker-oriented modality

[3] Exemples (63) and (64) are the result of our introspection.

and separates participant-internal from participant-external modality, considering deontic as a sub-domain of the participant-external value.

Huddlestone and Pullum (2002)

As Palmer (1986), Huddlestone and Pullum (2002) distinguish between epistemic, deontic and dynamic modality. However, they first present a contrast between weak and strong modality, built on the base of the strength of the commitment to the factuality or actualization of the situation. This contrast corresponds to the one between possibility and necessity, as possibility involves a weak commitment (67)[4], while necessity a strong one (68).

(67) *Se calhar vai chover.*
'It may rain'.

(68) *Não pode chover ao fim-de-semana!*
'It cannot rain in the weekend!'

The authors, then, explain that there is another distinction between modalities which cuts across the one based on strength: this is the one between epistemic and deontic modality. Both within epistemic and deontic modality we can, in fact, find strong and weak modality: within epistemic modality, sentence (69) is an example of weak modality, while sentence (70) is an example of strong modality; within deontic modality, (71) expresses weak modality and (72) the strong one.

(69) ***Pode*** *ser que chegue.*
'He may arrive'.

(70) ***Deve*** *chegar em breve.*
'He must arrive soon'.

(71) ***Podes*** *almoçar connosco.*
'You may have lunch with us'.

(72) ***Tens de*** *almoçar connosco.*
'You must have lunch with us'.

As Palmer (1986), Huddleston and Pullum also introduce dynamic modality, to refer

[4] All the examples in this section are constructed except for sentence (73).

to the notions of ability and disposition (73).

(73) *She **can** speak French.* (Huddlestone and Pullum, 2002: 178)

From an overall review of all the systems, we see that in all the classifications done by the authors we can find two constants. The first one, as we have seen at the beginning of this chapter, is that all the classifications are based on the same conceptual system built on the contrast between possibility and necessity. The second one is that they all identify the epistemic value; differences arise at the moment of choosing which are the values contrasting with epistemic modality.

Table 1 shows the modal values presented in the literature on modality in Linguistics by the different authors. The table is composed of three columns. In the first column, we show the distinction between strong and weak modality. In the second column, we report only epistemic modality, since it is the only value in common to all systems. In the third column, we report the values the different authors present in opposition to epistemic modality. In this way, we graphically show how all the authors identify epistemic modality but also how they all identify different values in opposition to epistemic modality. As we can see, three authors (Palmer, 1986; Oliveira, 1988; Huddlestone and Pullum, 2002) report the distinction between epistemic and deontic modality. Among them, Palmer (1986) and Huddlestone and Pullum (2002) also consider dynamic modality, while Oliveira (1988) does not. Palmer (1986) and Oliveira (1988) also consider volitives and evaluatives, while Huddlestone and Pullum (2002) do not. Oliveira (1988) is the only author presenting also temporal and causal modalities, and Palmer (1988) is the only author introducing commissives. Among the systems presented, two completely differ from all the others, Bybee (1994)'s one and van der Auwera (1998)'s one. Neither of them presents the opposition between epistemic and deontic modality; both identify epistemic modality but oppose different values to it: Bybee presents speaker-oriented and agent-oriented modality, while van der Auwera presents participant-internal and participant-external modality.

Since the systems of modality in Linguistics are often based on a different classification of modal values, we have decided to group the values presented in those systems in five big categories. Table 2 presents two columns: in the first, we show the five big categories of modal values; in the second, we present which are the values identified in the different systems in Linguistics included in the five categories.

TABLE 1. MODAL VALUES IN LINGUISTICS

STRONG MODALITY VS WEAK MODALITY	EPISTEMIC MODALITY Evidentiality Certainty Probability Possibility	DEONTIC MODALITY Obligation Permission VOLITIVE MODALITY (wants; fears; complaints) EVALUATIVE MODALITY CAUSAL MODALITY TEMPORAL MODALITY (Oliveira, 1988) DEONTIC MODALITY Obligation Permission DYNAMIC MODALITY COMMISSIVES VOLITIVES (hopes; wishes) EVALUATIVES (fears; complaints; surprise; disappointment; approval/disapproval) (Palmer, 1984) AGENT-ORIENTED MODALITY Obligation Necessity Ability Desire SPEAKER-ORIENTED MODALITY Imperatives Permissives Prohibitives Admonitives Optatives (volitives) Hortatives (Bybee, 1994) PARTICIPANT-INTERNAL MODALITY Necessity Possibility (capacity) PARTICIPANT-EXTERNAL MODALITY Necessity → DEONTIC MODALITY Possibility EVIDENTIALITY VOLITION (van der Auwera, 1998) DEONTIC MODALITY DYNAMIC MODALITY (Huddlestone and Pullum, 2002)

TABLE 2. COMPARISON SHOWING THE RELATION BETWEEN THE BIG CATEGORIES OF MODAL VALUES WE ESTABLISHED AND VALUES OF THE VARIOUS SYSTEMS IN THE LITERATURE ON MODALITY IN LINGUISTICS

EPISTEMIC MODALITY	EPISTEMIC MODALITY (Oliveira, 1988; Palmer, 1986; Bybee, 1994; van der Auwera, 1998; Huddlestone and Pullum, 2002)
DEONTIC MODALITY	DEONTIC MODALITY (Oliveira, 1988; Palmer, 1986; van der Auwera, 1998; Huddlestone and Pullum, 2002) AGENT-ORIENTED MODALITY (obligation) (Bybee, 1994) SPEAKER-ORIENTED MODALITY (imperatives; permissives; prohibitives; admonitives) (Bybee, 1994) PARTICIPANT-EXTERNAL MODALITY (van der Auwera, 1998) COMMISSIVES (Palmer, 1986)
PARTICIPANT-INTERNAL MODALITY	DYNAMIC MODALITY (Palmer, 1986; Huddlestone and Pullum, 2002) AGENT-ORIENTED MODALITY (Bybee, 1994) PARTICIPANT-INTERNAL MODALITY (van der Auwera, 1998)
VOLITION	VOLITIVE MODALITY (Oliveira, 1988) VOLITIVES (Palmer, 1986) SPEAKER-ORIENTED MODALITY (optatives) (Bybee, 1994) AGENT-ORIENTED MODALITY (desire) (Bybee, 1994) VOLITION (van der Auwera, 1998)
EVALUATION	EVALUATIVE MODALITY (Oliveira, 1988) EVALUATIVES (Palmer, 1986)

In Table 2, *epistemic modality* is the common value to all the systems.

Deontic modality is defined as such by most authors (Oliveira, 1988; Palmer, 1986; van der Auwera, 1998; Huddlestone and Pullum, 2002) and is the main value identified in opposition to epistemic modality; however, in the system of van der Auwera (1998), deontic modality is considered as a subvalue of participant-external modality, while in Bybee's system the category deontic modality is not considered but the subvalues included in it are considered within speaker-oriented or agent-oriented modality: the subvalue obligation is included within agent-oriented modality, and the subvalues imperatives, permissives, prohibitives and admonitives within speaker-oriented modality. The big category *deontic modality* also includes the value commissives identified by Palmer (1986), since this value expresses an obligation the speaker gives to himself, committing himself to do something.

We then identify *participant-internal modality*, which is the value expressing personal capacities and necessities of the participant. It is defined as participant-internal modality by van der Auwera (1998) and corresponds to what Palmer (1986) and Huddlestone and Pullum (2002) define as dynamic modality. In Bybee's system (1994), participant-internal modality is included within agent-oriented modality, since both agent-oriented and participant-internal modality include the subvalues of necessity and capacity.

Moreover, our big category *volition* corresponds to the modal value volition identified by most authors as a separate value, which includes wants, hopes and wishes of the speaker or participant. Even as concerns this value, Bybee's classification is different, since the author splits volitives in two categories: optatives, included in speaker-oriented modality, and desire, included in agent-oriented modality.

In the end, the category *evaluation*, including *evaluation* identified only by Oliveira (1988) and *evaluatives* identified by Palmer (1986), is used to identify the speaker's evaluation of facts and propositions.

So far, we have presented the systems of modality created by different authors in the literature on modality in Linguistics. We have especially paid attention to the modal values identified in the different systems, observing the most important similarities and differences. Through Table 2, we have also shown how in most cases the values identified are the same but are organized in a different way and have different names.

In the next section, we will present the existing literature on the annotation of modality.

2.2 Extended modality

In this second part of the chapter we introduce the existing works on the annotation of modality in corpora. First, we will explain how the notion of modality is "extended" in the computational field, if compared to that in Linguistics. Then, we will point out some of the reasons for the annotation of modality and explore the works already written in this field, observing which modal values the authors believe that are important to annotate and pointing out similarities and differences among the various annotation schemes proposed.

As concerns modality, we have noticed that the two approaches are based on a different notion of modality. Observing the proposals for the annotation of modality, we come across a broader concept of modality, which we define here as 'extended modality': as Annotation has practical purposes, the word 'modality' is not used only to refer to the attitude of the speaker towards what he is saying (restricted modality), but also indicates how the speaker expresses his opinions, feelings and emotions. Further on, often, in Annotation, the concept of 'modality' is associated to that of 'factuality', to mark if the event has really

happened, and to that of 'polarity', to mark if the modality is positive (74)[5] or negative (75).

(74) *Quero honrar o nome de Cubillas.*
'I want to honour Cubillas' name'.

(75) ***Não** quero honrar o nome de Cubillas.*
'I don't want to honour Cubillas' name'.

Moreover, the term 'extended modality' is used also to refer to all the elements that are involved in the expression of modality, such as the holder of the modality and the linguistic expression in the scope of the modal word. In the end, Matsuyoshi et al. (2010) include within the term 'extended modality' also the analysis of the mood, the time and the aspect of the verb in the sentence.

2.2.1 Why annotating modality?

The annotation of modality in corpora is a quite recent practice. In particular, the interest for the annotation of modality has grown in order to distinguish factual from non-factual information; annotating modality, in fact, helps in distinguishing if the information is presented as a certainty, as a possibility or as a probability, if it is presented as a personal opinion or belief or if it is an acquired knowledge.

In Linguistics, annotating modality in real texts is important for the study of a language, as it helps in identifying which expressions the speakers use to show their involvement towards what they are saying, or towards what others are saying, or even towards facts.

Moreover, annotating modality in a Portuguese corpus enables linguists to investigate modality in Portuguese, having the possibility to immediately identify modal values in context, to identify the various different expressions for modality and to observe the behaviour of lexical and syntactic items when expressing modality.

Apart from Linguistics, the annotation of modality can be useful in other investigation fields, such as, for example, politics, and the commercial and medical domains. Wiebe et al. (2005), for example, report that, in the political and commercial fields, information analysts want to automatically track attitudes and feelings in the news and in on-line forums. For example, automatic annotation of modality can be used to check which are the opinions and feelings of people about some important recent event such as the big meeting of NATO in Lisbon in November 2010.

Further on, the existence of corpora annotated with modal information would also

[5] All the examples without bibliographic references nor footnotes are from the corpus.

help in the development and evaluation of Natural Language Processing (NLP), a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. Within NLP, different applications are built in order to automatically perform tasks which are generally related with the analysis of language use in text (*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, by Daniel Juraksfy and James E. Martin. McGraw Hill, 2008 (2nd edition)). Here we mention the ones that may profit from having modality annotated in text (Baker et al., 2010; Matsuyoshi et al., 2010; Saurí et al., 2006; Wiebe et al., 2005); these are:

- Machine Translation, which automatically translates text from one language to another and would profit of having text where modality is annotated to improve translation as modality is expressed differently in different languages;
- Information Extraction, concerned with the extraction of important semantic information from the text, would profit of having modality annotated in text to distinguish relevant from irrelevant information about some specific topic;
- Recognizing Textual Entailment: given two text fragments, this task requires to recognize whether the meaning of one text is entailed (can be inferred) from the other text; even here, the annotation of modality would be helpful for the distinction between relevant and irrelevant information;
- Automatic Summarization, which produces a summary of a chunk of text, would profit from having text annotated with modal information to identify the most important information;
- Flame Detection or Sentiment Analysis: it extracts subjective information from a set of documents and is especially useful for identifying trends of public opinion in the social media, for the purpose of marketing; having text annotated with modal information would, in fact, help these systems in distinguishing between opinions and factual information;
- Question Answering, which, given a human-language question, determines its answer and is useful to automatically answer those questions which have a specific right answer, such as, for example, the question *What is the capital of Canada?* which can only have *Ottawa* as answer; as for Automatic Summarization, the annotation of modality in text would be useful for Question Answering to identify the most important information related to one topic.

2.2.2 Related work

Although the annotation of modality in corpora is a quite recent field of investigation, some important works have been written on this theme. All of them, however, report research

projects on the annotation of modality in languages other than Portuguese, as, till now, no research has been undertaken on the annotation of modality in a Portuguese corpus. In the following sections we will present each of these works.

Saurí, Verhagen and Pustejovsky (2006) and Saurí and Pustejovsky (2007)

One of the main reasons for annotating modality in corpora is that of distinguishing between factual and non-factual events. The work of Saurí and Pustejovsky aims at determining modality and factuality in text in order to recognize Textual Entailment. Textual Entailment, in fact, would benefit from a system classifying events on the basis of their degree of factuality, as an inference done on the basis of a factual event varies from one based on an event judged as possible or non-existent.

In their work, the authors identify events in the text and for each event build a factuality profile, which is a set of the factuality values that different sources[6] assign to the event: an event can be presented as corresponding to a real situation in the world (factual), to a situation that has not happened (non-factual), or to a situation of uncertain status (unknown). Sentences (76) and (77), for example, are factual, while sentences (78) and (79) are non-factual, because they denote something that did not happen; in sentences (80) and (81)[7], the factuality is unknown because there is no information whether the event took place or not.

(76) *Italian royal family returns home.* (Saurí and Pustejovsky, 2007: 2: 3H)

(77) *O Miguel já chegou.*
'Mike already arrived'.

(78) *The size of the contingent was* ***not*** *disclosed.* (Saurí and Pustejovsky, 2007: 2: 8)

(79) *O Miguel ainda* ***não*** *chegou.*
'Mike has not arrived yet'.

(80) *Mary did not say John is not sick.* (Saurí and Pustejovsky, 2007: 4: 14c)

(81) *Como ainda não falei com a minha mãe,* ***não sei se*** *ela já chegou a França.*
'As I haven't spoken with my mother yet, I don't know if she has already arrived in France'.

However, not all events can be classified on the basis of a binary value system: more

[6] Saurí et al. (2007) consider sources the "producers" of the event mention (speakers or writers).

[7] Examples (77), (79) and (81) have been constructed to show polarity contrast.

values can be identified between positive and negative factuality. Table 3 (Saurí and Pustejovsky, 2007:2:Table 1), presented at the beginning of the following page, shows the event factuality values the authors identify.

The table shows the relation between epistemic modality and polarity. As we already have explained in the first section of this chapter, epistemic modality is conveyed whenever the speaker shows his commitment to the truth of the proposition. In the table, the epistemic modality values are shown in the horizontal rows: besides the values *certain*, *probable* and *possible*, the authors consider the value *unknown*, in order to classify events in which modality is not shown. In the vertical columns, the authors show polarity, a component used to mark if the modal value expressed is positive or negative: besides positive polarity, which defines the factuality of the event, and negative polarity, which defines the non-factuality of the event, the authors also identify the value 'unknown', if the source does not know whether the event took place or not.

The authors also explain that factuality is not objective but depends on the commitment of the source: whenever there is a mention of an event, there is a commitment act towards the factuality of that event, performed by a particular source. Two kinds of sources can be identified: an 'anchor', which is the source of the event mention and is generally the speaker or writer of the sentence, and a 'cognizer', which is the source of the modality. This distinction is important because an event can have a different factuality value depending on the source.

(82) *Mary regrets that John does not know he is sick.* (Saurí and Pustejovsky, 2007: 5: 16b)

In sentence (82) three sources can be identified: the *speaker* or *writer*, which is the source of the event mention (anchor); *Mary*, which is the source of the modality expressed by *regret* (cognizer) in the main clause; and *John*, which is the source of the modality expressed by *does not know* (cognizer) in the first subordinate clause. According to the source of the event mention (speaker or writer) and to the source of the modality in the main clause (Mary), the fact of John being sick is factual, while according to the source of the modality in the first subordinate clause (John), the fact of John being sick is non-factual.

TABLE 3.

	Positive (+)	Negative (-)	Unknown
Certain	Fact: <CT,+>	Counter-fact: <CT,->	Certain but unknown output: <CT,UN>
Probable	Probable: <PR,+>	Not probable: <PR,->	(NA)
Possible	Possible: <PS,+>	Not certain: <PS,->	(NA)
Unknown	(NA)	(NA)	Unknown or uncommitted: <UN,UN>

According to the authors, the distinction between positive (83) and negative factuality (90)[8] is not always clear: in between the two values, there is a continuum of values that go from the most factual one (84) to the most counter-factual one (89):

- Factuality: the source presents the event as a fact (83);

  (83) *John **is** sick.*

- Evidentiality: the source gives some kind of evidence for what he is saying (84);

  (84) *Subcomandante Marcos **said** that the Mexican government is not interested in putting an end to the conflict.*
  (Saurí et al., 2006: 1: 2c)

- Command: the source gives an order (85);

  (85) *John Murtha **called** for the immediate withdrawal of U.S. troops from Iraq.*
  (Saurí et al., 2006: 1: 2f)

- Degrees of possibility: the source presents the event as possible or impossible, probable or improbable; it then is related to epistemic modality (86);

  (86) *These results indicate that Pb2+ **may** inhibit neurite initiation by inappropriately stimulating protein phosphorylation by CaM kinase.*
  (Saurí et al., 2006: 1: 2a)

[8] Examples (82) and (89) are constructed.

- Attempting: the source tries to do something (87);

(87) *George Mallory and Andrew Irvine first **attempted** to climb Everest in 1924.*
(Saurí et al., 2006: 1: 2e)

- Expectation: the source expects or wants something to happen (88);

(88) *Hans Blix **wants** the US to allow UN inspectors back into Iraq to verify any weapons found by coalition forces.*
(Saurí et al., 2006: 1: 2d)

- Belief: the source thinks or believes in something (89);

(89) *Chinese analysts **believe** that the United States will continue to provoke North Korea.*
(Saurí et al., 2006: 1: 2b)

- Non-factuality: according to the source, the event did not happen (90).

(90) *John **is not** sick.*

Baker et al. (2010)

Baker et al. (2010) also annotate modality. They especially aim at creating an annotation scheme of modality that could be useful for machine translation programs, which are programs used to automatically translate text from one language to another. In order to improve automatic translation, investigation in this field has started to combine machine translation programs with corpus linguistics and statistics, as it may help in handling differences in linguistic typology, in the translation of idioms and in isolating anomalies.

Machine translation programs often imply the use of taggers, which are machines that identify the linguistic elements in the text and assign them a label which gives information about that word. The most used tagging process is POS tagging, which labels parts-of-speech in the text. However, Baker et al. (2010) want to identify how modality is expressed and tag its components, as they believe that identifying modalities in text would help in improving translation. The tagger they use, in fact, has been trained in order to be able to identify the trigger for the modality, which is the linguistic modal expression, and its target, which is the event or relation in the scope of the modality, but was not trained to identify the source of the

modality (holder), as the identification of the sources was beyond their interests. The training for recognizing the modal values to associate to the triggers was done on a selection of modal values which was restricted only to eight modal values related to factuality. These are:

- Requirement: the source expresses a necessity (91);

  (91) *O João **precisa** que a Maria compre o açúcar para fazer o bolo.*
  'John needs Mary to buy sugar to make the cake'.

- Permissive: the source allows something to happen (92)[9];

  (92) ***Podes** entrar.*
  'You can come in'.

- Success: the source succeeds in something (93):

  (93) *A França **alcançou** a vitória já por 18 vezes.*
  'France reached the victory already 18 times'.

- Effort: the source tries to do something (94);

  (94) *Covilhã **tenta** revitalizer folclore.*
  'Covilhã is trying to revitalize folklore'.

- Intention: the source expresses the intention to do something (95);

  (95) *Facções rebeldes **pretendem** unificar-se em Cabinda.*
  'Rebellious factions pretend to unify themselves in Cabinda'.

- Ability: the source is capable or not to do something (96);

  (96) *O Governo não **soube** criar nem confiança nem credibilidade.*
  'The Government wasn't able to create trust or credibility'.

- Want: the source expresses a desire or a hope or a want (97);

[9] Examples (91) and (92) are constructed.

(97) *PSD **quer** que a Câmara de Braga assuma responsabilidades.*
'The PSD wants Braga's Town-Hall to shoulder its responsibilities'.

- Belief: the source expresses her belief in something (98).

(98) *Abraão **acreditou** e foi abençoado; e assim todos aqueles que crêem serão abençoados com ele.*
'Abraham believed and was blessed; so all who believe are blessed as he was'.

We can observe that the modal values they identify are more linked to action than to thinking: *permissive*, *success*, *effort* and *ability* imply that there is an action that has to be performed or that has been already performed. Only the values *belief* and *intention*, in fact, are more linked to epistemic modality.

Once the tagger has been trained and is able to recognize the components of modality and the modal value expressed, it can be used to improve machine translation output by imposing semantic constraints on possible translations when facing sparse training data.

A general overview of the two annotation schemes presented above shows that both schemes are centred on distinguishing factual from non-factual events: as this distinction is very important for translation, Baker et al. (2010) choose to train the tagger only with modal values centred on factuality; for Saurí et al. (2006 and 2007) the same distinction is important as an inference based on a real occurred fact is different from an inference based on a probable, possible or non-existing event.

However, there is a big difference too between the two annotation schemes: Saurí et al. (2006 and 2007) annotate also the sources, as different sources can present an event with a different factuality value, while to Baker et al. (2010), the identification of the sources is not important for the training of the taggers, because they want, first, to train the tagger to match the modal triggers to their targets.

Matsuyoshi et al. (2010)

Till now, we have reported works where the authors decided to annotate mainly modal values, triggers, targets (Baker et al., 2010), sources and polarity (Saurí et al., 2006 and 2007). Matsuyoshi et al. (2010) believe that NLP's applications such as Information Extraction, Question Answering and Recognizing Textual Entailment require analysing also the mood, the tense and the aspect of the modal verb, besides the components observed by the

other authors.

As concerns restricted modality, the authors select few modal types, which they include within the category called 'primary modality type'. These are:

- Assertion: the source presents something as a fact (99).

  (99) *He **said** he suffered from symptoms due to stopping steroid medicines at that time*. (Matsuyoshi et al., 2010: 1463: ID 2)

- Volition: the source expresses something he wants to do (100).

  (100) *If it is nice out tomorrow, I **will** go fishing in that lake*. (Matsuyoshi et al., 2010: 1463: ID 6-7)

- Wish: the source expresses a personal wish or desire (101).

  (101) *Taro said he **wanted** to go home soon*. (Matsuyoshi et al., 2010: 1463: ID 1)

- Imperative: the source gives an order (102).

  (102) ***Feel** for yourself the effects of restoration water!* (Matsuyoshi et al., 2010: 1463: ID 8)

- Permission: when the source gives the hearer the permission to do something (103).

  (103) *You **may** use a larger desk*. (Matsuyoshi et al., 2010: 1463: ID 11)

- Interrogative: when the source asks for some information or expresses a doubt (104).

  (104) ***Is it because Taro received appropriate nutrition that he recovered?*** (Matsuyoshi et al., 2010: 1463: ID 20)

The other components the authors choose to annotate are:

- Source, which expresses the agent or an organization that takes an attitude toward an event mention in a sentence; there can be three different kinds of sources, other than

the writer: an agent (105), an agent represented by a pronoun (106) or an arbitrary source, when the event mention is an overheard statement and, therefore, the source expresses second-hand or assumed evidence (107), being than related to evidentiality; if none of these is the source, the only source is the speaker or writer (108).

(105) ***Taro*** *said he wanted to go home soon.* (Matsuyoshi et al., 2010: 1463: ID 1)

(106) ***He*** *said he suffered from symptoms due to stopping steroid medicines at that time*. (Matsuyoshi et al., 2010: 1463: ID 2)

(107) ***I hear*** *that the medication is continued for regulating functions of three semicircular canals.* (Matsuyoshi et al., 2010: 1463: ID 3)

(108) *It is only a matter of time before progress of cloning technology leads to producing artificial organs.* (Matsuyoshi et al., 2010: 1463: ID 4)

- Time, defining the time in which the event occurred in relation with the time when the source took an attitude toward the event mention. It can be 'future', when the factuality of the event is not fixed in nature, meaning that the event has not happened yet (109), or 'notFuture', when the factuality is already fixed, which means that the event already happened (110).

(109) *It is only* ***a matter of time*** *before progress of cloning technology leads to producing artificial organs.* (Matsuyoshi et al., 2010: 1463: ID 4)

(110) *I guess he has been on steroids* ***since a month or two after birth***. (Matsuyoshi et al., 2010: 1463: ID 5)

- Conditional, which determines if the event mention is a proposition with or without a condition. Three values are selected by the authors: 'conditional' for event mentions that exist in a conditional clause (111), 'hasCondition' for event mentions that exist in the main clause of a conditional sentence (112), 'notConditional', when there is no condition (113).

(111) *If it* ***is*** *nice out tomorrow, I will go fishing in that lake.* (Matsuyoshi et al., 2010: 1463: ID 6)

(112) *If it is nice out tomorrow, I will* ***go*** *fishing in that lake.* (Matsuyoshi et al., 2010: 1463: ID 7)

(113) ***Feel*** *yourself the effects of restoration water!* (Matsuyoshi et al., 2010: 1463: ID 8)

- Actuality, the category expressing the degree of certainty toward an event mention in text, being, then, related to epistemic modality. Matsuyoshi et al. (2010) select five labels to define the degree of certainty: 'certain+', 'certain-', 'probable+', 'probable-', 'unknown'; the label 'possible' is excluded, because in Japanese it is not easy to distinguish between 'possible' and 'probable'. They also select some labels to annotate the transition from one state to another: 'certain- → +' indicates transition of a target event mention from uncertain to certain (114), while 'certain+ → certain-' indicates the opposite transition (115).

(114) *So, Taro **began** to use the toothpaste.* (Matsuyoshi et al., 2010: 1463: ID 13)

(115) *Jim decided to **stop** buying the weekly magazine.* (Matsuyoshi et al., 2010: 1463: ID 14)

The difference is that in (114), the action becomes factual, while in (115), the action passes from certain and regular to non-existent (certain+ → certain-). As we can see, this category corresponds to what Saurí et al. (2006 and 2007) and Baker et al. (2010) define as 'factuality' or 'factivity'.

- Evaluation, which indicates how the speaker evaluates a situation or an event mention: it is 'positive', when the source considers it as positive (116); it is 'negative' if the source evaluates it as negative (117), and 'neutral' if there is no source's evaluation (118).

(116) *Taro said he wanted to go home soon.* (Matsuyoshi et al., 2010: 1463: ID 1)

(117) *If I had known he would come to the party, I would not have been there.* (Matsuyoshi et al., 2010: 1463: ID 16)

(118) *It was not for you that he stayed.* (Matsuyoshi et al., 2010: 1463: ID 17)

Between (116) and (117), where the source is mentioned, the difference is that in (116) the event is considered positive by the source (Taro), while in (117) it is considered negative by the source (the speaker). In (118), evaluation is neutral because no source gives any kind of evaluation.

- Focus ("target" in Baker et al. (2010)'s terminology), the category determining the element on which negation, inference or interrogative scope over. In sentence (119), negation scopes over *for you*, while in sentence (120), it scopes over the conjunction

because:

(119) *It is not [for you] that he stayed.* (Matsuyoshi et al., 2010: 1463: ID 17)

(120) *It is not [because] he took the medicine that he recovered.* (Matsuyoshi et al., 2010: 1463: ID 18)

In sentence (121), inference scopes over "a month or two after birth":

(121) *I guess he has been on steroids since [a month or two after birth].* (Matsuyoshi et al., 2010: 1463: ID 5)

In sentence (122), interrogative scopes over "how":

(122) *Then, [how] is xylitol effective?* (Matsuyoshi et al., 2010: 1463: ID 12)

Comparing Matsuyoshi et al. (2010)'s scheme to those of Saurí et al. (2006 and 2007) and of Baker et al. (2010), we can immediately notice that Matsuyoshi et al. (2010) annotate more components than Baker et al. (2010) and Saurí et al. (2006 and 2007). In fact, Matsuyoshi et al. (2010) observe not only restricted modality, factuality and Textual Entailment, but also the temporal relations between the event mention and the moment when the source takes an attitude towards that event mention, as well as they annotate if there is some condition influencing the event. Moreover, Matsuyoshi et al. (2010) choose to observe and annotate all the components involved in the expression of modality (sources, triggers and targets), as they want their scheme to be useful for various NLP application aiming at solving different problems. On the other hand, Saurí et al. (2006 ad 2007), as well as Baker et al. (2010), build their annotation scheme with one specific purpose and therefore the components they annotate are chosen on the base of their utility for the final objective of the annotation scheme. Within the modal values Matsuyoshi et al. (2010) identify, they distinguish between *volition* and *wish*, while neither Saurí et al. (2006 and 2007), nor Baker et al. (2010) do: Saurí et al. (2006 and 2007) group the two within the value *expectation*, while Baker et al. (2010), distinguish between *intention* and *want*, but also introduce the value *requirement* which covers both volitive and deontic modality.

Wiebe et al. (2005)

As Matsuyoshi et al. (2010), Wiebe et al. (2005) explore the annotation of attitudes and opinions in language through a corpus study, as it may be helpful for various NLP

applications. They especially focus on four NLP applications: Information Extraction systems, Question Answering systems, Automatic Summarization and Flame Detection systems.

In their annotation scheme, the authors do not talk about modality but about 'opinions, emotions and feelings', which they join under the more general term *private states*, with which they also identify beliefs, thoughts, goals, evaluations, and judgments. Three types of private states are distinguished:

- explicit mentions of private states (123):

(123) *"The U.S. **fears** a spill-over" said Xirao-Nima.* (Wiebe et al., 2005: 5: 1)

In the example, *fears* is the verb expressing the private state: it is an explicit mention as there is a word directly expressing the private state;

- speech events expressing private states (124):

(124) *"The report is full of absurdities" Xirao-Nima **said.*** (Wiebe et al., 2005: 5: 2)

The term 'speech event' indicates any speaking or writing event: there is usually a specific word, such as *say*, *refer*, *report*, etc., which makes us understand that it is a speech event;

- expressive subjective elements (125):

(125) *"The report is **full of absurdities**" Xirao-Nima said.* (Wiebe et al., 2005: 5: 2)

Expressive subjective elements are those linguistic elements used by the speaker to express his emotional state (frustration, anger, wonder, positive sentiment, mirth, etc.): in sentence (125), the expression *full of absurdities* expresses the personal opinion of the speaker (Xirao-Nima) about the report.

The authors build two types of frames for the annotation of private states, one for explicit mentions of private states and for speech events expressing private states and the other for expressive subjective elements. We show them in Table 4 in the next page.

TABLE 4.

Frame for explicit mentions of private states and for speech events expressing private states:	Frame for expressive subjective elements:
Elements: - text anchor: the linguistic expression for the private state or for the speech event (*trigger* in the terminology used by Baker et al. (2010)); - source: who is expressing the private state or uttering the speech event; - target: the target or topic of the private state. Properties: - intensity: the intensity of the private state can be low, medium, high or extreme; - expression intensity: the intensity of the text anchor can be neutral, low, medium, high or extreme; - insubstantial: this attribute is used to distinguish relevant from irrelevant information.	Elements: - text anchor: the linguistic expression denoting the subjective element (*trigger* in the terminology used by Baker et al. (2010)); - source: who is expressing the private state. Properties: - intensity: the intensity of the expressive subjective element can be low, medium, high or extreme.

As we can see in the table, in the annotation of private states, they first identify the trigger (text anchor), the source and the target. They, then, annotate the intensity of the private state and the intensity of the linguistic expression, to see how the linguistic expression contributes to the overall intensity of the private state. According to the authors, in fact, the annotation of intensity ratings helps in assigning a degree of subjectivity to the expression, making possible to distinguish borderline cases from clear cases of subjectivity or objectivity.

The following sentence is reported to show how they annotate private states (126):

(126) *The US fears a spill-over.* (Wiebe et al., 2005: 10: 12)

Elements:

Text anchor: fears

Source: writer and US

Target: a spill-over

Properties:

Intensity: medium

Expression intensity: medium

As shown in Table 4 in the previous page, the annotation of expressive subjective elements focuses on less components than that of private states (127).

(127) *The report is full of absurdities.* (Wiebe et al., 2005: 11: 13)

Elements:

Text anchor: full of absurdities

Source: writer

Properties:

Intensity: high

As we can see, within the elements annotated, the target is not considered, while, within the properties, only the intensity of the subjective element is considered.

As concerns the annotation of objective speech events, the frame includes the same elements as the frame for private states, except for the properties. The resulting annotation scheme is shown in Table 5 at the beginning of the following page.

In sentence (128), the frame is created for the objective speech event represented by *said*.

(128) *Sargeant O'Leary said the incident took place at 2.00 pm.* (Wiebe et al., 2005: 7: 5)[10]

Elements:

Text anchor: said

Source: Sargeant O'Leary

Target: the incident took place at 2.00 pm

[10] In the original text 'sergeant' is written as we wrote it in the example 'sargeant'.

TABLE 5.

Frame for objective speech events:
Elements: - text anchor: a pointer to the span of text denoting the speech event; - source: the speaker or writer; - target: the target or topic of the speech event.

The annotation scheme of Wiebe et al. (2005) is detailed and rich of information and is not only limited to the annotation of modal values. As Matsuyoshi et al. (2010), Wiebe et al. (2005) identify more components than the ones generally identified in the literature on modality. As we have seen, within them, there are components which do not generally belong to 'restricted modality', but which can be useful for other purposes in Annotation. However, Wiebe et al. (2005) do not present a typology of modality: they do not make a classification of the modal values and of the private states they identify, but just show the components they annotate. As concerns terminology, we have noticed that Wiebe et al. (2005) use the term *anchor* to refer to the linguistic modal expression, the trigger, according to Baker's terminology, while in the work of Saurí et al. (2007), the same term, *anchor*, is used to refer to one of the sources, the author of the sentence.

Szarvas et al. (2008)

As we have seen, Matsuyoshi et al. (2010) decide to annotate even the elements on which negation, inference and interrogative scope over. Szarvas et al. (2008) too report a corpus annotation project based on the identification of the scope ("target" in Baker et al., 2010; "focus" in Matsuyoshi et al., 2010) of negation and uncertainty in biomedical texts. They underline the importance of annotating this information as it may help in avoiding errors of misinterpretation. For example, in the clinical coding of medical reports, the coding of a negative or uncertain disease may result in an over-coding financial penalty. Moreover, in the process of interaction extraction, which aims at determining text evidence for biological entities with certain relations between them, it is important to distinguish a factual relation from an uncertain one. In fact, identifying the target of some negative or speculative word may help in distinguishing factual from non-factual information. The following examples show how the authors annotate the target of some modal expressions:

(129) *(Alectasis in the right mid zone is, however, <possible>).* (Szarvas et al., 2008: 41)

(130) *These findings that (<may> be from an acute pneumonia) include minimal*

bronchiectasis as well. (Szarvas et al., 2008: 41)

(131) *Surprisingly, however, (<none> of this treatment is successful).* (Szarvas et al., 2008: 42)

In the examples, standard brackets are used to mark the target of the modal expression, which appears within angled brackets. In (129), the word *possible* focuses on the whole sentence, meaning that the value of 'possibility' is extended to the whole sentence; in (130), *may* focuses over the subordinate clause, meaning that the value of 'possibility' is only associated to the subordinate clause and not to the main clause; in (131), the negative particle *none* focuses on the following constituents.

2.2.3 General brief overview of the literature on modality in Annotation

Along the second part of this chapter, we have presented some proposals for the annotation of modality and have seen how each work has its own organization and its own list of elements to annotate, based on the objectives of its project. In this section, we will present Table 6, built on the basis of the one in the article of Matsuyoshi et al. (2010), showing the comparison among the annotated components of modality of the different authors we considered along this chapter.

From its analysis, we can see that the more constant components annotated are the trigger and the target of the modal expression. Moreover, along this chapter we have underlined how all the authors centre their attention on the distinction between factual and non-factual values, but only Saurí et al. (2006 and 2007) and Matsuyoshi et al. (2010) annotate the cases of transition of certainty[11]. Also sources are often annotated. We can observe that the less annotated components are temporal and conditional relations, only identified by Matsuyoshi et al. (2010), together with the intensity of the private state, the intensity of the linguistic expression and the attribute "insubstantial" for irrelevant information, which are only annotated by Wiebe et al. (2005). In the end, restricted modality is annotated in all the proposals, except for Wiebe et al. (2005) and Szarvas et al. (2008).

[11] In Saurí et al. (2006 and 2007), the transition of certainty is considered within the value "degrees of possibility", while, in Matsuyoshi et al. (2010), it is considered within the component "actuality".

TABLE 6.

	Trig.	Restr. Mod.	Certainty	Transition of certainty	Pol.	Targ.	Source	Time	Cond.	Priv. St.'s intensity	Ling. Expr.'s intensity	Insubstantial
Saurí et al., 2006 and 2007	*v*	*v*	*v*	*v*	*v*	*v*	*v*					
Baker et al., 2010	*v*	*v*	*v*			*v*						
Matsuyoshi et al., 2010	*v*	*v*	*v*	*v*	*v*	*v*	*v*	*v*	*v*			
Wiebe et al., 2005	*v*		*v*			*v*	*v*			*v*	*v*	*v*
Szarvas et al., 2008	*v*	*v*		*v*		*v*						

2.3 Modality in Annotation and modality in Linguistics

In this final section, our purpose is that of presenting a comparison between the modal values considered in Linguistics and the ones considered in Annotation.

We will point out mainly the differences as they are more salient than the similarities, but this choice is not done to mark a clear distinction between the two approaches. In fact, we want to immediately underline that the approach to the study of modality in Annotation is based on the study of modality in Linguistics. However, as in Annotation the identification of modalities is done with different purposes from those in Linguistics, there are some differences among the modal schemes presented.

First of all, we have noticed how the various schemes of modality created both in Linguistics and in Annotation are generally built on the basis of some conceptual system: in Linguistics, the conceptual system underlying the organization of the modal values generally faces the contrast between necessity and possibility; on the other hand, in Annotation, the system underlying the annotation schemes is based on the contrast between factual and non-factual information, in order to distinguish realized from unrealized events, and subjective from objective information.

Moreover, looking at the systems of modality presented in Linguistics and at the ones

presented in Annotation, there is a difference in the categorization used. Where in the literature on modality in Linguistics there are some big categories, distinguished on the basis of the values identified (epistemic, deontic, etc.), which include smaller categories, distinguished on the basis of the linguistic expressions conveying those modal values (declaratives, directives, etc.), in the literature in Annotation, the values identified are not organized in a hierarchical system. This lack of hierarchy is probably due to the fact that in most of the works done on the annotation of modality, the identification of the modal values is not the main task.

Further on, in the first part of this chapter, we have observed how in some cases some expressions are ambiguous, as the same linguistic expression can be used to express different modal values, as for *may* and *must*, in English, which can be both used to express epistemic or deontic possibility or necessity. However, the analysis of the context can make the expression less ambiguous. Both in Linguistics and Annotation, ambiguity is observed and different examples where the expression appears in context are presented to show all the various meanings. Differences in the approaches arise in the selection of the contexts: in fact, if in Linguistics the contexts are often fabricated, in Annotation, they are taken from examples of real language use.

Another main difference between the two approaches is that, in Linguistics, the analysis of the modal values is limited to the verb and to some other element that, in some cases, contributes to the expression of modality (adverbs, for examples), while, in Annotation, it is common to annotate every element of the discourse, specifying its function, whether modal or not, in the sentence. In fact, as we have seen, in a sentence, the annotators can be instructed to mark not only modal expressions (triggers) but also the sources and the targets of modality, of negation and of speculation.

A different approach to the annotation of modality is that of Wiebe et al. (2005), who, in the annotation of private states, identify attributes which, as we have already said, are not considered in the other proposals, such as the intensity of these 'private states' and the intensity of the linguistic expression, pointing out the importance of the annotation of these elements in domains related to politics and society.

In this section, in more detail, we report a comparison between the modal values identified in Annotation and the ones identified in Linguistics. Table 7 shows in the left column the five broad categories we have identified in Table 2, and in the right column the values identified in the different systems presented in the literature on the annotation of modality. The values shown in the left column include the values shown in the right column.

As we can see, the value *Requirement* is repeated several times; this is because in Baker's annotation scheme this tag is used to annotate expressions which in Linguistics are considered to have a different modal value, such as the ones carrying the values *epistemic*

interrogative, *participant-internal necessity*, *deontic obligation*, and *volition*.

Looking in detail at the relationship between the big categories presented in the left side of the table and at the values included in them in the right side of the table, some points can be made:

1. The big category *Epistemic Modality* includes all the values related to the factuality and certainty of the event identified in the annotation systems, presented in the right column: the values *Factuality*, *Evidentiality* and *Assertion* express the certainty of an event, while the values *Transition of certainty* and *Degrees of possibility* are used to annotate the different degrees of certainty of the event; and the value *Belief* and the values *Requirement* and *Interrogative* express the uncertainty of the event.
2. The big category *Deontic Modality* includes all the values in which there is an imposition of an entity on another; in it, we find the values *Command*, *Requirement* and *Imperative* expressing obligation, and the values *Permissive* and *Permission* expressing permission.
3. *Participant-internal modality* includes two of the values identified in Annotation: the value *Ability* used by Baker et al. (2010) to annotate when the speaker expresses his or someone else's capacities and related to the value Participant-internal modality identified by van der Auwera (1998); and the value *Success*, used by Baker et al. (2010) to annotate the success of the participant in making something happen, also related to van der Auwera's participant-internal modality since the results of the action can depend on a personal capacity or necessity of the participant.
4. The big category *Volition* includes all the values related to personal desires, hopes and wishes of the participant, such as the values *Want*, *Requirement* (Baker et al., 2010) and *Volition* (Matsuyoshi et al., 2010), which are used to convey what the participant wants; the value *Wish* (Matsuyoshi et al., 2010), used to convey personal wishes of the participant; and, the value *Intention*, used to annotate that the participant wants to do something. Moreover, we also include in the category Volition, the value *Effort* (Baker et al., 2010), which expresses the effort the participant does to make something happen, and is related to volition since if the participant makes an effort to make something happen it is because he wants it to happen.
5. In the end, the big category *Evaluation* includes values that are considered only in the systems of modality in Linguistics, since in the annotation systems we have presented, no one annotates the personal evaluation of facts and proposition.

TABLE 7. MODAL VALUES IN ANNOTATION

CATEGORIES OF MODALITY IN LINGUISTICS	MODAL VALUES IN ANNOTATION
EPISTEMIC MODALITY	FACTUALITY (Saurí et al., 2006; Baker et al., 2010)/ACTUALITY (Matsuyoshi et al., 2010) EVIDENTIALITY (Saurí et al., 2006) ASSERTION (Matsuyoshi et al., 2010) TRANSITION OF CERTAINTY (Matsuyoshi e al., 2010)/DEGREES OF POSSIBILITY, (Saurí et al., 2006) BELIEF (Saurí et al., 2006; Baker et al., 2010) REQUIREMENT (Baker et al., 2010)/INTERROGATIVE (Matsuyoshi et al., 2010)
DEONTIC MODALITY	COMMAND (Saurí et al., 2006) REQUIREMENT (Baker et al., 2010) IMPERATIVE (Matsuyoshi et al., 2010) PERMISSIVE (Baker et al., 2010) PERMISSION (Matsuyoshi et al., 2010)
PARTICIPANT-INTERNAL MODALITY	ABILITY (Baker et al., 2010) SUCCESS (Baker et al., 2010) REQUIREMENT (Baker et al., 2010)
VOLITION	EXPECTATION (Saurí et al., 2006 and 2007) WANT (Baker et al., 2010)/VOLITION (Matsuyoshi et al., 2010) REQUIREMENT (Baker et al., 2010) WISH (Matsuyoshi et al., 2010) INTENTION (Baker et al., 2010) ATTEMPTING (Saurí et al., 2006 and 2007)/EFFORT (Baker et al., 2010)
EVALUATION	

In this chapter, we have shown some of the systems of modality presented in the literature in Linguistics for Portuguese and English and some of the systems built for the annotation of modality by various authors for English. As we have seen, in Linguistics there is not a unique modal system: different authors select different values or, in some cases, select the same values but give them different names or organize them in a different way. In Annotation, as well, there are differences among the various systems built for the annotation of modality. Here, the values and components of modality to annotate in corpora change from author to author, but, as we have seen, there are still points in common. This variety in the selection of the modal values to annotate in corpora might depend not only on the fact that the different proposals have different purposes, but also on the fact that the interest for the investigation of modality in Annotation is quite recent. Moreover, even corpora are quite recent means for the study of the language and their annotation at the linguistic level is even

more recent. In European Portuguese, for example, there are still no corpora annotated with modal information; this is the reason why the purpose of this work is that of building the basis for creating an annotation scheme for modality in a European Portuguese corpus.

3. Proposal for the annotation of modality in Portuguese

In the previous chapter, we have presented the existing literature on modality in Linguistics and the literature on the annotation of modality. In this chapter, we present our proposal for the annotation of modality in corpora. The chapter is divided in four main sections: in the first section, we present the data we worked on to test our annotation scheme and the methodology of annotation; in the second section, we introduce the list of modal values we decided to observe and the motivation for choosing them; in the third and fourth sections, we present the other components of our annotation scheme (Polarity, Trigger, Source and Target) and some criteria to annotate them, showing also some difficulties found in their annotation. Finally, we present the scheme we built for the annotation of modality in Portuguese as a whole.

We decided to create a new annotation scheme and avoided the use of the already existing annotation schemes for English, since we wanted to create a scheme specifically built for the annotation of modality in Portuguese, in order to have a tool to be used to observe which are the most expressed modalities and which expressions are used to convey them in Portuguese. In fact, the existing practical annotation schemes described in the previous chapter, as they have been built in order to improve NLP applications, do not present a detailed typology of modal values, but just identify the modal values that are important for the NLP application for which the scheme is created (Baker et al., 2010; Matsuyoshi et al., 2010; Saurí et al., 2006 and 2007; Wiebe et al., 2005)[12].

3.1 Methodology

The creation of our scheme for the annotation of modality followed two main steps.

First, we made a list of the modal values to be used in the annotation of the corpus. This list has been built on the basis of the most frequent modal values presented in the literature on modality in Linguistics but includes also some values presented in some of the works on the annotation of modality. The list has also been modified during the annotation of the corpus on the basis of the problems encountered in the identification of the modal value;

[12] In this chapter, all the examples are taken from our corpus, except for the ones which are signalled with bibliography and for those carrying the note to say that they have been constructed.

the final list of modal values is presented in section 3.2, where we define each modal value.

Second, we selected the components we considered to be directly involved in the expression of modality on the basis of the components presented in the literature on the annotation of modality described in the previous chapter. These are: the *Trigger*, which is the lexical element conveying the modal value; the *Target*, the linguistic expression in the scope of the trigger; the *Source of the event mention*, which is the producer of the event mention; the *Source of the modality*, which is the holder of the modality; the *Modal value*, the component where to tag the modal value expressed; and the *Polarity*, used as a place to mark if there is negation scoping on the modal value.

The annotation of each sentence of the corpus follows two main steps: first, we identify all the triggers in the sentence; and second, we apply the full annotation scheme to each trigger. Fore each trigger we annotate all the components of the scheme in the following order: 1) the Trigger; 2) the Target; 3) the Source of the event mention; 4) the Source of the modality; 5) the Modal value; and 6) the Polarity.

The corpus

In order to check the utility of our annotation scheme, we applied it to a restricted corpus of approximately 2000 sentences, extracted from the written part of the Corpus de Referência do Português Contemporâneo (CRPC)[13], a Portuguese corpus on which the Centro de Linguística da Universidade de Lisboa (CLUL) is working since 1988, collecting various types of written and oral texts belonging to different Portuguese speaking countries, from 1970 onwards, for a total amount of 311 millions of words (Bacelar de Nascimento, 2000; Généreux et al., to appear). The sentences were extracted doing queries for a set of verbs we had previously established in order to check if they carried modal meaning. We only extracted the sentence in which the verb for which we had done the query appeared and avoided extracting further context, as, in this initial phase, we limit our annotation to sentences.

To select the corpus of sentences to test our annotation scheme, we used the Corpus Query Processor (CQPWeb)[14], a software created by Andrew Hardie from Lancaster University, which gives us the possibility to search in a particular subset of the corpus, choosing the country of origin of the text and the type of text. We chose to restrict the geographical origin of the texts to Portugal, as we want, for the moment, to limit our annotation to European Portuguese, and we restricted the text type to the following textual categories: correspondence, pamphlet, newspaper, book (fiction, technical, didactic),

13 http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_crpc.php

14 http://alfclul.clul.ul.pt/CQPweb/

magazines and *varia*, which includes texts belonging to different categories. As our interest for investigating modality was directed especially to common language, we chose to exclude the transcriptions of the Parliament sessions (text type: Politics) and the laws established by the Supreme Court of Justice (text type: Law), because in these two fields the language used is more formal and, being a specific jargon, it is not a sample of everyday's language. We, then, made a selection of modal verbs whose use we wanted to explore and extracted the contexts in which those verbs appeared. In the query, we put the lemma of the verb (e.g. *saber* 'to know') whose contexts we wanted to extract, in order to retrieve all word forms belonging to the lemma (e.g. *sei*, *soube*, *sabe*, *saber*, 'know', etc.). Choosing to extract hits only for verbs was a decision mainly taken in order to restrict the number of hits for the annotation of modality, as the annotation regarded not only modal verbs but also nouns, adverbs and, in some cases, adjectives (e.g. *necessário* 'necessary'; *fundamental* 'fundamental'; *evidente* 'evident', etc.), and entire sentences (e.g. interrogatives). Of course, extracting from the corpus hits only for verbs leads to having a restricted amount of examples where modality will be mostly carried by modal verbs and might miss other linguistic expressions for modality. However, each search query on a particular verb in the CRPC still retrieved a large number of hits, in most cases more than 100,000 hits. We, therefore, used the function "Thin it" to restrict the selection to 5% of the hits and selected the option that arranges the examples in random order, in order to have a list of hits belonging to different texts and topics. We, then, selected only the first 50 hits and extracted the sentence in which the verb was contained.

In the end, we had approximately 2000 sentences to which we could apply our annotation scheme to check if it worked and to observe possible problems.

3.2 Choice of the Modal Values

In this section we present the list of modal values we selected for the annotation of our corpus (Table 8). This selection was based on the observation of the values presented in the different schemas proposed in the literature. We especially follow the modal systems presented by Palmer (1986), Oliveira (1988) and van der Auwera (1998) in Linguistics, but also consider some values presented by some authors in the literature on the annotation of modality (Baker et al., 2010). Our list of modal values includes epistemic and deontic modality, as they are the most presented values in the literature (Palmer, 1986; Oliveira, 1988): epistemic modality is related to the validity of the proposition on the basis of the commitment of the speaker to it, while deontic modality is related to the use of language to make action start.

TABLE 8

Epistemic knowledge Epistemic belief Epistemic doubt Epistemic interrogative Epistemic possibility
Deontic permission Deontic obligation
Participant-internal capacity Participant-internal necessity
Evaluation
Volition
Effort
Success

Moreover, we consider participant-internal modality, as it focuses the attention on the kind of factors influencing the realization of the action, distinguishing especially between needs and capacities of the participant (van der Auwera, 1998); and volitives and evaluatives (Palmer, 1986; Oliveira, 1988), as they focus the attention on the kind of commitment of the participant to the action or situation. Further on, we annotate two other interesting values presented in the literature in Annotation, effort and success, as they are related to the commitment of the participant to the accomplishment of actions (Baker et al., 2010).

As we can see from Table 8, some values, such as epistemic modality, deontic modality, and participant-internal modality include sub-values, while other values, such as effort, success, volition and evaluation, do not.

Within epistemic modality, we identify the following sub-values:

- *epistemic knowledge* to annotate when the speaker presents his or someone else's knowledge (132) or when he expresses some degree of understanding about something (133);

(132) *Em St. Louis, no Missouri, por exemplo, os habitantes* ***sabem*** *de antemão qual vai ser a qualidade do ar no dia seguinte.*
'In St. Louis, in Missouri, for example, the inhabitants know in advance the quality of the air in the following day'.

(133) *Assim que comecei a pensar nisso* ***percebi*** *imediatamente que ia ser muito complicado.*
'As soon as I started thinking about it, I understood that it would be very complicated'.

- *epistemic belief*, when the speaker expresses his or someone else's belief in something (134) or opinion on something;

(134) ***Acreditava*** *nos fantasmas, no sobrenatural e na metempsicose, tecia considerandos sobre os elementos, terra, água ou fogo.*
'He believed in ghosts, in the supernatural and in metempsychosis, and built considerations on the elements, earth, water or fire'.

- *epistemic doubt*, when the speaker doubts something or expresses someone else's doubts (135);

(135) *Quando* ***duvidaram*** *da sua cegueira , os polícias estavam a cumprir aquelas ordens restritas que incluem não deixar entrar guarda-chuvas que não sejam de bolso , obrigam toda a gente a deixar as tampas das garrafas de água - de plástico - no caixote do lixo, a tirar os paus das bandeiras.*
'When they doubted his blindness, the police was executing those restricted orders such as not letting in big umbrellas, obliging everyone to leave the water bottles' plastic caps in the garbage, taking off the sticks from the flags'.

- *epistemic interrogative* for interrogative sentences (136);

(136) ***É uma questão de imagem ou de dinheiro?***
'Does it depend on image or on money?'

- *epistemic possibility*, to annotate when the speaker presents what he or someone else is saying as a possibility or probability (137);

(137) *No sossego do lar, e depois de introduzir uma imagem sua no computador, o cliente* ***pode*** *experimentar e combinar tudo o que lhe*

apetecer.

'In the tranquillity of home, and after having introduced an image in the computer, the client can experiment and combine whatever he wants'.

In our annotation scheme, we avoided the annotation of declaratives as they do not express any type of involvement of the source and because, as Oliveira (1988) says, if we consider declaratives as modal, then all language is modal. We also did not consider evidentials, as they are very similar to declaratives and do not really express modality, but only express the kind of evidence given by the speaker to support what he is saying (138). Together with evidentials, Palmer (1986) presents judgments. With this term he refers to all propositions expressing doubts, hypothesis or some possibility. However, in our annotation, we did not consider this kind of propositions as they are too general and express jointly the values *epistemic doubt*, *epistemic belief* and *epistemic possibility*. Initially, we had also considered the contrast between epistemic possibility and epistemic necessity, but along the annotation, we decided not to consider epistemic necessity as in many cases it was difficult to distinguish if the necessity was epistemic or if it depended either on some condition internal to the participant, being then participant-internal necessity (139), or on some external factor, being then deontic (140).

(138) *Mas **dizem** que ele a queria trazer como recordação.*

'But they say he wanted to bring it/her as a souvenir'.

(139) ***Preciso** de espaço para revelar tudo o que desejo.*

'I need space to reveal all I desire'.

(140) ***Deve** dizer-se, aliás, que a revisão a que o autor submetia o passado pátrio, dramatizando-o de um modo crescente, de meados do séc. XVI até atingir o clímax do declínio no primeiro quartel do séc. XIX, constituía um meio seguro de conquistar um novo público.*

'It has to be said that the constantly dramatizing revision of National history, from the mid XVI century to the apotheosis of the decline in the XIX century, was a safe way for the author to conquer a new public'.

Following Palmer (1986) and Oliveira (1988), we chose to oppose deontic modality to epistemic modality, as deontic modality deals with the relation between language and action, and joins together all those cases in which there is an imposition by an entity on another. We distinguish between the sub-values *deontic obligation*, when the speaker or

participant forces the hearer to do something (141), and *deontic permission*, when the speaker or participant allows the hearer to do something (142).

(141) *Recorda que a imprensa chegou a noticiar a recondução de Esmeraldina Brandão no cargo, após que «o PS da Guarda se agita e* ***obriga*** *a ministra da Saúde a dar o dito por não dito e, em nome da pressão partidária, a violar grosseiramente a lei».*
'He reminds that the press announced the re-election of Esmeraldina Brandão for the position, just after that «the PS of Guarda gets worried and obliges the health minister to unsay something said and to violate the law in the name of the pressure of the party»'.

(142) *Então o director do circo fez-lhe um discurso e disse que se ele prometesse não arranjar sarilhos, o* ***deixaria*** *montar, isto no caso de ele achar que seria capaz de se aguentar em cima de um cavalo.*
'Than, the circus manager talked to him and told him that if he promised not to create problems, he would let him mount the horse, if he thought he was able to stay on the horse'.

In contrast with van der Auwera (1998), who considers deontic modality as a subtype of participant-external modality, in our modal system, we consider participant-external modality within deontic modality, but we decided not to create a separate label for it. Therefore, when an expression conveys participant-external modality we tag it with the labels we created for deontic modality. In fact, observing samples of the real use of Portuguese, in many sentences, it is really difficult to establish if the external factor influencing the participant to act is another person or a superior authority or if it is the situation. In sentence (143), we do not know what kind of external factor imposes the action expressed by *avançar* 'advance', if it is the situation or if it is a person or an institution. To avoid this distinction, we chose to only tag deontic modality, mainly because in the literature on modality the most reported value in opposition to the epistemic one is deontic modality.

(143) *Desse ponto de vista penso que é* ***preciso*** *avançar.*
'From that point of view I believe we must go on'.

At the beginning we had also planned to annotate two more deontic sub-values, deontic suggestion, used to tag when the speaker is giving a suggestion (144), and deontic request, used to tag when the speaker wants the hearer to do something but does not impose himself (145). However, along the annotation, we did not find many expressions conveying

these values as in many cases the expression was ambiguous. In fact, even the examples (144) and (145) are not clear examples the first one of a suggestion and the second of a request but could also be interpreted as obligations.

(144) *Não **deixe** que nada saia do seu controle.*
'Don't let anything get out of your control'.

(145) *Não **deixe** de me escrever e de me dizer o que vai tentando e o que vai conseguindo.*
'Don't stop writing me and telling me what you are up to and what you are achieving'.

Following van der Auwera (1998), we consider participant-internal modality, to annotate if the speaker expresses a personal need (*participant-internal necessity*) (146) or capacity (*participant-internal capacity*) (147), as it focuses the attention on what influences the involvement of the participant towards the realization of the action.

(146) *O meu programa para Atenas estava preparado mais para os 200m e os 400m, mas agora vou **ter de** treinar mais a velocidade.*
'For Atenas I was more prepared for the 200m and 400m, but now I will have to train especially velocity'.

(147) *"Não conseguimos falar com a mãe, que **poderia** explicar melhor o sucedido, porque ela está em estado de choque", sublinhou.*
'"We could not speak to the mother who could have better explained what happened, as she is in shock", he underlined'.

As we have seen in chapter 2, Palmer includes in his system of modality evaluatives and volitives. In our annotation scheme, we consider the two of them as they express some kind of commitment of the participant or speaker towards actions and facts. Evaluatives are considered modal as they express the speaker or participant's evaluation of facts and propositions (148): we, in fact, tag evaluatives with the value *evaluation*.

(148) *Penso que a opção tomada foi **um erro**.*
'I think the choice taken was a mistake'.

Volitives, tagged with the modal value *volition*, on the other hand, are included within

modality as they are related to possible but unrealized facts since they express the speaker's or someone else's wants (149), wishes (150) and hopes (151).

(149) *Xabier Arzalluz, presidente do PNV, disse, em entrevista publicada no sábado em Le Monde, que «desta vez, creio que a ETA* ***quer*** *negociar», acrescentando que «muitos (no interior da ETA) estão convencidos de que a luta actual não tem futuro».*
'Xabier Arzalluz, the president of PNV, said, in an interview published on Saturday in Le Monde, that «this time, I believe that ETA wants to negotiate», also adding that «lots of people (within ETA) believe that the present struggle has no future»'.

(150) *Pires da Fonseca, administrador da CP, responsável pela Unidade de Transporte de Mercadorias e Logística (UTML),* ***espera*** *chegar no final de 2004 ao número recorde de 70 milhões de euros de volume de negócios, atingindo pela primeira vez resultados positivos.*
'Pires da Fonseca, CP's administrator, responsible for the Unidade de Transporte de Mercadorias e Logística (UTML), hopes to get in the end of 2004 to the record number of 70 millions of euros in business, reaching for the first time positive results'.

(151) *Eu* ***gostava*** *que ela tivesse dormido com alguém...*
'I wished she had slept with someone'.

However, our value *volition* covers also two more modal values which are presented by Baker et al. (2010), *intention* (152) and *expectation* (153); these two sub-values are not considered in restricted modality but we can however include them within volition as they express different types of volition and because they are related to possibility and non-factuality.

(152) *São Pedro do Sul seria recolocado no mapa da visita depois de se saber que a Secretaria de Estado da Juventude tinha comprado em hasta pública o velho Hotel das Termas, que* ***tenciona*** *converter em pousada para os jovens.*
'São Pedro do Sul would be included again within the plan of the travel once it was known that the Secretaria do Estado da Juventude had bought in a public auction the old Hotel das Termas, which it intends to turn into a youth hostel'.

(153) *É de* ***esperar*** *que o primeiro passo dado agora pela Microsoft - e que se crê venha a ser anunciado na próxima semana - seja no sentido de pedir ao Tribunal Europeu que suspenda os efeitos de qualquer decisão negativa até ao julgamento de recurso.*

'We expect that the first step now given by Microsoft – and that we believe will be announced next week – will be that of asking to the European Court to suspend the effects of any negative decision till the trial'.

Palmer (1986) also considers commissives, which express the commitment of the speaker or participant to make something happen. The value *commissive* differs from epistemic modality as it expresses commitment towards action while epistemic modality expresses commitment towards speech. However, we decided not to include it in our annotation scheme, since it was not a frequent value and, in many examples, it overlapped with deontic modality: in sentence (154), for example, the verb *prometer* 'to promise' expresses the obligation of the source of the modality to support the fight (*apoiar luta anti-aterro na Figueira da Foz*). In our modal system, commissives are, therefore, annotated with the value *deontic obligation*, as the trigger that could be annotated with the modal value commissive, very often also expresses an obligation given by the speaker or participant to himself.

(154) *Dirigente do Bloco de Esquerda* ***promete*** *apoiar luta anti-aterro na Figueira da Foz.*
'The leader of the Bloco de Esquerda promised to support the fight against landfill utility at Figueira da Foz'.

Following Baker et al. (2010), we decided to observe two more modalities which could be related to volition and to participant-internal modality but which we prefer to consider on their own. These are:

- *effort*, to annotate the commitment and application of the participant to make something happen or to do something; this value is particularly useful to annotate those cases in which from the context we do not know if the action depends or not on a personal necessity or capacity (participant-internal necessity or participant-internal capacity) but lexically it expresses the 'effort' that the participant does to make the action happen (155):

(155) *A Inspecção Geral da Administração Interna também está a* ***tentar*** *perceber o que se passou.*
'The Inspecção Geral da Administração Interna too is trying to understand what happened'.

Effort is also related to volition, since the effort of the participant to make something happen depends on how much the participant wants it to happen; in fact, in (156), the effort the architect does to negotiate the time he wants to spend with the child depends on how strong his desire to stay with the child is. However, the verb *tentar* 'to try' more clearly expresses the value effort and therefore we annotate it with the label *effort*.

(156) *Inconformado com a decisão judicial, o arquitecto **tenta** negociar mais tempo com a criança e acaba por, à custa de algumas cedências, conseguir também jantar com ela às quartas e domingos.*
'Not accepting the court decision, the architect tries to negotiate to spend more time with the child and, giving up on some things, succeeds in also having dinner with him/her on Wednesdays and Sundays'.

- *success*, which expresses whether the results of the commitment of the participant were successful or not; they may be successful or not depending not only on the capacity of the participant (participant-internal capacity) (157), but also on other factors, such as, for example, the effort the person does or the level of commitment of the participant to make the event succeed (158);

(157) *Segue-se então uma segunda fase, muito mais lenta, que corresponde à eliminação progressiva das células infectadas que **conseguem** coexistir com o vírus e que os cientistas pensam ser células imunitárias chamadas macrófagos.*
'There is than a second phase, much slower, which consists of the gradual elimination of infected cells which are able to coexist with the virus and that the scientists believe to be immune cells called macrophages'.

(158) *Trabalhou muito para alcançar a medalha mas não a **conseguiu**.*
'He/she worked hard to reach the medal but he/she did not reach it'.

The value *effort* can be distinguished from the value *success* as it points out that the participant is making some kind of effort to make the event happen, but we do not necessary know which is the result of the effort (155), while the value *success* focuses the attention on the result, marking if the action was accomplished (159) or not (160).

(159) *A HB, que não deixa passar uma única oportunidade para mover os peões da sua*

estratégia, lançou dúvidas sobre as causas desta morte e ***conseguiu*** *que a assembleia municipal de Arrigorriaga, a localidade onde vivia o Etarra, aprovasse uma moção apresentada pelos seus vereadores em que culpava o Governo e a sua política de dispersão dos presos pela trágica ocorrência.*

'HB, who does not let any opportunity pass to move the pawns of her strategy, doubted the causes of this death and obtained that the assembly of Arrigorriaga, the place where Etarra lived, approved a motion presented by his city councillors which blamed the government and its politics of dispersion of prisoners for the tragic happening'.

(160) *É o caso deste estudante, que participa pela terceira vez: «No primeiro ano ficámos antes do meio da tabela, mas o ano passado só por pouco é que não* ***alcançámos*** *a final», acrescenta.*

'It is the case of this student, who is participating for the third time: «The first year we did not reach the middle of the table, but last year we did not reach the final only for few points», he continues'.

The two modal values *effort* and *success* can appear in the same sentence (161)[15]: the value effort is expressed by the verb *esforçar* 'to work' in the main clause, while the value success is conveyed by the verb *conseguir* 'to succeed' in the subordinate clause.

(161) ***Esforçou****-se tanto para aprender a nadar mariposa que em menos de um ano* ***conseguiu****.*

'She worked so much to learn how to swim butterfly stroke that in less than a year she succeeded'.

Ambiguity in the modal value

In the literature on modality in Linguistics, some authors face the problem of the identification of the modal value in expressions that seem to carry more than one modal value. These expressions are generally defined as ambiguous, as they carry different modal values at the same time in the same context. In the literature on the annotation of modality we have explored, on the contrary, the authors do not create a label to annotate ambiguity as the expression is analysed in its original context, which, in most cases, helps in disambiguating its modal meaning. However, along the annotation of our corpus, ambiguity in the modal value was detected, as there were different expressions that could be interpreted as conveying more than one modal value. In order to annotate ambiguous expressions, in our annotation scheme,

[15] Example (161) is a product of our introspection to show how the two modal values *effort* and *success* can co-occur in the sentence.

we use two criteria: 1) we created the component *Ambiguity* as a place to refer that the expression is ambiguous; and 2) in the component *Modal value*, we write all the modal values that can be associated to the expression in that context, listed from the most immediate and natural one to the most peripheral one, as it is shown in example (162), where *exigir* 'to require' expresses mainly *deontic obligation* but can also be interpreted as expressing *volition*.

(162) *PS de Esposende **exige** que presidente da câmara explique retirada de pelouro.*
'Esposende's PS requires that the town hall president explains the withdrawal of function'.
Modal value: deontic_obligation; volition
Ambiguity: yes

Annotating ambiguities between modal values is an important process for the definition of our list of modal values. In fact, if some value is always ambiguous with another, we can decide to eliminate one of the two values. The choice is done based on the frequency of the two modal values alone and on the frequency of the ambiguity between the two modal values. For example, as we have said, we do not consider the value *commissive*, proposed by Palmer (1986), since, after a first annotation of our corpus using also this value, we realized that it was very rare and that in most cases it was ambiguous with the value *deontic obligation*, as shown in the previous example (154).

3.3 Polarity of the modal value

Negation is a phenomenon that interacts with modality, even though it cannot be considered as part of it. In this interaction between negation and modality, the modal meaning of the expression can be drastically modified, since it can turn from positive to negative. In order to signal such cases, we decided to introduce the component *Polarity* in our scheme and attribute two values to it: positive, when the modality conveyed by the trigger is positive (163), and negative, when the modality conveyed by the trigger is negative (164)[16].

(163) *Alain Pinto, o dono do veículo, queria oferecer o animal ao filho Rohan.*
'Alain Pinto, the owner of the car, wanted to offer the animal to his son Rohan'.

(164) *Alain Pinto, o dono do veículo, **não** queria oferecer o animal ao filho Rohan.*
'Alain Pinto, the owner of the car, did not want to offer the animal to his son Rohan'.

[16] Sentence (164) is a product of our introspection and as been built on the base of sentence (163), which is taken from the corpus, in order to show polarity contrast.

The annotation of the polarity of the modal value is important especially for some NLP applications, such as, for example, the Sentiment Analysis application, which needs to know if the public opinion on some specific topic is positive or negative (Wiebe et al., 2005). In sentence (165), for example, the speaker (Sue) is giving a positive evaluation on the election, while in sentence (166) the word *criticized* conveys negative evaluation on the US report's criticism of China's human rights record.

(165) *Sue said: 'The election was **fair**'*. (Wiebe et al., 2005: 9: 8)

(166) *China **criticized** the U.S. report's criticism of China's human rights record.* (Wiebe et al., 2005: 9: 11)

From the annotation of polarity in our corpus, we have observed that negative polarity can be carried by a negative particle focusing on the modality trigger, as for *não*, scoping on *queria* 'wanted' in the previous sentence (164), or it can be carried by the trigger itself, as for *proibir* 'to forbid' in sentence (167)[17].

(167) *O João **proibiu** as filhas de sair à noite com os amigos por terem chegado a casa tarde demais no fim-de-semana anterior.*
'John forbade his daughters to go out at night with friends as they had arrived at home too late in the weekend before'.

Within the triggers we identified, various verbs convey themselves negative polarity even if there is no negative particle or word focusing on them in the sentence. Besides the verbs *impedir* 'to prevent', *proibir* 'to forbid' and synonyms, expressing negative polarity of the modal value *deontic permission*, there are also other verbs, such as, for example, *falir* and *falhar* 'to fail', which express negative polarity of the modal value *success*. In sentence (168), for example, negative polarity is conveyed by *faliu* 'failed', which means 'do not succeed'.

(168) *Na opinião deste académico, os fogos da Serra da Caldeirão vieram demonstrar que o "modelo **faliu**: todos os anos temos mais fogos, cada vez maiores".*
'According to this academician, the fires in the Serra do Caldeirão demonstrated that "the model failed: every year there are more and bigger fires"'.

These verbs that carry negative polarity are very interesting since if they are in the scope of some negative particle, their negative meaning is cancelled by the preceding negative

17 Sentences (167), (168) and (169) have been based on our introspection, in order to show polarity contrast.

particle, and the whole sentence acquires an overall positive polarity. If in sentence (167) the polarity of the modal value is negative, in sentence (169) it is positive, as there is the negative particle *não* scoping on the trigger *proibir* 'to forbid': the negative polarity of the particle *não* cancels the negative polarity of the trigger *proibir* 'to forbid', turning the polarity of the modal value into positive.

(169) *O João **não** proibiu as filhas de sair à noite com os amigos por terem chegado à hora combinada no fim-de-semana anterior.*
'John did not forbid his daughters to go out at night with friends as they had arrived home on time the weekend before'.

The same occurs with the verbs *falir* and *falhar* 'to fail'. Sentence (170) is a new version of sentence (168) but with some changes. In (170), the negative particle *não* focusing on the verb *falir* 'to fail' turns the negative polarity of the modal value *success* into positive. In order to make the sentence coherent, the second clause has been changed.

(170) *Na opinião deste académico, os fogos da Serra da Caldeirão vieram demonstrar que o "modelo **não** faliu: todos os anos temos menos fogos, cada vez menores".*
'According to this academician, the fires in the Serra do Caldeirão demonstrated that "the model did not fail: every year there are less and smaller fires"'.

In both the examples (169) and (170), we annotate the polarity as positive.

3.4 Triggers, Sources and Targets

In the literature on the annotation of modality we have seen that the annotation schemes presented do not only consider modality in the restricted sense, but also analyse other components involved in the expression of modality. The most annotated elements are generally triggers, sources and targets. The *trigger* is the word or string of words carrying modal meaning, the *source* is the producer of the event mention or the subject of the modality and the *target* is the expression in the scope of the trigger. In this section, we will look at each of them in detail.

3.4.1 Triggers

We have already anticipated that the trigger is the lexical expression carrying the modal value. In language, modality can be expressed at different levels, especially at the

lexical level, at the syntactic one and at the morphological one. In this work, we mainly focus on the lexical level.

In the lexicon, there are various kinds of trigger for modality: these are verbs, nouns, adjectives and adverbs.

The most frequent triggers are auxiliary modal verbs, which are verbs that modify the meaning of the main verb in the sentence, giving modal information about it. In Portuguese, the most observed modal verbs are *poder* 'can/may', *dever* 'must/may', and *ter de/que* 'must/have to', which are very much explored in Linguistics, (see especially Oliveira (1988) and Costa Campos (1989) on the uses of the modal verbs *dever* 'must/have to' and *poder* 'can'). However, many other verbs can express modality in Portuguese. Along the annotation, we have observed not only auxiliary modal verbs but also other verbs, which, according to us, carried some modal meaning in the sentence, and have found different triggers for each modal value. For epistemic modality, for example, we annotated all the verbs related to knowledge and understanding, when they conveyed some modal value in the sentence, such as *saber* (171)[18] and *conhecer* 'know', *perceber* 'understand' and synonyms, the ones related to thinking and belief, such as *considerar*, *achar*, *pensar* 'think', *crer, sustentar*, *acreditar* 'believe' (172), the ones related to doubt, such as *duvidar* 'doubt' (173), but also the ones that express possibility or probability, such as, for example, *poder* 'can' (174) and *dever* 'must' (175).

(171) *Em St. Louis, no Missouri, por exemplo, os habitantes* ***sabem*** *de antemão qual vai ser a qualidade do ar no dia seguinte.*

'In St. Louis, in Missouri, for example, the inhabitants know beforehand what the quality of the air will be in the following day'.

(172) *Aliás, Machado* ***acredita*** *que depois de abandonar a reitoria da Universidade, em meados de 98, Alarcão poderá ser muito útil a Coimbra, designadamente na sua projecção no exterior.*

'Indeed, Machado believes that after abandoning the rectorate of the University, in middle 98, Alarcão might be very useful to Coimbra, specifically regarding its projection to the exterior'.

(173) *Quando* ***duvidaram*** *da sua cegueira, os polícias estavam a cumprir aquelas ordens restritas que incluem não deixar entrar guarda-chuvas que não sejam de bolso, obrigar toda a gente a deixar as tampas das garrafas de água - de plástico - no*

[18] In the examples in this section, we only highlight in bold the trigger we mention in the section.

caixote do lixo, a tirar os paus das bandeiras.

'When they doubted his blindness, the police was executing those restricted orders such as not letting in big umbrellas, obliging everyone to leave the water bottles' plastic caps in the garbage, taking off the sticks from the flags'.

(174) *A PRINCÍPIO **podia** pensar-se que a vida amorosa do novo chanceler da Alemanha, Gerhard Schröder, ia prejudicá-lo durante a campanha eleitoral, mas o resultado foi o inverso.*

'At the beginning, one could think that the love life of the new German counsellor Gerhard Schröeder would prejudice him during the electoral campaign, but the result was the opposite'.

(175) *O novo instrumento **deverá** ser assinado em 1999.*

'The new instrument should be signed in 1999'.

For deontic modality, we found the deontic verbs *obrigar* 'oblige', *exigir* 'urge', *permitir* and *conceder* 'allow' and synonyms (176).

(176) *Mais recentemente, as investigações em curso **obrigaram** ao encerramento de algumas salas.*

'More recently, the on-going investigation obliged the closing of some rooms'.

Other triggers for modality are nouns, such as *esperança*, conveying the modal value *volition* in (177), and adjectives, such as *fundamental* 'fundamental' and *essencial* 'essential', conveying the modal value *deontic obligation* in (178). As concerns nominal triggers, if the noun is part of a larger nominal phrase, we tag as trigger only the head noun, as for *esperança* 'hope' in (177)[19].

(177) *O Belenenses venceu a União de Leiria e também tem uma réstia de **esperança** de chegar a um lugar que dê viagens ao estrangeiro na próxima temporada.*

'Belenenses won on the União de Leiria and also has some hope to get to a position that allows to travels abroad in the next season'.

[19] In sentence (177), in this section, we only report on the annotation of the nominal trigger, and do not consider the other modality triggers in the sentence.

(178) *É **fundamental** resolver o problema da sobrelotação, mas, por outro lado, também é **essencial** perceber que tipo de delinquência temos, para saber, por exemplo, se no nosso país se justifica a existência de uma prisão de alta segurança.*
'Solving the overload problem is fundamental, but, on the other side, it is also essential to understand what kind of delinquency we have, in order to know, for example, if our country needs a high-security prison'.

Adjectives are annotated as part of the trigger only when they are part of a verbal phrase and create a compound predicate, as for *fundamental* 'fundamental' and *essencial* 'essential', which are part of the verbal phrases *é fundamental* 'it is fundamental' and *é essencial* 'it is essential' in the previous example (178) and for *necessário* 'necessary' included in the verbal phrase *é necessário* 'it is necessary' in the following example (179). In these cases, however, we do not tag the verb *é* as part of the trigger, but only consider the adjective as trigger since it is the element carrying modal information. Adjectives, in fact, are not annotated as part of the trigger if they are predicates of some entity, as for *o apoio necessário* 'the necessary help' in sentence (180).

(179) *Por tudo isto, é **necessário** (re) pensar a lógica interna do MPT, definindo não só os objectivos como também os métodos a seguir para os alcançar.*
'For all this, it is necessary to think (again) the internal logics of MPT, defining not only the objectives but also the methods to follow in order to reach them'.

(180) *De qualquer forma, garantiu, «estão já preparadas patrulhas de três militares que prestarão o apoio necessário» e avançou com a ideia dessas patrulhas passarem a integrar um elemento da GNR .*
'However, he guaranteed, «patrols composed of three soldiers each are already prepared to give the necessary help» and advanced with the idea that those patrols could integrate a GNR element'.

As concerns adverbial triggers, we tag the complete adverb, including further adverbs scoping on the modal adverb: these adverbs are, in fact, modifiers of the modal adverb and can influence its meaning, as for *pelo menos* 'at least', which scopes over *aparentemente* 'apparently' in sentence (181)[20].

[20] In sentence (181), in this section, we only report on the annotation of the adverbial trigger, and do not consider the other modality triggers in the sentence.

(181) *Para os que não sabem, ele é proprietário de uma editora discográfica e,* ***pelo menos aparentemente****, dedica-se em exclusivo a gravar a música de que gosta e a viajar.*
'For those who do not know, he owns a label and, at least apparently, he dedicates his time to record the music he likes and to travel'.

In some sentences, the adverb is used to emphasize the modal meaning of the verb, and therefore we have to include the adverb in the trigger, as in example (182), where *obrigatoriamente* 'obligatorily' is included in the trigger together with *devem* 'must'.

(182) *Faltas sucessivas aos julgamentos contribuem para a morosidade da justiça e,* ***obrigatoriamente****,* ***devem*** *ser contrariadas.*
'Successive absences from the trials contribute to the justice arrearage and must, obligatorily, be opposed'.

As in our annotation system, we consider as trigger only the element which lexically expresses the modality (verbs, nouns, adjectives, adverbs), many elements are not tagged as trigger; these are: a) negative particles; b) prepositions; and c) articles.

As shown in example (183), negative particles are not included in the trigger as they do not convey modality but polarity.

(183) *Não* ***creio*** *que haja ameaça espanhola.*
'I don't believe there is some Spanish threat'.

Prepositions as well are never included in the trigger as they generally belong to the target (184), except for those cases in which the preposition is part of the verbal phrase, as for the verb *ter de* 'have to' in sentence (185).

(184) *Antes julgo estar* ***ciente*** *[das dificuldades do próximo ano e de que o sector em que trabalhamos será porventura o último a sentir os efeitos da retoma que parece emergir].*
'First of all, I believe I know next year's difficulties and that the area which we work in will probably be the last to feel the effects of the recovery that seems to emerge'.

(185) *Além disso,* ***tenho de*** *[ver como está a concorrência].*
'Furthermore, I have to see how the competition is'.

Articles are generally found before nouns and are not included in the nominal trigger, since we only consider the head noun as trigger, as we can see from example (186), in which the nominal trigger is *desejo* 'desire' and the article *o* 'the' is not tagged.

(186) *Conclusão dos especialistas: o **desejo** de alcançar a igualdade no local de trabalho está a fazer com que as mulheres imitem os homens nos aspectos mais negativos.*
'Conclusion of the specialists: the desire to reach equality in the work place is making women imitate men in the worst aspects'.

In some cases, we identify as trigger the whole sentence. This happens specifically in the annotation of interrogative sentences (187).

(187) ***Quem sabe se não está nela o traço de ligação entre os discos que eu mais ouvia há 50 anos?***
'Who knows if it is not in it/her that lays the link between the records I heard 50 years ago?'

A frequent situation faced during the annotation of modality in the corpus was that in the same sentence there were various different triggers for modality, each one expressing a different modal value. In these cases, we annotated separately each clause containing a trigger for modality (188).

(188) *Walter Veltroni, ministro da Cultura, **acha** a censura um anacronismo, mas até agora não **conseguiu** demover a comissão.*
'Walter Veltroni, the culture minister, thinks that censorship is an anachronism, but till now he could not dissuade the commission'.

In sentence (188), we have two triggers, one in the main clause (*acha* 'thinks'), expressing *epistemic belief*, and the other in the coordinate clause (*conseguiu* 'could'), expressing *success*, and therefore for each trigger we annotate all the components involved in the expression of the modality.

We also annotate discontinuous triggers: in sentence (189), the trigger for the value deontic obligation (*teve de* 'had to') is discontinuous, since there is the adverb *ontem* separating it; to annotate this discontinuity we use an @ between the two parts belonging to the trigger.

(189) *A violência sectária no Punjab, onde a polícia (na foto)* ***teve*** *ontem* ***de*** *intervir para acalmar os protestos, envolve grupos rivais da maioria sunita e da minoria xiita que se responsabilizam mutuamente pelos ataques.*

'The sectarian violence in Punjab, where the police (in the picture) had yesterday to intervene to calm the protests, involves rival groups of the Sunnis majority and the Shiite minority who mutually invest each other with responsibility for the attacks'.

Trigger: teve@de

In the section on targets, we will see how we use the same symbol to mark discontinuity in the target.

Finally, punctuation marks are never included in the trigger, except for three cases: a) in interrogative sentences, where the question mark is tagged as part of the trigger as it marks itself the modality *epistemic interrogative*, as in the previous sentence (187) and in the following example (190); b) in imperatives, where the exclamation mark is considered in the trigger as, even in this case, it helps in marking the fact that it is an imperative, expressing deontic modality (191); and c) in those cases in which the expression tagged as trigger includes punctuation, as for commas (,) parentheses and inverted commas ("), as in (190).

(190) ***E queres agora que vá assim arriscar o meu futuro, o futuro do meu coração, para satisfazer esta loucura?***

'And now you want me to risk my future, the future of my heart, in order to satisfy this madness?'

(191) ***Aguenta****-te****!***

'Resist!'

Difficult cases in the annotation of triggers

As concerns triggers, doubts in their identification arose especially in those cases in which there were two modal verbs one next to the other, both carrying modal meaning. Oliveira in Mira Mateus et al. (2003) explains that, when two modal verbs appear one next to the other, their interpretation is different from when they appear alone. As concerns the verbs *poder* 'can' and *dever* 'must/shall', Oliveira (2003) says that when they co-occur in the sentence, the first generally has an epistemic value. For example, in sentences (192), (193) and (194), the first modal verb (*deve* 'must' in (192) and (194); *pode* 'may' in (193)) affects

the certainty of the following modal verb, conveying a meaning of possibility or probability to it. On the contrary, when the verb *ter de* 'must/have to' precedes another modal verb, it keeps its deontic meaning, marking what follows as something necessary, as in sentence (195), where *tem de* 'must' marks as necessary the possibility expressed by *poder* 'be able to'.

(192) *Ele **deve** **[ter de** [chegar amanhã]].*(Oliveira in Mira Mateus et al. (2003): 248: 16)
'He must have to arrive tomorrow'.

(193) *Ele **pode** **[ter de** [chegar amanhã]].* (Oliveira in Mira Mateus et al. (2003): 247: 15)
'He may have to arrive tomorrow'.

(194) *Ele **deve** **[poder** [chegar amanhã]].* (Oliveira in Mira Mateus et al. (2003): 247: 14)
'He may be able to arrive tomorrow'.

(195) *Ele **tem de** **[poder** [chegar amanhã]].* (Oliveira in Mira Mateus et al. (2003): 248: 17)
'He must be able to arrive tomorrow'.

Cases of co-occurrence between different modal verbs were very frequent in our corpus. Along the annotation, we have observed that generally the first modal verb gives information on the degree of certainty of the second modal verb. For example, in sentence (196), the verb *acreditar* 'believe' restricts the degree of certainty of the modal verb in the subordinate clause marking it as a belief of people. In sentence (197), the value *deontic obligation* conveyed by *ter de* 'must' marks as necessary the value *epistemic knowledge* carried by *conhecer* 'know': the sentence expresses the obligation of the voted people to know what the public dislikes and the problems to solve.

(196) *A maior parte das pessoas **acreditou** que a palavra **deve** vir primeiro.*
'Most people believed that the word should come first'.

(197) *Quem tem de dirigir politicamente a Câmara são os eleitos que **têm de** **[conhecer** [os descontentamentos concretos ou os pontos de bloqueio]] para poderem intervir e procurar resolver com autoridade e em tempo útil os problemas existentes.*
'The ones that must politically lead the Town-Hall are the elected ones, who must know the concrete complaints and the blocking issues in order to intervene and solve with authority and in the right time the existing problems'.

The full annotation scheme, in fact, is applied to both the verbs.

3.4.2 Sources

In some of the works written on the annotation of modality, the authors underline the importance of identifying sources. Generally, in annotation, the authors consider as source the experiencer or cognizer of the modality (Saurí et al., 2006 and 2007, Matsuyoshi et al., 2010). In sentence (198), the source is clearly the producer of the sentence, therefore the speaker or writer of the sentence, which is also the source of the modality, as he speaks in the first person.

(198) *Precisava de consultar um advogado, mas não conheço nenhum pessoalmente.*
'I needed to see a lawyer, but I know none personally'.

However, there are cases in which the producer of the sentence is not the same as the experiencer or cognizer of the modality. In sentence (199), for example, the producer of the event mention is the speaker or writer and the experiencers of the modality conveyed by *necessitam* 'need' are *os portugueses* 'Portuguese people'[21].

(199) ***Os portugueses*** *necessitam, em média, de 180 contos por mês para a manutenção de uma família de quatro pessoas.*
'Portuguese people need, on average, 180 contos per month to support a family of four people'.

In order to report this distinction, we have created two tags for the source in our annotation scheme. These are:

- the *Source of the event mention*, to tag the producer of the event mention containing the modality trigger, generally the speaker or writer;
- the *Source of the modality*, to tag the experiencer or cognizer of the modality, which is prototypically expressed by a nominal phrase;

In sentence (198), therefore, the source of the event mention and the source of the modality are the same (the speaker or writer), while, in sentence (199), the source of the event mention is the speaker or writer and the source of the modality are Portuguese people (*os portugueses*).

[21] In this section, sources are highlighted in bold.

The identification of the source of the modality implies the definition of the linguistic elements included in it and of the ones left apart. We tag as source all the elements which help in defining the semantic domain of the source. These are: a) nouns; b) articles; c) adjectives; d) prepositional phrases; and e) restrictive clauses.

When the source of the modality is a simple nominal phrase, composed of a noun or of a noun and an article, we tag both the noun and the article as *Source of the modality*, as for *O ministério* 'the Ministry' in (200).

(200) ***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.*
'The Ministry, announcing that it is going to take off the label superior to the nine institutes, and offering them the possibility to give graduation through the polytechnic school, tries to restore the law, minimizing the damages to the students who chose those private institutions'.

When there are adjectives scoping on the noun carrying the function of Source of the modality, we tag in the component *Source of the modality* also the adjective, as for *nenhuma* 'no' and *internacional* 'international' in (201).

(201) *O governante português afirmara que **nenhuma força internacional** será capaz de fazer aplicar os acordos se as partes não os cumprirem.*
'The Portuguese minister had said that no international force would be able of applying the agreements if the parts do not fulfil them'.

The same occurs if there are prepositional phrases restricting the domain of the source of the modality, as in the case of *do clube alemão* 'of the German team' in sentence (202).

(202) *Enquanto isso, **o médico do clube alemão** referiu que o jogador precisa de descansar.*
'In the meantime, the doctor of the German team reported that the player needs to rest'.

If the noun in the source of the modality is restricted by a restrictive clause, the

restrictive clause too falls within the source of the modality together with the noun, as it is important information defining the semantic domain of the noun. In (203), for example, the clause *que têm sido utilizadas no estudo do cérebro* 'used to study the brain' defines a specific type of approach, meaning that not all approaches need a common base, but only the ones used to study the brain, and therefore the restrictive clause is tagged within the source of the modality together with the nominal phrase *as diferentes abordagens* 'the different approaches'.

(203) ***As diferentes abordagens que têm sido utilizadas no estudo do cérebro*** *precisam de um denominador comum, declarou à revista britânica new scientist guy mckhann*[22]*, da universidade johns hopkins.*

'The different approaches used to study the brain must have a common base, declared guy mckhann from the university johns hopkins to the British journal new scientist'.

Difficult cases in the annotation of sources

Doubts about the annotation of sources emerged in seven main cases: a) when the source of the modality is not expressed; b) in impersonal constructions; c) with pronouns which are not subjects; d) when the source is a non animated entity and the trigger requires an animated source; e) in passive sentences; f) in interrogative sentences; and g) in imperative sentences.

Being Portuguese a null subject language, in many sentences the source of the modality is not expressed: we, then, annotate as source of the event mention the speaker or writer and as source of the modality the verb itself as it carries the marks of the subject (person and number), as in example (204), where *sabemos* 'we know' is annotated both as trigger and source of the modality.

(204) *Quanto a este fragmento* ***sabemos*** *que quem o profere é Tiestes, e que o faz em diálogo com Atreu, pois disso nos informa o escólio à Retórica de Aristóteles que o regista.*

'About this fragment we know that it is said by Tiestes in a dialogue with Atreu, as reported by the scholium to the Rhetoric of Aristotle'.

Through this criterion of annotation we are also able to maintain the ambiguity of the source of the modality when it is conveyed by verbal forms that can be associated to more

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

subjects, as for *sabia* 'knew' in sentence (205), which can be associated to the speaker or writer or to another person whom the speaker or writer is talking about.

(205) *Já **sabia** que o senhor Barbosa Ribeiro [presidente da concelhia do PS-Gaia] era politicamente uma nulidade, uma pessoa desequilibrada.*
'I/he/she already knew that Mister Barbosa Ribeiro [the president of the PS of Gaia] was politically a nullity, an unbalanced person'.

As concerns impersonal constructions with the element *se*, we decided to annotate the speaker or writer as source of the event mention and the clitic *se* as source of the modality (206), since it carries the impersonal reading. Consequently, *se* is not considered within the trigger.

(206) *Sabe-**se**, entretanto, que Miranda apresentou a sua demissão à Caixa de Crédito Agrícola Mútuo, em Ovar, onde ocupava o cargo de subgerente, a 8 de Julho, um dia antes de partir de férias e da apresentação da equipa de futebol, onde já não esteve presente.*
'Meanwhile, it known that Miranda resigned from the Caixa de Crédito Agrícola Mútuo, in Ovar, where he held the position of sub-chief, on the 8th of July, one day before leaving for his holidays and before the presentation of the football team, to which he was not present'.

In the third case, when the source of the modality is not expressed by the subject of the modal verb but it is expressed by an indirect object, we tag as *Source of the modality* the indirect object as it is the only element in the sentence referring to the holder of the modality: in sentence (207), we tag the indirect complement *me* as *Source of the modality*.

(207) *Apetecia-**me**, neste caso, trabalhar com actores, não como tipos mas como personagens.*
'I felt like working with actors, not as types but as characters'.

In the fourth case, the source of the modality is expressed by an inanimate subject, even though the modal verb typically requires an animated agentive subject. In (208), for example, we consider as source of the modality *o tecido*, 'the tissue', which is presented as the subject and agent of the verb *obrigar* 'have to', and the modal value expressed is *participant-internal necessity*.

(208) ***O tecido*** *para algumas destas peças - o burel - obrigava à sua passagem pelo pisão para, numa última operação, lhe ser garantida a consistência e durabilidade que, de outro modo, não possuiria.*

'The tissue of some of these pieces – the burel – had to be passed in the fulling mill so that, in the last operation, he would have the consistency and durability that he wouldn't have in other ways'.

Another example is shown in sentence (209), where the subject of the verb *permitir* is *o computador* 'the computer'. Here, we annotate *me* as source of the event mention as it represents the speaker or writer and *o computador* as source of the modality.

(209) *Agora, embora não seja capaz de pintar porque não tenho técnica para o fazer, descobri que* ***o computador*** *me permite transformar as minhas imagens de tal maneira que ficam a parecer autênticas pinturas.*

'Now, although I cannot paint as I don't have the technical capacity to do it, I discovered that the computer allows me to transform my images so that they look like real paintings'.

In passive sentences, the source of the modality does not correspond to the grammatical subject of the modal verb but is the agent and is, therefore, expressed by a complement. We, then, annotate the real semantic agent of the modal verb in the passive form as the source of the modality and the grammatical subject as the target of the modality: in (210), the source of the event mention is *o acordão* 'the judgment' and the source of the modality expressed by *foram consideradas* 'considered' is the court (*Tribunal de Execução de Penas*).

(210) *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".*

'The judgment reminds that "in the official communication presented, signed by Ricardo Cardoso, it was not possible, nor is possible now, to give relevance to the same circumstances used in another process to which the offender was condemned and considered existent by the Tribunal de Execução de Penas, in order to give conditioned freedom to the offender"'.

However, in some passives the agent is not expressed. In these cases, we annotate as source of the event mention the producer of the event mention and as source of the modality the verbal trigger: in (211), in fact, the source of the event mention is the speaker or writer, and, as the source of the modality is not expressed, we tag in it the verb *foram obrigados* 'were forced'.

(211) *Não se poderá responsabilizar os portugueses que por lá passaram e viveram, nem os macaenses, filhos da terra, que ficaram ou* ***foram obrigados*** *a imigrar, nem os chineses, porque esses investem ainda tudo o que têm no futuro.*
'It won't be possible to invest with responsibility the Portuguese people who passed and lived there, nor the people from Macau, sons of the land, who stayed or were forced to immigrate, nor Chinese people, because they invest all they got in the future'.

In 3.4.1, we have reported on the identification of the trigger in interrogatives and have said that we consider the whole sentence as trigger. The question now is what to tag as source of the modality. Observing the interrogative (212), we can see that there is no source lexically expressed for the modal value *epistemic interrogative* conveyed by the whole sentence; therefore, the only source is the speaker or writer, which is both the source of the event mention and the source of the modality.

(212) *Qual será o mecanismo que permite transferir directamente a energia do interior da estrela para a sua coroa externa?*
'Which will be the mechanism allowing to transfer directly the energy from the interior of the star to its external crown?'

In imperatives as well, doubts arise as concerns the source of the modality. However, we adopt the same criteria as in interrogatives and tag as source of the modality the speaker or writer, which, here again, is also the source of the event mention, as in (213).

(213) *Esperem lá!*
'Wait!'

In the end, in the case of finding discontinuous sources we use the same symbol as for discontinuous triggers, the @, in order to mark their discontinuity.

3.4.3 Targets

At the beginning of section 3.4, we have defined as target the expression in the scope of the modal element. While Szarvas et al. (2008) include within the target[23] the biggest syntactic unit possible in the scope of the trigger, we decided to limit the target to the complements of the verb.

In general, identifying the target and its boundaries is easy, as, in most cases, the target of the modal expression is located in the right periphery of the trigger, as in the left one there is the source: in sentence (214), the target is the expression *a reabertura da Rua do Comércio* 'the re-opening of Rua do Comércio' and, in (215), the target is *criar nem confiança nem credibilidade* 'to create neither trust nor credibility'. When the target is nominal (214), we consider as target the whole nominal phrase including prepositional phrases or sub-clauses; when the target is an entire clause, we tag as target the whole clause (215), including the conjunctions and the prepositions introducing subordinate clauses (216).

(214) *Comerciantes da Guarda exigem [a reabertura da Rua do Comércio].*
'Merchants in Guarda urge the re-opening of Rua do Comércio'.

(215) *O Governo não soube [criar nem confiança nem credibilidade].*
'The Government was not able to create neither trust nor credibility'.

(216) *Ele acha [que a Igreja fica favorecida pelo facto de estar aliada ao poder político].*
'He thinks the Church is advantaged because of the fact that she is allied with the political power'.

In some cases, the target is placed before the trigger, as in sentence (217), where the target is the subordinate clause *onde arranja dinheiro* 'where he finds money'.

(217) *[Onde arranja o dinheiro], não sei, nem interessa.*
'I don't know where he finds money, neither I care'.

There are also cases in which the target is split in two parts that are placed in the sentence in two separate positions, as in (218), where the target is composed of *lhe* 'him' and *se condenava as relações extra-conjugais* 'if he condemned extramarital relationships'. In this case, both parts are annotated in the target as they are both complements of the trigger and to mark the fact that they are discontinuous but syntactically linked by concordance we use the

[23] In this section, targets are marked within square brackets.

sign @ in the annotation, as we do for discontinuous triggers and sources.

(218) *Logo a seguir, ensaiou o primeiro «drible», quando [lhe] perguntaram [se condenava as relações extra-conjugais].*
'Immediately after, he tested the first dribble, when they asked him if he condemned extramarital relationships'.
Trigger: perguntaram
Target: lhe@se condenava as relações extra-coniugais

Finally, 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 in (219), in which we tag as target only *sensações tácteis parecidas* 'similar tactile sensations', and not *os livros de papel* 'paper books', as they are semantically but not syntactically linked to each other; in (219), the complement of *parecidas* 'similar' is inferred and would be the prepositional phrase *com as dos livros de papel* 'to those of paper books' as it is shown in sentence (220)[24].

(219) *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.*
'They are similar to paper books: they want to give similar tactile sensations and give to the reader something more – the possibility to have a library and a moving databank, weighting a bit more than a kilogram'.
Trigger: permitir
Target: sensações tácteis parecidas

(220) *Parecem-se com os livros de papel, querem permitir [sensações tácteis parecidas com as dos livros de papel].*
'They are similar to paper books, as they want to give tactile sensations similar to those of the paper books'.
Trigger: permitir
Target: sensações tácteis parecidas com as dos livros de papel

Difficult cases in the annotation of targets

Till now, we have presented cases in which the target was easily identified. However,

[24] Example (220) is a product of our introspection to show how the semantic link between *parecidas* and *livros the papel* should be syntactically expressed.

the identification of targets is not always immediate. We especially had problems in the annotation of those clauses in which the trigger had no target, and in identifying the target of interrogative and imperative sentences.

In those sentences in which the target is not lexically expressed, we do not annotate it. In (221), the target of the triggers *proibissem* 'would forbid' and *proibir* 'forbid' is something that has been expressed in a previous sentence, and could only be identified extracting a bigger context. We, therefore, do not annotate it.

(221) *Quando trespassei a minha casa calculei que, mais cedo ou mais tarde, o nosso ramo ia falir e não era lá porque os governos proibissem ou deixassem de proibir.*
'As I transferred my house I calculated that, sooner or later, our branch would fail and it was not because the governments would forbid or stop to forbid'.

In interrogative sentences, the target is the core sentence, the doubted part, the part about which the speaker is asking information or about which he is expressing some doubt. In many cases, in order to extract only the target from the sentence, we should have modified the whole sentence. Since in the annotation we do not want to modify anything in the sentence, we decided to mark as target the whole sentence, including the question mark. In this way, we show that the value epistemic interrogative applies to the whole sentence. In sentence (222), for example, the value *epistemic interrogative* applies to the expression *Ensaiou as ampolas* 'Did he test the ampoules', as well as in sentence (223), it applies to the whole expression *Sete vezes sete* 'Seven times seven'. In fact, it can be observed that, as concerns interrogatives, in most cases, we annotate the same elements both as triggers and as targets.

(222) *[Ensaiou as ampolas?]*
'Did he test the ampoules?'
Trigger: Ensaiou as ampolas?
Target: Ensaiou as ampolas?

(223) *[Sete vezes sete?]*
'Seven times seven?'
Trigger: Sete vezes sete?
Target: Sete vezes sete?

Also in the annotation of sentence (224)[25], we annotate as target the whole question.

[25] Sentence (224) is the result of our introspection on the topic.

(224) *[Quanto é sete vezes sete?]*
'How much is it seven times seven?'

Interrogatives are also annotated when indirect. Even in this case, the target is the core sentence, but in indirect interrogatives we do not need to annotate the whole sentence as target as generally the target is the constituent in the scope of the lexical question trigger (225).

(225) *E assim chegamos ao cerne da questão, que é saber [quem nomeou Manuel dos Santos].*
'So we got to the core of the problem, which is to know who nominated Manuel dos Santos'.

In sentence (225), the trigger for the interrogative is *saber* 'to know' and the target is what is contained in its scope (*quem nomeou Manuel dos Santos* 'who nominated Manuel dos Santos').

In imperatives, the target is the person to whom the order is directed, the hearer. Generally the hearer is not lexically expressed (226). We, then, tag as target the verb in the imperative form (*esperem* 'wait' in (226)), as it contains the marks of the person to whom the imperative is directed.

(226) *[Esperem] lá!*
'Wait!'

However, in those imperative sentences in which the target is lexically expressed, we mark as target not the verb but the real target, as for *Ó Virginazinha* 'Virginazinha' in sentence (227) and for *Ó seu árbitro* 'Referee' in sentence (228):

(227) *[Ó Virginazinha]*, *deixa o teu marido!*
'Virginazinha, leave your husband!'

(228) *[Ó seu árbitro], ponha esse malandro fora do campo!*
'Referee, kick that villain out!'

The final problem we found in the annotation of targets regards such verbs as *ter*

de/que ‘must’ and *obrigar* ‘to oblige’, and other verbs as *precisar* and *necessitar* ‘to need’. In some sentences these verbs have two interpretations and the annotation changes on the base of the interpretation. The first two verbs can convey participant-internal necessity or deontic obligation in sentences as (229), as well as the second two verbs can be associated to the same two values in sentences as (230).

(229) *O Estado tem de perceber que já não há um modelo único cultural.*
‘The State must understand that there is no more a unique cultural model’.

(230) *Corretja também precisou de eliminar compatriotas.*
‘Corretja too had to eliminate compatriots’.

In sentence (229), if we consider as source of the modality the nominal phrase *O Estado* ‘The State’, the target is *perceber que já não há um modelo único cultural* ‘understand that there is no more a unique cultural model’, and the modal value is participant-internal necessity; however, we can also consider the two expressions *O Estado* ‘The State’ and *perceber que já não há um modelo único cultural* ‘understand that there is no more a unique cultural model’ all as one discontinuous target, being the speaker or writer the only source.

The two annotations would be as follows:

Trigger: tem de Target: perceber que já não há um modelo único cultural Source of the event mention: sp/wr Source of the modality: O Estado Modal value: participant-internal_necessity Polarity: positive	Trigger: tem de Target: O Estado@perceber que já não há um modelo único cultural Source of the event mention: sp/wr Source of the modality: sp/wr Modal value: deontic_obligation Polarity: positive

In sentence (230), we can consider as source of the modality the proper noun *Corretja* and the expression *de eliminar compatriotas* ‘eliminate compatriots’ as target, in which case the modal value is participant-internal necessity, or we can annotate the speaker or writer both as source of the event mention and as source of the modality, being the target composed of the two elements *Corretja* and *de eliminar compatriotas* ‘eliminate compatriots’.

The two annotations would be as follows:

Trigger: precisou	Trigger: precisou
Target: de eliminar compatriotas	Target: Corretja@de eliminar compatriotas
Source of the event mention: sp/wr	Source of the event mention: sp/wr
Source of the modality: Corretja	Source of the modality: sp/wr
Modal value: participant-internal_necessity	Modal value: deontic_obligation
Polarity: positive	Polarity: positive

In the annotation of this type of sentences, we observed the sentence and chose one interpretation according to the context. In (231), for example, the speaker or writer presents the necessity to understand the functioning of Portuguese public life as a personal necessity; we, then, annotate the sentence with the modal value participant-internal necessity, tagging the verb *temos de* 'must' as source of the modality, since it is the only element in the sentence carrying the subject information. In (232), the source presents the payment of the bill as an obligation imposed by rules or laws, and therefore we annotate it with the modal value deontic obligation, tagging the speaker or writer as source of the modality.

(231) *Para perceber toda esta situação temos de compreender o funcionamento da vida pública portuguesa*.

'In order to understand the whole situation we must comprehend the functioning of Portuguese public life'.

Trigger: temos de

Target: compreender o funcionamento da vida pública portuguesa

Source of the event mention: sp/wr

Source of the modality: temos de

Modal value: participant-internal_necessity

Polarity: positive

(232) *Caso seja outra vez considerado culpado, terá de pagar uma multa no valor máximo de 500 dólares (390 euros), ou 30 dias na cadeia.*

'If he is considered guilty once again, he must pay a bill for a maximum of 500 dollars (390 euros), or he must spend 30 days in jail'.

Trigger: terá de

Target: culpado@pagar uma multa no valor máximo de 500 dólares (390) euros), ou 30 dias na cadeia

Source of the event mention: sp/wr
Source of the modality: sp/wr
Modal value: deontic_obligation
Polarity: positive

As we have seen in section 3.4, the annotation of modality and of its components is not always easy; problems arise not only as concerns the identification of the modal value expressed by the trigger in the sentence, but also as concerns the identification of each component and of its boundaries, and of the relationships between one component and another when observed and annotated in real text.

3.5 Final annotation scheme

Till now we have presented the components we wanted to identify through our annotation scheme, observing in detail the function of each of them. In the end of the chapter, we present a table (Table 9) where all the components of our final annotation scheme are shown.

The final annotation scheme includes also the components *Ambiguity* and *Comment*, which are not always annotated. The first, as we have seen, is used only in the annotation of ambiguous expressions; the second is used to annotate any comment about the annotation of the sentence, such as, for example, the problems we had in the annotation of a sentence or observations about the identification of the modal value or of the other components.

We also use two graphical criteria for the annotation of the components and of the modal values of the schema: a) for the components of the schema, we use capital letters, while for the modal values tagged in the component *Modal value* we don't; b) in the annotation of the modal value, we use an underscore (_) to mark the relation between value and sub-value, as shown for the value expressed by the trigger of the subordinate clause in the following example (233).

Example (233) shows the application of our annotation scheme to a sentence extracted from the CRPC:

(233) *Todos querem saber o que é que acontece entre Oliveira e Marcello Mastroianni, no Norte de Portugal, em "Viagem ao Princípio do Mundo", que neste momento está a ser rodado.*
'Everyone wants to know what happens between Oliveira and Marcello Mastroianni, in the North of Portugal, in "Voyage to the Beginning of the World", currently being

filmed'.

Trigger: querem
Target: saber o que é que acontece entre Oliveira e Marcello Mastroianni, no Norte de Portugal, em "Viagem ao Princípio do Mundo", que neste momento está a ser rodado
Source of the event mention: sp/wr
Source of the modality: Todos
Modal value: volition
Polarity: positive
Trigger: saber
Target: o que é que acontece entre Oliveira e Marcello Mastroianni, no Norte de Portugal, em "Viagem ao Princípio do Mundo", que neste momento está a ser rodado
Source of the event mention: sp/wr
Source of the modality: Todos
Modal value: epistemic_knowledge
Polarity: positive

As we have said in the section on triggers in this chapter, we consider *querem* 'want' and *saber* 'to know' as two separate triggers belonging to two separate clauses, as they express two different modal values; we, in fact, apply the full annotation scheme to each trigger.

While in (233) we only annotate the main components of modality, the annotation of (234) shows an example of how we annotate the component *Comment*.

(234) *No caso do aborto eugénico, ou seja, havendo má formação do feto, ou motivos para crer que este venha a sofrer de doença grave e incurável, a interrupção da gravidez pode acontecer até cerca dos cinco meses e meio, em vez do limite actual de quatro meses (dezasseis semanas).*
'In the case of an eugenic abortion, i.e. when there is fetus' malformation, or reasons to believe that it might be affected by some serious and untreatable illness, the abortion can be done till the middle of the sixth month, instead of usual limit of fourth months (sixteen weeks)'.

Trigger: crer
Target: que este venha a sofrer de doença grave e incurável
Source of the event mention: sp/wr

Source of the modality: sp/wr

Modal value: epistemic_belief

Polarity: positive

Trigger: pode

Target: a interrupção de gravidez@acontecer até cerca dos cinco meses e meio

Source of the event mention: sp/wr

Source of the modality: sp/wr

Modal value: deontic_permission

Polarity: positive

Comment: the modal value is deontic_permission as the amount of time to do abortion is imposed by biological conditions

As we can see, our annotation scheme aims to identify only the elements which are directly involved in the expression of modality, such as the modal value and its polarity, the lexical element conveying the modal value (Trigger), and the other elements affected by the modality (Source of the event mention; Source of the modality; and Target).

In fact, we do not annotate many of the components introduced in the works on the annotation of modality. In the section on the literature on the annotation of modality, we have observed how most of the authors annotate factuality. In their work, the annotation of factuality is necessary as the applications that will use annotated text generally need to immediately identify what is presented as certain from what is presented as uncertain. On the contrary, in our annotation scheme, factuality is not considered as we are not concerned with the distinction between realized and unrealized events, but with the way how the speaker shows his commitment to speech and action.

We also do not consider two more components presented in the annotation scheme of Matsuyoshi et al. (2010): the component *Time*, which is used to annotate temporal relations between event mentions, and the component *Conditional*, which is used to annotate if there is some condition for the realization of the event; in our annotation scheme, temporal relations are not annotated, while conditions are annotate under the modal values *epistemic possibility* or *epistemic doubt*.

In the next chapter, we will observe the results of the annotation of our corpus of 1935 sentences.

TABLE 9. COMPONENTS OF THE ANNOTATION SCHEME

Trigger
Target
Source of the event mention
Source of the modality
Modal value
Ambiguity
Polarity
Comment

4. Results

In the previous chapter we have explained how we worked to annotate modality in our corpus of sentences, showing the list of modal values used for the annotation and explaining how we annotate the different components involved in the expression of modality in a sentence. In this chapter, we show the results of the use of our scheme for the annotation of modality in a corpus. This annotation provides us with insights in the phenomenon of modality in the current use of Portuguese.

We have two main objectives: the first is that of analysing which are the most expressed modal values in our corpus, and to understand if our list of modal values is detailed enough or if it is too much detailed for the annotation of modality in real text; the second is to discuss in detail the difficult cases and ambiguities encountered in the corpus.

We can already anticipate that, looking only at modal values and not considering their division in sub-values, the most expressed modal values in our corpus were epistemic and deontic modality, immediately followed by participant-internal modality and volition, while the less expressed values were effort and success. Moreover, in the whole corpus, we found 2377 triggers for modality.

4.1 Implications of the corpus on the results

As we have explained in chapter 3, we have applied the annotation scheme to a sub-corpus composed of 1935 sentences extracted doing queries for specific modal verbs in the

Corpus de Referência do Português Contemporâneo (CRPC) in a set of European Portuguese texts belonging to determined categories. The list of verbs, shown in Table 10 in the following page was made choosing 4 or 5 verbs, which could be associated to each type of modality; for example, the verbs *saber* 'know', *pensar* 'think', *crer* 'believe', *perceber* 'understand' and *julgar* 'judge' generally have an epistemic meaning so they were chosen to see if they really represent epistemic modality, as well as the verbs *permitir* 'allow', *obrigar* 'oblige', *exigir* 'require', *conceder* 'allow', *deixar* 'allow' were chosen to see if they always carry a deontic value or may also carry other values. The query was done for the lemma of the verb, in order not to have only sentences in which the verb was in the infinitive but also sentences in which the verb was conjugated. As, in most cases, the resulting number of hits extracted was very high, we applied the option of the software *CQP web* that allows us to select only 5% of the contexts and to order them randomly, so that we would have hits from different texts belonging to different categories. We, then, selected the first 50 hits and extracted only the sentence in which the verb was included. This method of selection of sentences for our corpus has some implications. First of all, deciding to extract contexts for a specific set of defined modal verbs and in a set of texts belonging to defined categories leads to influencing the frequency of each modal value, which means that the frequency values may be completely different in the annotation of another sub-corpus of sentences or in non-sampled text. Moreover, choosing to annotate sentences and not full texts has two main implications: first, we cannot treat co-referential relations between linguistic elements outside the sentence, so that, when we face an ellipsis in a clause, we cannot recover the omitted element, unless it is expressed within the sentence containing the clause; and second, annotating sentences without their original context leads to find more ambiguities, while a larger context could disambiguate their meaning.

In the rest of the chapter, we will present general observations and interesting findings about the annotation of each component of our annotation scheme, from the modal values and the polarity, to the triggers, sources and targets.

TABLE 10

Portuguese verbs	**English translation**
Saber	To know
Pensar	To think
Crer	To believe
Perceber	To understand
Julgar	To judge
Permitir	To allow
Deixar	To let/allow
Conceder	To let/allow
Exigir	To require
Obrigar	To oblige/force
Vedar	To forbid
Tentar	To try/attempt
Experimentar	To try
Arriscar	To risk
Procurar	To try
Ensaiar	To try out
Conseguir	To succeed
Ser capaz de	To be able to
Chegar a	To reach/manage
Precisar	To need
Necessitar	To need
Alcançar	To reach/succeed
Atingir	To reach/succeed
Falir	To fail
Falhar	To fail
Querer	To want
Apetecer	To fancy/enjoy
Desejar	To desire/wish
Esperar	To hope
Aspirar	To aspire to
Poder	Can/may/be able to
Dever	Must/have to/should
Ter de/que	Must/have to
Prometer	To promise
Comprometer-se	To commit
Responsabilizar-se	To take responsibility for
Haver de	Should/must
Garantir	To promise/guarantee

4.2 Modal values

In the previous chapter we have presented the modal values we consider in our annotation scheme. The annotation of the whole corpus allows us to understand the effectiveness and functionality of these modal values. From an overall view, we can affirm that our list of modal values was effective: each modal value, in fact, was annotated several times in our corpus. In Table 11, shown at the beginning of the following page, we report the list of modal values annotated in the left column, with the occurrences and percentage of frequency they had in our corpus in the middle and right columns. The table shows that the most expressed modalities are deontic and epistemic modality, as they occur 740 times (≈ 28,84%) and 739 times (≈ 28,81%), respectively. In more detail, within deontic modality, 581 (≈ 22,65%) occurrences correspond to *deontic obligation* and the remaining 159 (≈ 6,20%) to *deontic permission*. Within epistemic modality, the most expressed value is *epistemic possibility*, with 279 occurrences (≈ 10,87%), while the less expressed one is *epistemic doubt*, with only 29 occurrences (≈ 1,13%). *Volition* is very frequent too, counting 396 occurrences (≈ 15,44%), while the frequency of participant-internal modality is quite lower than that of epistemic and deontic modality and than that of volition, as this value only occurs 248 times (≈ 9,67%). The results match with our expectations as concerns the frequency of epistemic and deontic modality, while we did not expect volition to be so frequent. Moreover, the values *effort* and *success*, if compared to such values as *epistemic possibility*, *deontic obligation* and *volition*, had a low frequency value, but, if compared to *participant-internal necessity* and *capacity*, had a quite high frequency value and were more frequent than the value *epistemic interrogative*. As we have specified at the beginning of the chapter, this frequency values have been calculated in a specific corpus of sentences chosen extracting specific modal verbs; in another corpus of sentences or in non-sampled text, in fact, the frequency values may vary consistently.

From the overall annotation of our corpus, we can say that, on the one hand, it was easier to assign the modal value to the verbs expressing deontic modality, since deontic modality covers all the expressions conveying permission or obligation. On the other hand, within the verbs expressing epistemic modality it was more difficult to establish which sub-value they carried. This is due to the fact that epistemic modality covers more sub-values and that the difference between one sub-value and another is less visible. However, within epistemic modality, there were some sub-values whose annotation was straightforward. For example, the annotation of the expressions conveying the value *epistemic interrogative* was rapid, since this value is generally expressed through a question mark (235) or, in the case of indirect interrogatives, through a word introducing the question, such as *perguntou* ‘asked’ in (236).

TABLE 11. FREQUENCY OF THE MODAL VALUES.

MODAL VALUES	**NUMBER OF OCCURENCES**	**PERCENTAGE (%)**
EPISTEMIC	**739**	**28,81**
possibility	279	10,87
knowledge	183	7,13
belief	161	6,27
interrogative	87	3,39
doubt	29	1,13
DEONTIC	**740**	**28,84**
obligation	581	22,65
permission	159	6,20
PARTICIPANT INTERNAL	**248**	**9,67**
capacity	126	4,91
necessity	122	4,75
VOLITION	396	15,44
EVALUATION	159	6,20
SUCCESS	119	4,64
EFFORT	110	4,29

(235) *Julgas que ataquei o teu país e o destruí sem a ajuda do Senhor**?***
'Do you think I have attacked and destroyed you country without God's help?'

(236) *O social-democrata Miguel Macedo foi mais longe e **perguntou** ao ministro como é que se procede se um preso falecer dois dias antes de o chefe de Estado conceder os indultos.*
'The Social Democratic Miguel Macedo went more far and asked the minister what is the procedure when a prisoner dies two days before the day in which the president pardons him'.

In the whole, the annotation and identification of the modal values expressed by the triggers was quite straightforward. This derives from the fact that our list of modal values is precise but not too much detailed, being the values clearly defined, so that we were able to

distinguish one from another. However, often, difficulties were faced when we encountered expressions that could be interpreted as conveying multiple modal values. We will further discuss these kind of problems in the section related to the ambiguity of the modal value.

4.3 Polarity of the modal value

In the previous chapter, we have presented *Polarity* as one of the main components of our annotation scheme. This component is used as a place to mark if there is negation scoping on the modal value. We, in fact, assign to polarity two values: positive and negative.

In the whole corpus *positive* was the predominant value for polarity as we found 1902 modal values carrying positive polarity and 450 carrying negative polarity. Among them, positive polarity rarely is lexically marked. The only example in our corpus in which the positive polarity of the modality trigger was marked is sentence (237).

(237) *Sabemos **sim**, senhor Little Axe…*
'Yes, we know it, sir Little Axe…'

In Portuguese, in fact, positive polarity is not generally marked by a specific lexical element, while negative polarity is always marked. The use of the particle *sim* to mark positive polarity is exceptional and only occurs in specific contexts, such as answers to questions, and coordinated sentences or phrases presenting polarity contrast (Oliveira, 2003). In (238), the *sim* underlines the contrast between the positive polarity of the coordinated clause and the negative polarity of the first member of the coordination (*não acredito que seja tanto a presença ou ausência de petróleo ou urânio o mais importante* 'I don't believe that the presence or absence of petrol or uranium is the most important thing').

(238) *Não acredito que seja tanto a presença ou a ausência de petróleo ou de urânio o mais importante, mas **sim** a progressiva aproximação do Ocidente ao Terceiro Mundo e, de uma maneira geral, da população mundial.*
'I don't believe that the presence or absence of petrol or uranium is the most important thing, but it is instead the progressive approximation of the West to the Third World and, in a more general manner, of the world population'.

Except for these cases, the lexically marked polarity is the negative one. The most common and frequent element used in Portuguese to mark negative polarity is the particle *não* which when preceding a modal verb focuses on it and negates its modal meaning. In sentence (239), the negative polarity affects the modal value *epistemic knowledge* expressed by the trigger *compreendia* 'understand'.

(239) *Ali Alatas não pôs qualquer obstáculo e disse mesmo que* ***não*** *compreendia a política europeia que até agora vedava deslocações ao território a nível de embaixadores.*
'Ali Alatas did not oppose any obstacle and also said that he did not understand the European policy forbidding the ambassadors' displacements in the territory'.

In the whole corpus, out of the total of 450 expressions carrying negative polarity, 378 were the ones in which negative polarity was conveyed by *não*, representing *não* the most frequent word used to express negation.

However, other elements can convey negative polarity. First of all, the modal verb itself can convey negative polarity to the modal value. In Portuguese, it occurs with the verbs *vedar* 'to forbid', *impedir* 'to prevent' and other synonyms, which generally express *deontic permission* with negative polarity and with the verbs *falir* and *falhar* 'to fail', which may express *success* with negative polarity. In our corpus, we found 41 occurrences in which the negative polarity was conveyed by the verb *vedar* 'to forbid' (240), while only 4 occurrences in which it was conveyed by the verb *impedir* 'to prevent'. The predominance of examples in which negative polarity is conveyed by *vedar* 'to forbid' is due to the fact that vedar was one of the verbs for which we did queries in the CRPC, while *impedir* 'to prevent' was not. We also found around 9 occurrences of negative polarity conveyed by the verbs *falir* and *falhar* 'to fail'.

(240) *Mas logo que a água atingiu uma cota razoável "começaram a* ***vedar*** *[o acesso às águas públicas]", queixa-se o dirigente associativo.*
'But as soon as the water reached a reasonable cost "they started interdicting the access to public water", the associative director complains'.

In (240), the verb *vedar* 'to interdict' is itself the trigger for negation and focuses on the target *o acesso às águas públicas* 'the access to public water': the negation is, in fact, lexically expressed by the verb itself and not by other particles.

The same polarity and value is expressed by sentence (241)[26]. However, in (241), the trigger for negation is *não* focusing on the modality trigger *permitir*, which itself alone carries positive polarity. The difference is that in (240) the negation is lexically incorporated in the verb, while in (241) the negation is external to it. In the corpus, we found 23 examples in which the polarity of the value *deontic permission* was negated through an external negative

[26] Sentence (241) is the result of our introspection in order to show some polarity issues.

particle, while 47 were the examples in which the polarity was marked as negative by the verb itself. As we can see, as concerns the value *deontic permission*, we have found more cases of internal negation than of external negation, which fact we would not expect, especially knowing that we extracted 50 contexts both for the verb *permitir* 'to allow' and for the verb *vedar* 'to forbid'. This might depend on the fact that *vedar* 'to forbid' was generally annotated as deontic, while *permitir* 'to allow' sometimes also conveyed the value *epistemic possibility*.

(241) *Mas logo que a água atingiu uma cota razoável "**não** [permitiram mais o acesso às águas públicas]", queixa-se o dirigente associativo.*
'But as soon as the water reached a reasonable cost "they did not allow anymore the access to the public water", the associative director complaints'.

Another trigger conveying negative polarity is the adjective *difícil* 'difficult', which in sentence (242) marks the polarity of the modal value *evaluation* as negative.

(242) *Por mais forte que uma pessoa tente ser, é muito **difícil** [continuar com o mesmo estado de espírito com que se estava anteriormente].*
'Even if we try to be strong, it is very difficult to continue with the same state of mind we had before'.

Even in this case, if we insert a negative particle (*não*) before the negative trigger *difícil* 'difficult', the polarity of the modal value *evaluation* becomes positive (243)[27].

(243) *Por mais forte que uma pessoa tente ser, **não** [é muito difícil continuar com o mesmo estado de espírito com que se estava anteriormente].*
'Even if we try to be strong, it is not very difficult to continue with the same state of mind we had before'.

As we have seen, polarity is an important component of our annotation scheme, as it helps in distinguishing when the modal value conveyed is positive or negative. Moreover in this chapter, we will present the difficulties we found along its annotation.

4.4 Triggers

We define the trigger as «the word or string of words that expresses modality» (Baker et al., 2010). Different kinds of expressions can be identified as triggers for modality. In our

[27] Example (243) is a product of our introspection.

corpus, since we selected contexts from the CRPC doing queries for specific modal verbs, the most frequent triggers for modality were verbs. However, we also found nominal and adverbial triggers.

As we said in the previous chapter, we collected 50 sentences for each verb query so the minimum frequency for each verb is 50. However, certain verbs occurred more times in our sample. Among the verbs for which we did queries in the CRPC, the most frequent modal verbs were *poder* 'may/can' and *dever* 'must', the first counting 210 occurrences with modal meaning and only 11 without, and the second counting 218 occurrences with modal meaning and 10 without. Even *querer* was quite frequent, counting 103 occurrences, while the verb *ter de* occurred 78 times. The rarest verbal triggers were the verbs *falhar* 'to fail' with only 9 occurrences with modal meaning and 41 without, *garantir* 'to guarantee', which only carried modality in 3 sentences, *responsabilizar-se* 'to be responsible' with only 2 occurrences with modal meaning, and *falir* 'to fail', which only carried modality in one sentence.

On the one hand, the identification of the modal value for the different triggers was straightforward especially for some verbs, such as *saber* 'to know/to be able to', generally linked to the modal value *epistemic knowledge*, but which sometimes also conveys *participant-internal capacity* (244); *perceber* 'to understand', conveying *epistemic knowledge*; with the verbs expressing *epistemic doubt*, such as *duvidar* 'to doubt'; and with the verbs *acreditar* and *pensar* 'to believe' expressing *epistemic belief.*

(244) *Os investigadores não **sabem** explicar como chegou ao continente americano.*
'Researchers cannot explain how he arrived to the American continent'.

Deontic modality too was often easily recognizable, since it was expressed by verbs conveying either obligation or permission.

The annotation also was very rapid and immediate when we faced verbs expressing volition. In general, in fact, someone's desire is expressed through such verbs as *querer* 'want', *esperar* 'hope', *desejar* 'wish', *apetecer* 'want' (245) and *aspirar* in the sense of 'aim at' (246).

(245) *Nessa noite, em Carreiros, **apeteceu**-me cantar sòzinho*[28].
'That night, in Carreiros, I wanted to sing alone'.

(246) *Há quem **aspire** a ser ministro.*
'There are people who aim at being minister'.

[28] The examples extracted from the corpus are copied and pasted in our text, respecting the orthography they have in the original text.

Also the annotation of the values *effort* and *success* was quite straightforward, especially when they were conveyed the first by the verbs *tentar* and *procurar* 'try to' and the second by the verbs *conseguir* and *atingir* 'to succeed'. In more detail, if the trigger for the modal value success follows a trigger for the modal value effort, the identification of both the values is much easier: in (247), the trigger for the modal value success is *conseguir* 'succeeds', which follows the trigger for the modal value effort *tenta* 'tries'.

(247) *Inconformado com a decisão judicial, o arquitecto* ***tenta*** *negociar mais tempo com a criança e acaba por* ***conseguir*** *também jantar com ela às quartas e domingos.*
'Unwilling to accept the judiciary decision, the architect tries to negotiate more time with the child and also succeeds in having dinner with her on Wednesdays and Sundays'.

On the other hand, difficult verbs are *precisar* and *necessitar*, since it is difficult in some cases to distinguish if the necessity is internal to the participant or if it is imposed by others; and *ser capaz* 'be able to', which can express both *success* and *participant-internal capacity*. Finally, the verb *exigir* 'to require' too is often ambiguous as it can be associated to the two modal values *deontic obligation* and *volition.*

As concerns nominal and adverbial triggers, their frequency in our corpus was very reduced, since we did queries only for modal verbs and not for nouns and adverbs.

In our corpus of sentences, we found 28 nominal triggers. Among them, the noun *desejo* 'desire', expressing volition, was found 7 times and the noun *tentativa* 'attempt' occurred 4 times to convey the value effort. The nouns *objectivo* 'objective' and *capacidade* 'capacity' occurred 3 times each, the first conveying volition and the second participant-internal capacity. Moreover, we found 2 occurrences of the noun *possibilidade* 'possibility', linked to the modal value epistemic possibility, 2 occurrences of the noun *conhecimento* 'knowledge', conveying epistemic knowledge, 2 of the noun *opinião* 'opinion', conveying epistemic belief, and 2 of the noun *esperança* 'hope', conveying volition. Nouns occurring only once were *compreensão* 'comprehension', expressing epistemic knowledge, and *necessidade* 'necessity', conveying participant-internal necessity.

As concerns adverbs, in our corpus, we only found a few expressing modality and all of them were included within the verbal trigger conveying more force to the modality expressed by the modal verb. We especially found two occurrences of the adverb *obrigatoriamente* 'obligatorily', conveying deontic obligation, (248), one occurrence of the adverb *necessariamente* 'necessarily', also conveying deontic obligation (249), one of the adverb *provavelmente* 'probably', conveying epistemic possibility (250), and one of the

adverb *totalmente* 'totally' (251), conveying more strength to some personal evaluation.

(248) *A feminilidade tem de ser* ***obrigatoriamente*** *excluída do feminismo?*
'Does femininity have to obligatorily be excluded from feminism?'

(249) *O desenvolvimento de uma agricultura específica e de qualidade vai* ***necessariamente*** *obrigar a uma redução da população que ainda trabalha neste sector, considera a entrevistada cuja passagem pelo Ministério da Agricultura no tempo dos governos de Cavaco Silva a obrigou a um estudo profundo sobre o sector.*
'The development of a specific and qualitative agriculture will necessarily require a reduction of the population that still works in this area, says the interviewed who's passage in the Agriculture Minister during Cavaco Silva's government obliged her to deeply study the sector'.

(250) *E que deverá* ***provavelmente*** *ganhar, a julgar pelo acolhimento que o seu programa recebe.*
'And who probably will win, according to the approval his program receives'.

(251) *É* ***totalmente*** *irrealista pensar que pode ser de outra maneira.*
'It is totally unrealistic to think that it can be in another way'.

In our corpus, we did not find any example of the adverbs *talvez* and *possivelmente*, with the function of emphasizing the strength of some modal verb.

However, we have done small queries in the CRPC for such adverbs in order to observe their position in relation to their targets. We have extracted 5 sentences for each of the adverbs *necessariamente* 'necessarily', *obrigatoriamente* 'obrigatorily', *possivelmente* 'possibly', *provavelmente* 'probably' and *talvez* 'maybe', and annotated the adverb itself as trigger for modality. We have seen that the target of the modal adverb generally follows the adverbial trigger, as in sentence (252), but it can also precede it, as in sentence (253).

(252) *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].*
'Some people say that I should not pay attention to the topic, since, possibly, the seven teens of lines that the trustee of the reader dedicated to the topic passed unnoticed'.

(253) *Se os partidos políticos não desempenharem o seu papel e não se auto-reformarem, [outros surgirão]* ***provavelmente.***

'If the political parties do not accomplish their functions and do not auto-reform themselves, others will arise probably'.

From the observation of the examples reported, we can see how modality can be conveyed by different words belonging to different grammatical categories. In Table 12, in the next page, we present all the triggers found along the annotation of our corpus.

TABLE 12. TRIGGERS IN OUR CORPUS

VERBAL TRIGGERS	Aspirar ‘to aspire/desire’ Apetecer ‘to enjoy/want’ Arriscar ‘to risk/try’ Atingir ‘to reach/succeed’ Achar ‘to think’ Alcançar ‘to reach/succeed’ Acreditar ‘to believe’ Avaliar ‘to evaluate’ Autorizar ‘to authorize’ Aprender ‘to learn’ Aperceber ‘to understand/to realize’ Crer ‘to believe’ Conseguir ‘to succeed’ Comprometer to commit’ Compreender ‘to understand’ Concluir ‘to conclude’ Considerar ‘to think’ Ser capaz ‘to be able to’ Conhecer ‘to know’ Conceder ‘to allow’ Convencer ‘to convince’ Conquistar ‘to conquer’ Chegar a ‘to reach/succeed’ Criticar ‘to criticize’ Dedicar-se ‘to dedicate yourself to’ Defender ‘to defend’ Deixar ‘to let/allow’ Descobrir ‘to discover’ Desconhecer ‘to ignore’ Descrer ‘to disbelieve’ Desdenhar ‘to disdain’ Desejar ‘to wish’ Desprezar ‘to disdain’ Dever ‘must/to have to’ Duvidar ‘to doubt’ Ensaiar ‘to try’ Entender ‘to understand’ Esperar ‘to hope’ Exigir ‘to require’	Experimentar ‘to experiment/try’ Falhar ‘to fail’ Falir ‘to fail’ Garantir ‘to guarantee’ Gostar ‘to like/to enjoy’ Haver de ‘must/have to’ Ignorar ‘to ignore’ Imaginar ‘to imagine’ Impedir ‘to prevent’ Impor ‘to impose’ Importar ‘to import’ Insistir ‘to insist’ Interessar ‘to interest’ Interrogar ‘to question/to ask’ Julgar ‘to judge’ Necessitar ‘to need’ Obrigar ‘to oblige/to force’ Perceber ‘to understand’ Permitir ‘to permit/allow’ Precisar ‘to need’ Procurar ‘to try’ Poder ‘can/may’ Prometer ‘to promise’ Querer ‘to want’ Questionar ‘to doubt’ Reclamar ‘to complain’ Rejeitar ‘to reject’ Responsabilizar ‘to get responsible’ Saber ‘to know’ Supor ‘to suppose’ Suspeitar ‘to suspect’ Sustentar ‘to think/believe’ Tentar ‘to try’ Ter de ‘must/have to’ Ter que ‘must/have to’ Vedar ‘to forbid’ Visar ‘to aim at’ Valorizar ‘to value’
NOMINAL TRIGGERS	Ambição ‘ambition’ Capacidade ‘capacity’ Compreensão ‘comprehension’ Confiança ‘trust’ Conhecimento ‘knowledge’ Desejo ‘desire’ Dúvida ‘doubt’ Esperança ‘hope’ Ideia ‘idea’ Impressão ‘impression’ Insatisfação ‘dissatisfaction’ Intenção ‘intention’ Objectivo ‘objective’ Opinião ‘opinion’ Plano ‘plan’ Possibilidade ‘possibility’ Suspeita ‘suspect’ Tentativa ‘attempt’	
ADVERBIAL TRIGGERS	Obrigatoriamente ‘necessarily’ Necessariamente ‘necessarily’ Totalmente ‘totally’ Provavelmente ‘probably’	
COMPLEX TRIGGERS	Fazer questão ‘require’	

4.5 Sources

As we have reported in the previous chapter, we consider two types of source. These are: the source of the event mention, which is the producer of the sentence containing the trigger for modality; and the source of the modality, which is who experiences the modality.

In most cases, we do not know the identity of the source of the event mention and assume that it is the speaker or the writer, as we have seen in paragraph 3.4.2.

However, in other cases, we know who the source of the event mention is, especially when there is direct speech, as in (254), where the source of the event mention is *Robert Thomas*, and in (255), where the source of the event mention is *Maria Antonia Palla*.

(254) *"Quando fazemos a demonstração isto fica cheio de gente, mas quando perguntamos se querem experimentar fogem todos", descrevia* ***Robert Thomas****.*
'"When we do the demonstration this place is full of people, but when we ask them if they want to try they all run away", Robert Thomas said'.

(255) ***Maria Antónia Palla*** *explicou que "a troika de observadores falhou por parcialidade e nunca conseguiu gerar a confiança e a vontade indispensável para ultrapassar as dificuldades inevitáveis" de um processo de paz.*
'Maria Antónia Palla explained that "the troika of observers failed because of some partiality and never succeeded in building the necessary trust and desire to pass over the unavoidable difficulties" of a peace process'.

In our corpus, we tagged 2244 sources of the event mention as speaker or writer, since the identity of this source was not known. In the remaining 133 triggers, the source of the event mention had an identity. Of these known sources of the event mention, 60 are identified through human proper nouns and 65 through human common singular nouns; 3 are human common plural nouns (*reformadores iranianos* 'Iranian reformers'; *os ingleses* 'English people'; *os acusadores* 'the offenders'); 2 are non-human proper collective nouns (*Agência Lusa*; *DoseOne*) and other 3 are non-human common nouns (*a carta dirigida a Paulo Mendo* 'the letter to Paulo Mendo'; *um documento* 'a document'; *Isaias 66.3*).

Together with the source of the event mention, we identify the source of the modality.

In the case of nominal sources, we have observed that all types of nouns (singular and plural, common or proper nouns) can carry the function of source of the modality. As we have said in the previous chapter, also pronouns can occupy the role of source of the modality: the most common ones are personal pronouns (*você* 'you', *ele* 'he', *nós* 'we', *a gente* 'we', etc.) (256); however, also demonstrative pronouns (*este* 'this', *esse* 'that', *aquele* 'that', etc.),

(257), relative pronouns (*quem* 'someone who' in (258)), and indefinite pronouns (*alguém* 'somebody', *ninguém* 'nobody', *todos* 'everybody', etc.) can carry the function of source of the modality.

(256) *Só espero que **a gente** consiga recuperar fisicamente.*
'I just hope we can recover physically'.

(257) *Estes e outros casos de falta de sensibilidade e tacto constituem argumento para **aqueles** que em alguns países africanos ainda procuram criar dificuldades quanto às relações privilegiadas com Portugal.*
'This and other cases of lack of sensibility are an argument for those who in some African countries still try to create obstacles to the privileged relations with Portugal'.

(258) *E nas ruas, olho atentamente as fachadas das casas, como **quem** numa multidão perscruta os rostos e tenta descobrir aquele dum ser há muito perdido.*
'And in the streets, I attentively look at the fronts of the houses, as someone who, in a crowd, observe faces and tries to discover one that has been lost since long'.

As explained in chapter 3, the source of the modality can be carried also by pronouns which do not have the function of subjects. In our corpus, these cases occur mainly with the verb *apetecer* 'feel like', even if there are other verbs with the same construction, as for example *agradar* 'appreciate/enjoy' (259).

(259) *Um dos aspectos que também **lhe** agradou foi o "respeito com que trataram estas coisas".*
'One of the aspects he also enjoyed was the "respect used to threat these things"'.

With these verbs, the source of the modality is not the grammatical subject of the verb but the prepositional phrase with the function of indirect complement, generally expressed by pronouns in the dative paradigm. The grammatical subject, instead, carries the modal function of the Target.

Along the annotation we also faced modal verbs in impersonal constructions with SE. As reported in the previous chapter, this construction is characterized by the fact that it has no subject but always presents the particle *se*, which can be considered as the source of the modality, since it expresses the impersonality of the verb. In fact, in our corpus, we annotated the *se* of impersonal verbs as source of the modality and encountered it 112 times (260).

(260) *Em 1998 Aguardela lançava o projecto Megafone, esse sim, à época um cometa: procurava-**se** fazer convergir, num mesmo ponto, as linhas de orientação da electrónica com as fontes do folclore português.*
'In 1998 Aguardela launched the project Megafone, which at the time was a comet: the objective was to make the orientation lines of synthesizer music converge with Portuguese folklore music'.

In Portuguese, the source of the modality can also be expressed by treatment pronouns, which are forms often used to address someone. In sentence (261), the source of the modality *epistemic knowledge* expressed by *saber* 'know' is *a senhora* 'Madame', which is a treatment pronoun used to refer to the addressee of the speech event.

(261) *Então **a senhora** não sabe que não se pode andar por aí a fazer propaganda dos beligerantes?!*
'Does not Madame know that it is forbidden to advertise belligerents?!'

Other treatment pronouns can appear as source of the modality: firstly, the masculine form of the pronoun *a senhora*, *o senhor*; secondly, all the professional titles, such as *o doutor*, *o engenheiro*, *o professor* and others (262)[29].

(262) ***O doutor João** quer falar consigo relativamente à venda da casa.*
'Doctor John wants to talk to you about the sale of the house'.

In some sentences, the source of the modality is expressed by free relative clauses or by bound relative clauses: in sentence (263), we tag as source of the modality the whole free relative clause *Quem venha da CREL e use via verde* 'Who comes from CREL and uses via verde', since it defines who is obliged to turn left, while in sentence (264), we consider as source of the modality of the trigger *obrigue* 'imply' the bound relative clause *a razão que os levou a desistir* 'the reason why they quit'.

(263) ***Quem venha da CREL e use via verde** é obrigado a guinar para a esquerda para, logo a seguir, ter que guinar à direita e atravessar quatro faixas de rodagem se quiser virar para o IC17 e auto-estrada de Loures.*
'Who comes from CREL and uses via verde is forced to slue left and, immediately after, slue right and cross four roadways if he wants to ride towards the IC17 and the

[29] Example (262) is the product of our introspection.

Loures highway'.

(264) *Caso* ***a razão que os levou a desistir*** *não os obrigue a internamento hospitalar, então têm de pedalar como se estivessem em prova.*
'If the reason why they quit does not imply any hospital admission, they have to spin as in a competition'.

Among these types of sources, in our corpus, the most frequent source of the modality was the speaker or writer: we, in fact, encountered 762 sources of the modality tagged as speaker or writer. In 167 cases the source of the modality was expressed by proper nouns, while in 501 cases by common nouns. All proper nouns are human. Among the common nouns, 341 are human. Of these, 137 are singular common nouns and 204 are plural common nouns. Still among human nouns, we found 160 collective nouns tagged as source of the modality. The total amount of human sources was of 668, while non-human sources were 184. As we can see, most sources of the modality are human: in fact, we have defined modality as the expression of the speaker or participant's attitude towards what he or others are saying or doing, which leads the source of the modality to be human in most cases. However, as we have seen in the previous chapter, sometimes, sources can also be non-human, since the speaker or writer can present as source of the modality non-human and inanimate elements (*o tecido* 'the tissue', *o computador* 'the computer', *os dados obtidos* 'the obtained data', *os equipamentos electrónicos* 'the electronic equipment'). Finally, in the whole corpus 63 relative clauses were tagged as source of the modality: most of them (52) were bounded relative clauses, while only 11 were free relative clauses.

4.6 Targets

The target is the linguistic expression in the scope of the trigger. It, therefore, includes the elements affected by the modality expressed by the trigger. Along the annotation, we have come across different types of target, and from a general overview, we can divide them in three main categories. These are: a) nominal phrases; b) subordinate clauses; and c) main clauses.

In our corpus, we found 284 nominal targets, of which 44 were human and 240 were not. Among the human nouns, we found 4 proper nouns (e.g. *Samuel*) and 40 common nouns (e.g. *o homem* 'the man'); all human proper nouns were masculine, while among human common nouns only one was feminine (*a ministra* 'the ministry'). Among the non-human nouns, we found 88 nominal phrases composed of an article and a feminine common noun (e.g. *a vitória* 'the victory') and 69 nominal phrases composed of an article and a masculine common noun (e.g. *o triunfo* 'the triumph'). 4 nominal phrases were composed of a common

noun with no article (e.g. *tempo* 'time'), while 8 nominal phrases included a demonstrative adjective instead of the article (e.g. *este código* 'this code'). Only 6 targets were composed of a demonstrative pronoun alone (e.g. *isto* 'this', *isso* 'that', *aquilo* 'that').

In the whole, the most frequent type of target was a subordinate clause. We have classified four main types of subordinate clauses: the ones carrying a finite verb, continuous (265) and discontinuous; and the ones carrying the verb in the infinitive form, continuous and discontinuous (266).

(265) *Em todo o caso, sustentam [que é preferível saber a localização precisa dos dejectos].*
'In any case, they claim that it is better to know the precise location of the evacuation'.

(266) *[Por mais forte que] uma pessoa tente [ser], é muito difícil continuar com o mesmo estado de espírito com que se estava anteriormente.*
'However strong a person tries to be, it is very difficult to continue with the same state of mind of before'.

In our corpus we found approximately 1526 subordinate clauses as targets. Among them, 1125 sentences carry the verb in the infinitive form and 401 carry a finite verb. Looking at the subordinate clauses with the verb in the infinitive form, we distinguish 817 continuous clauses and 308 discontinuous ones. Among the 401 subordinate clauses with a conjugated verb, 363 were continuous and 38 discontinuous.

In some cases, also main clauses can be targets. This occurs especially when there are two juxtaposed clauses, one of which has a parenthetical function as in (267), where *como sabem* 'as you know' is within the main clause *só que no futebol tudo pode acontecer* 'in football everything can happen', being separated from it by two comas.

(267) *Só que [no futebol], como sabem, [tudo pode acontecer], pelo que vamos esperar para ver.*
'It is just that in football, as you know, everything can happen, so we will wait to see what happens'.

The same occurs in example (268), in which *julgavam eles* 'they thought' is the clause with parenthetical function containing the trigger *julgavam* 'thought' whose scope is the whole main clause (*mas na Ânglia devia-se viver perpetuamente à luz artificial* 'in Anglia

people probably always lived using artificial light').

(268) *[Mas na Ânglia], julgavam eles, [devia-se viver perpetuamente à luz artificial], pois a luz natural era coada por nevoeiros eternos que se assemelhavam aos que cobriam sempre as costas da Nova Vtopia.*
'But in Anglia, they thought, people probably always lived using artificial light, since natural light was strained by eternal fogs which were similar to the ones covering always New Utopia's coasts'.

To get an indication of which is the most frequent position of the target, we observed the position of the target in a sub-corpus of sentences extracted from our corpus. We analysed the position of the target in the first sentence of each of the fifty hits extracted for each verb. The total sub-corpus is composed of 50 sentences. In these 50 sentences, we found 51 modality triggers and 51 targets. Among these targets, 39 were in the right periphery of the trigger, 3 in the left one and 7 both in the right and left peripheries. Of course, these counts depend on the type of corpus we annotate.

The most frequent type of target, besides subordinate clauses, are nominal targets. In fact, as most triggers are verbs, the elements in their scope are most frequently subordinate clauses or nominal complements of the verb. In sentence (269), we have two modal verbs, one in the first subordinate clause (*tentar*, 'try'), expressing the modal value *effort*, and another in the third subordinate clause (*permita*, 'let'), expressing the modal value *deontic permission*: the target for the trigger *tentar* 'try' is the subordinate clause *provocar um conflito* 'to provoke a conflict', while the target for the trigger *permita* 'allowing' is the nominal expression *uma intervenção estrangeira* 'a foreign intervention'.

(269) *O Presidente georgiano acusou Abachidze de "estar a tentar [provocar um conflito para criar uma base jurídica que permita [uma intervenção estrangeira]]".*
'The Georgian President accused Abachidze of having been trying to provoke a conflict for creating a juridical base allowing a foreign intervention'.

From the observation of triggers and targets, we have seen that verbal triggers are the triggers that select more types of target, as they can select nominal (270) and prepositional targets (271), and verbal phrases (272).

(270) *Nossa Senhora de Fátima não quer [a incineradora].*
'Nossa Senhora de Fátima does not want the incinerator'.

(271) *Emparedado entre a glória e o ascetismo, aspira [a uma cristianidade homogénea, de que seria a trave mestra].*
'Imprisoned between glory and asceticism, he aspires at a homogeneous Christianity, of which he would be the main beam'.

(272) *Estes e outros casos de falta de sensibilidade e tacto constituem argumento para aqueles que em alguns países africanos ainda procuram [criar dificuldades quanto às relações privilegiadas com Portugal].*
'These and other cases of lack of sensitivity and touch are the argument for those African countries which are still trying to create obstacles in the privileged relation with Portugal'.

In this section, we have analysed the existing types of triggers, sources and targets, also observing the relations between triggers and targets and triggers and sources. In the next section, we will present all the difficult cases and ambiguities faced during the annotation of each component.

4.7 Ambiguities and difficult cases in annotation

The annotation of modality in the corpus has not always been an easy task: problems emerged especially as concerns the identification of the modal value conveyed by the trigger, since in some cases the expression can be associated to more than one modal value; however, we also found difficulties in the annotation of the other components of our scheme, specifically the polarity, the triggers, the sources and the targets.

Ambiguities and difficult cases in the annotation of the modal value

Since in the previous chapter we have shown how we annotate those expressions whose modal meaning is ambiguous, in this section, we decided to present the most frequent ambiguities in the modal value faced along the annotation of our corpus.

Before annotating our corpus, we already knew that we were going to find cases of ambiguity between values. The literature on modality in Linguistics especially reports on the ambiguity between epistemic and deontic modality. Palmer (1986), for example, shows how in English the auxiliary verbs *may* and *must*, typically associated to the epistemic value and to the deontic one respectively, in many cases are used in contexts where they can be interpreted either deontically or epistemically (273 – 274).

(273) *He **may** come tomorrow*. (Palmer, 1986: 121)

(274) *He **must** be in his office*. (Palmer, 1986: 121)

Along the annotation of modality in our corpus, the ambiguity between epistemic and deontic modality was very recurrent. In particular, we have found ambiguity between the sub-values *epistemic possibility* and *deontic permission*. This ambiguity is mostly conveyed by the verb *permitir* 'to allow', which can be interpreted as expressing a possibility, or a permission, if there is some external factor allowing someone to do something or allowing something to happen (275).

(275) *Um trio de criadores americanos (quarteto, com o coreógrafo Mark Morris) tinha o apoio repartido de três instituições europeias e três americanas, não escondendo alguns ser este um desenho estratégico em que «a canção da Europa lírica 'de referência' para esta co-produção internacional vai **permitir** a entrada da criação [contemporânea] nas grandes instituições americanas, muito conservadoras» (brochura da temporada da Ópera de Lyon).*

'A trio of American creators (quartet, including the choreographer Mark Morris) was supported by three European and three American institutions, being a strategic plan in which the «European 'reference' lyric song for this international co-production allows the entrance of [contemporary] creation in the big very conservative American institutions» (brochure of the Opera of Lyon')'.

In (275), *permitir* 'to allow' is ambiguous: on the one hand, it expresses *epistemic possibility*, if we interpret the sentence as meaning that having the European lyric song in the international co-production **makes** the entrance of contemporary music into the American institutions **possible**; on the other hand, it expresses *deontic permission*, if we interpret it as meaning that the European lyric song in the international co-production is a **necessary condition** for the entrance of contemporary creation into the American institutions.

This ambiguity is the most frequent and occurs 34 times in the whole corpus. If compared to the frequency of epistemic possibility [279], the ambiguity has not a very high frequency value, but if compared to the frequency of deontic permission [159], we can see that it represents 21% of the total occurrences of deontic permission (Table 11). It is, then, important to keep the ambiguity annotated.

In few sentences, epistemic modality also overlaps with participant-internal modality: we found only 6 cases of ambiguity between the values *participant-internal capacity* and *epistemic possibility*, the first occurring 126 times alone and the second 279 times; the

ambiguity is conveyed especially by the verb *poder* 'can/may', which can be interpreted as expressing capacity or possibility in sentence (276):

(276) *A água das fontes que brota entre pedras, bebida de noite, quando mais fria,* ***pode*** *conceder a juventude eterna, talvez - quem sabe? - a imortalidade.*
'The water spreading from the fountains between stones, when drunk at night, if cold, can give eternal youth, and maybe – who knows – immortality'.

In (276), *poder* 'can/may' is ambiguous as the water **has the capacity** to give eternal youth (participant-internal capacity) or **may possibly** give eternal youth (epistemic possibility).

As we have seen, this ambiguity is very rare, proving the existence of both the values participant-internal capacity and epistemic possibility effective.

As concerns deontic modality, in our corpus, the sub-value *deontic obligation* overlaps 25 times with the value *volition*. This ambiguity is especially conveyed by the verb *exigir* 'to require', which, in fact, can indicate that the subject imposes his will or desire on someone else (277).

(277) *Médicos de hospitais do Norte* ***exigem*** *pagamento de horas extras.*
'The doctors in the hospitals in the north require the payment of extra hours'.

The same ambiguity is present also in (278), where the parents of the girls 'wanted' and 'required' the formation of a new class only for girls.

(278) *De acordo com o director do liceu, foram os próprios pais das raparigas que* ***exigiram*** *a constituição de uma classe especial.*
'According to the director of the high school, it were the parents of the girls who required the constitution of a special class'.

Moreover, we found 14 cases of ambiguity between the value *success* and the value *participant-internal capacity*. Both the values are conveyed by the triggers *conseguir* 'succeed' and *ser capaz* 'be able to/can', and in some sentences it is not clear what is the reason why the source succeeds or not in something, if it depends on a personal capacity or not. For example, in sentence (279), we do not know what is the reason why the speaker cannot face what is done.

(279) *Não somos* ***capazes de*** *encarar o que se faz hoje.*

'We cannot face what is done today'.

Although both the ambiguities (the one between deontic obligation and volition and the one between success and participant-internal capacity) occur, their frequency is still very low if compared to the frequency of the values alone. In fact, if the ambiguity between deontic obligation and volition occurs 25 times, the value *deontic obligation* alone occurs 581 times and the value *volition* 396 times, which makes us understand the importance of keeping both the values in our modal system. As concerns the second ambiguity, the values alone are very frequent too, compared to the occurrences of the ambiguity: if the ambiguity between success and participant-internal capacity is presented 14 times, the value *success* alone occurs 119 times, while the value *participant-internal capacity* alone occurs 126 times.

In the end, the expressions conveying the value *effort* can also be interpreted as expressing the value *volition*. The verbs *tentar* and *procurar* 'to try', for example, can be interpreted as expressing both values in most cases, since if a person tries to do something it is because he wants it to happen. Sentences (280) and (281), for example, can be both considered as ambiguous.

(280) *Covilhã **tenta** revitalizar folclore.*
'Covilhã tries to revitalize folklore'.

(281) *Marcos **procurava** rezar.*
'Marcos tried to pray'.

However, as we explain in chapter 3, we annotate the verbs *tentar* 'to try' and *procurar* 'to try' with the value *effort*, as they highlight more the effort of the person to do the action, while the verbs *querer* 'to want', *desejar* 'to desire' and other synonyms, with the value *volition*, as they highlight more the will and desire of the participant. In fact, along the annotation, this ambiguity was encountered only 7 times, while the values alone occur the first (*effort*) 110 times and the second (*volition*) 396 times, showing the importance of considering both the values in our modal system.

From an overall observation of the ambiguities found in the modal value, we can say that the ambiguity depends on how we interpret the sentence. Moreover, further context in most cases can lead to disambiguate the expression. For example, if the trigger *ser capaz* 'can' is ambiguous in the previous sentence (279), in sentence (282)[30] it can be associated

[30] Example (282) is a product of our introspection and shows how the expression *ser capaz* can

only to the modal value *participant-internal capacity* thanks to the expression *porque perdemos as nossas energias todas ontem*.

(282) *Não somos **capazes de** encarar o que se faz hoje porque perdemos as nossas energias todas ontem.*
'We cannot face what we do today because we have lost all our forces yesterday'.

As we can see, in the whole corpus, we only found 135 ambiguous triggers, which correspond to 5,68% of the total number of triggers (2377). The fact of having found only few ambiguities can be interpreted as an indication of the functionality of our system of modal values for the annotation of modality in text.

Difficult cases in the annotation of the polarity of the modal value

Along the annotation, identifying the polarity of the modal value was sometimes difficult. We especially found six cases in which the identification of the polarity of the modal value was more complicated. These were: a) interrogative sentences; b) when the trigger in the scope of the negative particle scopes over another modality trigger; c) when the trigger in the scope of the negative particle conveys itself negative polarity; d) when the negation is contained in the target; e) when the negation is contained both in the trigger and in the target; f) when the negation is carried by the source of the modality.

As we explained in the previous chapter, in our annotation scheme we only consider the values *positive* and *negative* for polarity. Therefore, we mark the polarity of interrogatives as positive, since positive is the unmarked value. However, this is a decision taken for convenience. When, in fact, the sentence is either a direct or indirect interrogative clause, it is difficult to mark the polarity as positive or negative, since, in reality, the modal value has no polarity at all. A possible solution to capture this information is to create a value *neutral*, which would allow marking the neutrality of interrogatives (283).

(283) *Julgarão que violar promessas aumenta a confiança dos cidadãos?*
'Do you think that violating promises make the citizens' trust grow?'

Moreover, the value *neutral* for polarity may also be used to define the polarity of the modal value *epistemic doubt*, since a doubt means that there is no certainty about something (284).

---

be disambiguated by expressing further context.

(284) *Não sei se eu tenho se não, mas o que é certo é que desde o princípio tentei não ser igual a ele.*
'I don't know if I have it or not, but certainly since the beginning I tried not to be like him'.

Sentence (284) is interesting as it could be annotated in two different ways. If we consider as trigger *sei* 'know', the modal value is *epistemic knowledge* and the polarity is negative. However, the most immediate interpretation is to consider as trigger the whole expression *não sei se eu tenho se não* 'I don't know if I have it or not' whose modal value is *epistemic doubt* with positive polarity.

Even though the use of the value *neutral* for polarity would be appropriate for these two modal values, it could make the task more complex for other modal values where it is difficult to distinguish between positive or neutral polarity. In sentence (285)[31], for example, the polarity of the modal value *epistemic possibility* conveyed by the trigger *for possível* 'it is possible' could be marked as *neutral* since we do not know if it is possible or not to catch a taxi, but we mark it as *positive*, since *positive* is the unmarked value for polarity.

(285) *Se for possível, prefiro apanhar o taxi.*
'If it is possible, I prefer to go by taxi'.

The second interesting aspect about the polarity of the modal value is found in those sentences where there are two triggers, the second of which is in the target of the first. In these cases, we have observed that the negative particle scoping on the first modality trigger conveys negative polarity to both the following triggers. In sentence (286), for example, the negative particle *nunca* 'never' conveys negative polarity to the trigger *conseguir* 'manage', expressing *success*, and to the trigger *crer* 'believe', expressing *epistemic belief.*

(286) *É este um vício que sempre atinge os míseros:* ***nunca*** *conseguir crer na felicidade!*
'And this is a vice that always affects poor people: to never manage to believe in happiness!'

The third interesting case regards those sentences in which the polarity of the modal value is positive, even if there is a negative element scoping on the modality trigger. This occurs especially when the modality trigger itself carries a negative meaning. In sentence

[31] Example (285) is not from the corpus but has been constructed to show neutral polarity.

(287) we have two separate triggers: the first, annotated with the modal value *evaluation*, is *difícil* 'difficult', and carries positive polarity since it is in the scope of the negative particle *não*; the second is *saber* 'know', which conveys the modal value *epistemic knowledge* and carries positive polarity, as it is in the scope of the modal expression *não é muito difícil* 'it is not very difficult'. The real meaning of the sentence, in fact, is that 'it is easy' to know which are the real ones and which depend on the fact that there has been the 1st of April.

(287) *Não é muito difícil saber quais as verdadeiras e quais as que têm origem no facto de termos atravessado o 1º de Abril.*
'It is not very difficult to know which are the real ones and which are the ones born from the fact that we have been living the 1st of April'.

The fourth interesting case is when the negation of the sentence is expressed in the target of the modal verb. In this case, the modal verb does not carry negative polarity, but the sentence conveys anyway a negative meaning (288).

(288) *O João obrigou [a Maria a* ***não*** *entrar na casa da amiga com aquela raiva] para evitar que a discussão acabasse em luta.*
'John obliged Mary not to enter in her friend's house with that anger in order to prevent that the discussion ended in a fight'.

However, our annotation scheme does not yet cover the observation of the polarity expressed in the target, since in this initial phase, we only look at the polarity of the modal value.

In some cases, the negation can be carried both in the trigger and in the target. The negation in the trigger focusing on the negation in the target conveys to the sentence an overall positive polarity. In sentence (289), for example, the value *evaluation* expressed by the adjective *impossível* 'impossible' focuses on the expression *não tínhamos capacidade* 'we had no capacity' in the subordinate clause. We annotate both the polarity of the first trigger and of the second trigger as negative, since the two triggers alone carry negative polarity, but the overall polarity of the sentence is positive. However, since through our scheme, we only analyse the polarity of each modality trigger alone, we do not annotate the polarity of the whole sentence.

(289) *Era* ***impossível*** *dizer que* ***não*** *tínhamos capacidade para crer, para amar ou para adorar.*

'It was impossible to say that we had no capacity to believe, to love or to worship'.

Last, negative polarity can also be conveyed by the source of the modality. Since the source of the modality scopes on the trigger, if the source of the modality contains a negative element, the negative element conveys negative polarity to the trigger and to the modal value. Some linguistic elements tagged as source of the modality that can express negative polarity are: a) negative pronouns, such as *ninguém* 'no one' in sentence (290), conveying negative polarity to the modal value *epistemic belief* expressed by *acredita* 'believes'; and b) negative specifiers, such as *nenhuma* 'no' in sentence (291), conveying negative polarity to the modal value *participant-internal capacity* expressed by *consegue* 'can'. In our corpus, we found 11 occurrences of the pronoun *ninguém* 'no one' and 6 of the adjective *nenhum* 'no'.

(290) *Hoje* ***ninguém*** *realmente acredita que os políticos que elege sejam capazes de fazer alguma coisa em relação aos problemas que estão no topo das preocupações: o desemprego, a droga, a insegurança.*
'Today no one really believes that the politicians one votes for can do something as concerns the most worrying problems: unemployment, drug, insecurity'.

(291) ***Nenhuma*** *máquina de focagem fixa consegue fazer fotos focadas a essa distância.*
'No fixe focusing machine can take focused pictures with that distance'.

Difficult cases in the annotation of triggers

The identification of triggers was in most cases immediate and did not create big problems. The only difficult case we found was the complex trigger *fazer questão*, which can be translated in English as 'to insist'. We consider this expression as a complex trigger for the modal value *volition*, since in sentences as (292) it means that the source truly wants something, while *questão* alone does not carry any modal value at all.

(292) *A parada* ***fez questão*** *de tentar demonstrar o contrário, com participantes de todo o mundo.*
'The parade really insisted in trying to demonstrate the opposite, with participants coming from all over the world'.

Ambiguities and difficult cases in the annotation of sources

In the annotation of sources, we faced difficulties especially as concerns the

identification of the elements to be tagged as source of the modality when in the sentence there was no source of the modality expressed.

Since Portuguese is a null-subject language, very often the source of the modality is not lexically expressed in the sentence. In some cases, however, it can be recovered through the form of the verb. In the previous chapter, we have explained that when the source of the modality is not lexically expressed we tag the verb both as trigger and as source of the modality, since it carries the subject information. In example (293), the source of the modality is not expressed but we know that the speaker or writer is talking about another person from the fact that the verb *precisar* is conjugated as the singular third person (*precisa* 'must').

(293) ***Precisa*** *de ser mais competitivo, de ter mais continuidade durante os jogos.*
'He/she must be more competitive and have more continuity during games'.

However, in other cases, the source of the modality cannot be recovered immediately. In fact, some finite forms of the verb can be associated to multiple sources since they have the same ending as other finite forms. It is very frequent with regular verbs for the first and third singular person forms of the indicative imperfect and past perfect, of the subjunctive present, imperfect and future, and of the personal infinitive. Example (294) shows how the ambiguity of the source of the modality is conveyed by the past perfect tense of the verb; example (295) shows how the ambiguity occurs also with the subjunctive future and with the present tense of the indicative; and in sentence (296)[32] the ambiguity is conveyed by the personal infinitive.

(294) ***Permitira*** *que os miúdos jogassem fora mesmo que se sujassem por causa da terra molhada.*
'I/he/she let the kids play outside even if they would get dirty because of the wet soil'.

(295) *Se* ***tentar****, pode ser que* ***consiga****.*
'If I try, I might succeed'.
'If he/she tries, he/she might succeed'.

(296) *O João disse para* ***deixar*** *os miúdos brincarem no quarto dele.*
'John said to let the kids play in his room'.

As we can see, in all these examples we can interpret as source of the modality the speaker or writer or another person whom the speaker or writer is referring to, but which is

[32] Example (294), (295) and (296) are not sentences extracted from the corpus but have been constructed to show the ambiguity of the source conveyed by the subjunctive future.

not the hearer.

Ambiguities and difficult cases in the annotation of targets

The annotation of our corpus led us to face several difficult cases as concerns the identification of targets. The most frequent difficult cases we encountered were: a) the sentences in which the target is not expressed; b) the ones in which the target is itself a trigger for another modal value; c) the sentences in which the target is the same as the trigger; and d) the sentences in which the target is the same as the source of the modality.

As we have said in the previous chapter, when there is no expression in the scope of the trigger, we do not annotate it. In our corpus, we found 44 examples in which the target is not expressed in the sentence and can only be recovered from a larger context. In sentences (297) and (298), for example, there is no target for the modality triggers: in sentence (297), there is no expressed target for the verb *saber* 'know' conveying the modal value *epistemic knowledge*; in sentence (298), there is no expressed target for the trigger *escape fácil* 'easy escape' conveying the value *evaluation.*

(297) *Felizmente ninguém na plateia respondeu: "Sabemos sim, senhor Little Axe…"*
'Fortunately no one in the audience answered: "Yes, we know, sir Little Axe…"'

(298) *Mas foi escape fácil perante a dificuldade de relacionamento com a longa e silenciosa travessia da América que antecede essa sequência, contornos de um "road-movie" que Gallo torna numa obra profundamente melancólica, mais tradução plástica de estares de alma do que procura narrativa.*
'Anyway it was an easy escape to the difficult relations with the long and silent crossing of America preceding that sequence, like a road-movie that Gallo turns into a profoundly melancholic work, which is a plastic translation of the states of the soul more than a narrative search'.

From the observation of the data we annotated, we have observed that all triggers can appear in the sentence without their targets expressed, even deontic verbs, whose target can be present outside the sentence so that in the annotation at sentence level it cannot be recovered (299)[33].

(299) *Eles não pedem, exigem.*

[33] Sentence (299) is a made-up example to show how the target can be present outside the sentence.

'They do not ask for it, they require it'.

However, in our corpus, the targets of the triggers conveying deontic modality were always lexically expressed within the sentence.

A recurrent situation faced during the annotation was that there were two triggers for two different modal values and the second was also part of the target of the first trigger, as in (300), where *permitir* 'to allow' is itself a trigger for modality but is also part of the target of the preceding trigger *querem* 'want'.

(300) *Parecem-se com os livros de papel,* ***querem*** *[****permitir*** *[sensações tácteis parecidas e dar ao leitor algo mais]].*
'They are similar to paper books, they want to allow similar tactile sensation and give the reader something more'.

In the whole corpus, we found different examples in which this situation occurred, and we could observe some typical patterns in the modal behaviour of some verbs when co-occurring in the sentence.

First of all, if the first verb conveys the value epistemic possibility, the certainty of the modal value conveyed by the second verb is conditioned. In sentence (301), for example, the verb *poder* 'can', conveying the modal value *epistemic possibility*, influences the degree of certainty of the modal value *success*, conveyed by the verb *conseguir* 'achieve': the 'success' conveyed by *conseguir* is, in fact, not as certain as in sentence (302)[34].

(301) *A protecção que as plantas* ***podem*** *[****conseguir*** *quanto a uma evaporação excessiva] é modificar a sua estrutura de modo a diminuir, ou mesmo evitar, a transpiração.*
'Plants can achieve protection from excessive evaporation modifying their structure to reduce or avoid transpirations'.

(302) *A protecção que as plantas conseguem quanto a uma evaporação excessiva é modificar a sua estrutura de modo a diminuir, ou mesmo evitar, a transpiração.*
'Plants achieve protection from excessive evaporation modifying their structure to reduce or avoid transpirations'.

As we have explained in the previous chapter, through our scheme, we apply the full annotation to the two triggers: we annotate the first trigger *podem* 'can' with the modal value

[34] Example (302) is a product of our introspection and has been constructed on the base of example (301) to show the contrast between the expression *podem conseguir* and *conseguem*.

epistemic possibility; the second trigger is annotated twice, the first time as part of the target of the first trigger, and the second time as trigger itself. In this way, the annotation expresses that the value of possibility conveyed by the verb *poder* 'can' influences the value success conveyed by the verb *conseguir* 'achieve' in its target. The resulting annotation of sentence (301) is the following:

Trigger: podem
Target: a protecção que@conseguir quanto a uma evaporação excessiva
Source of the event mention: sp/wr
Source of the modality: as plantas
Modal value: epistemic_possibility
Polarity: positive
Trigger: conseguir
Target: a protecção
Source of the event mention: sp/wr
Source of the modality: as plantas
Modal value: success
Polarity: positive

In our annotation, there were several examples in which the value *epistemic possibility* influenced the certainty of other values, specifically of the values *evaluation* (303) and *deontic obligation* (304).

(303) *Se o aluno se perde,* ***pode*** *[ser* ***difícil*** *voltar a apanhar].*
'If the student looses himself, it can be difficult for him to catch up again'.

(304) *Neste sentido, a APAF afirma recear que estejam a ser criados ambientes que* ***poderão*** *[****obrigar*** *à não comparência dos árbitros em campos onde as suas integridades fisica e moral possam ser postas em causa].*
'In this sense, APAF expresses its concern about the possibility that the environments being created might prevent the referees to show up in soccer fields, where their physical and moral integrity might be in danger'.

Other values that influence the certainty of the modal verb in the sentence are the values *epistemic interrogative* and *epistemic doubt*. In fact, if the triggers for these two modal values scope over another modality trigger, the certainty of the modality expressed by the second trigger is conditioned by the first trigger. Sentences (305) and (306) are examples of

this process.

(305) ***[E és capaz de conjugar o verbo gostar no presente do indicativo?]***
'And are you able of conjugating the verb gostar in the present of the indicative?'

(306) *Também* ***não sei*** *[se* ***precisamos*** *de dez orquestras semelhantes para a europa de 94], conforme pressenti nas preocupações do maestro, mas tenho a infeliz certeza de que precisamos de ordenados compatíveis com o não morrer à fome nos próximos meses.*
'I also don't know if we need ten similar orchestras for the Europe of 94, as I have felt from the worries of the conductor, but I have the unhappy certainty that we need sufficient salary not to die of hunger in the next months'.

In sentence (305), the whole interrogative sentence carrying the value *epistemic interrogative* marks as uncertain the value *participant-internal capacity* expressed by the trigger *ser capaz*; in sentence (306), it is the expression *não sei* 'I don't know' which marks as uncertain the value *participant-internal necessity* expressed by the verb *precisamos* 'we need'.

We frequently also observed that the verbs expressing *participant-internal capacity* (*ser capaz*, *saber*) are often in the target of the verbs *dever* and *ter de* 'must', expressing *deontic obligation*, as in (307) and (308).

(307) *As suas mãos* ***terão de*** *[****saber*** *exprimir os sentimentos da alma] e* ***deverão*** *[****ser capazes*** *de transmitir ternura].*
'His hands must be able to express the soul's feelings and have to be able to communicate tenderness'.

(308) *Espero bem que sim, mas para isso Portugal* ***terá de*** *[****ser capaz*** *de se mostrar humilde], estar concentrado e sobretudo errar menos que o adversário.*
'I hope it so, but for this Portugal must be able to show humility, must be concentrated and especially must make less errors than the opponent'.

Even in these cases, each trigger maintains its modal value, but the presence of a deontic verb scoping on the verb expressing *participant-internal capacity* marks the capacity as a necessary condition for the realization of some action or event.

Moreover, in our corpus we have extracted 50 hits for the verb *aspirar* 'to aspire' and have annotated them when the verb conveys the value *volition*. Along the annotation, we have

observed that in 21 hits, *aspirar* 'to aspire' is in the target of the verb *poder* 'have the possibility' (309). This is the only case in which the same two verbs co-occur several times in different sentences.

(309) *Apenas três ou quatro pilotos nacionais* ***poderão [aspirar*** *a classificações honrosas].*
'Only three or four national pilots will have the possibility to aspire for honourable classifications'.

In our corpus, we also observed those cases in which the trigger and the target for one modal value were expressed by exactly the same linguistic elements. This occurred only in the annotation of two types of sentences, interrogatives and imperatives. As we have seen in the previous chapter, in both these types of sentences, we annotate as trigger and as target the same linguistic elements, which in interrogative are the entire interrogative clauses (310), and in imperatives are the verbs in the imperative form (311).

(310) ***[Há o risco do povo ficar afectado por este abalo no reino vitoriano?]***
'Is there the risk that the population is affected by this trembling of the Victorian kingdom?'

(311) ***[Digam]*** *missas pela minha salvação!*
'Say masses for my salvation!'

In the whole corpus, in 36 of the 50 imperatives, the trigger was annotated also as target, while in 70 of the 87 interrogatives, the whole sentence was annotated both as trigger and as target. Therefore, in the whole, we found 106 cases in which we annotated as trigger and as target exactly the same linguistic elements.

We also observed that in 5 sentences of our corpus the target was the same as the source: in 4 of them, the modal value is carried by *falhar* 'to fail' and in one it is carried by *falir* 'to fail'. Sentence (312) is composed of two coordinated clauses both expressing the modality *success* with negative polarity through the verbs *falhar* 'to fail' in the first clause and *falir* 'to fail' in the second clause. In the first clause, the nominal phrase *os modelos de socialização antigos* 'the old socializing models' carries both the function of source of the modality and of target of the modality trigger *falharam* 'failed', while, in the second clause, the nominal phrase *os mecanismos de transmissão de saber* 'the mechanism of transmission of knowledge' is both the source of the modality and the target of the triggers *faliram* 'failed' and *estão a falir* 'are failing'.

(312) *A única coisa que se sabe de certeza é que* ***os modelos de socialização antigos*** *falharam e que* ***os mecanismos de transmissão de saber*** *faliram ou estão a falir.*
'The only certain thing we know is that the old socializing models failed and that the mechanisms of transmission of knowledge failed or are failing'.

Another interesting case of ambiguity between target and source of the modality was found in examples with the verbs *precisar* 'to need', *necessitar* 'to need', *ter de* and *ter que* 'have to', and synonyms. These verbs, in fact, can be associated to two different modal values according to which is the element that we consider as source. In sentence (313), for example, we can consider the constituent *muitas destas páginas* 'many of these pages' as source of the modality, or we can decide that it is part of the target, being then the speaker or writer the only source: in the first case, considering *muitas destas páginas* 'many of these pages' as source of the modality, the modal value is *participant-internal necessity*; in the second case, if we consider *muitas destas páginas* 'many of these pages' as part of the target, the modal value is *deontic obligation*, and the speaker or writer is both the source of the event mention and the source of the modality.

(313) ***Muitas destas páginas*** *precisam de ser preparadas para o efeito de modo a prevenir danos.*
'Many of these pages need to be prepared in order to prevent damages'.

Also with the verb *ter de* or *ter que* 'have to', this ambiguity is very frequent, as we can see from example (314), in which if we consider as target both the expressions *o campeão do mundo* 'the World's Champion' and *ganhar com justiça* 'win with justice', the modal value is *deontic obligation*, while if we consider *o campeão do Mundo* 'the World's Champion' as source of the modality and *ganhar com justiça* 'win with justice' as target, the modal value is *participant-internal necessity*.

(314) *Até* ***o campeão do Mundo*** *tem que ganhar com justiça.*
'Even the World's Champion must win with justice'.

In our corpus, we found 70 examples in which this phenomenon occurred with the verb *ter de/que* 'have to'.

As we can see, this phenomenon was very frequent but, as our scheme does not yet have a system to annotate these structural ambiguities, we were not able to keep the ambiguity in the annotation of the corpus; however, in further research, we can investigate this

ambiguity and find a solution for its annotation.

In this chapter, we have presented what we discovered about modality from the annotation of our corpus. First, we have focused on some typical patterns of behaviour of the components of our annotation scheme, observing the relation between the component and the linguistic expressions used to convey it. Second, we have focused on the difficulties and problems we have found along the annotation of each component, reporting on which were the most frequent ambiguities encountered.

As we have seen, the annotation scheme has proved an efficient instrument to study the expression of modality in corpora. A further confirmation has been given by an experiment we have done in order to measure the inter-annotator agreement (IAA) for tasks of modality annotation. The experiment consisted of using the scheme to annotate 50 sentences extracted randomly from the CRPC. Two linguists annotated the same 50 sentences independently from each other, using our annotation scheme and our system of modal values. We computed IAA using the kappa-statistic for each field in the annotation. For the Trigger the kappa value was .65 and for the accompanying Modal value a kappa of .85 was obtained, similar to the reported IAA for English (Matsuyoshi et al., 2010). As we can see, the results of the application of the annotation scheme to the same set of sentences by two different annotators resulted in a quite high degree of concordance as concerns the identification of modal values, but not as concerns the identification of triggers, as the two annotators did not always identify the same triggers. The kappa value for the modal value is a further confirmation of the efficiency of our list of modal values, but the lower kappa value for the trigger might indicate that we should better specify what to consider as modality trigger.

5. Conclusion

In this work, we have proposed a scheme for the annotation of modality in a Portuguese corpus. This scheme was created as an instrument to study the phenomenon of modality in the real use of Portuguese.

The research project followed three main steps. First, we presented and compared the existing literature on modality and on its annotation. The analysis of the existing previous studies on the topic is important in order to know what has been already done and what is still missing. When we studied the existing literature on modality in Linguistics, we found two main points that the different approaches have in common: first of all, they are built on the basis of the distinction between the concept of necessity and possibility; secondly, they all identify epistemic modality: differences between the approaches can be found in the values that oppose epistemic modality.

Studying the literature on the annotation of modality, we have seen that no proposal has been presented yet for the annotation of modality in corpora with the main objective of studying the linguistic expression of modality in Portuguese. In fact, as we have seen in chapter 2, the current practical annotation schemas are built for English with the main purpose of improving NLP applications. As these schemas are application-oriented, their choice of modal values is usually more coarse-grained than the proposed typologies that can be found in the Linguistics literature.

However, the proposals for the annotation of modality reported in the literature present a more detailed list of the elements in the sentence that are actively involved in the expression of modality. The authors especially identify, besides the modal value, the element that actually triggers the modality, the linguistic expression on which it has some influence, the linguistic element expressing the holder of the modality, and if there is negation scoping on the modality trigger.

The second step was the creation of our scheme for the annotation of modality. In it, we aimed to combine the more detailed and theoretical linguistics modal value typology with the practical schema of annotation which also captures the other elements involved in the expression of modality. In fact, the scheme we built is composed of components and modal values that have been taken from the different systems created for the annotation of modality in corpora and from the systems of modal values created in Linguistics by the different authors who have investigated on modality.

The third step of our research process is the most innovative aspect of our approach to the study of modality: in order to analyse how the speakers convey modality in Portuguese, we use corpus annotation. The scheme we have created has been tested applying it to a corpus of 1935 sentences extracted from a bigger corpus, the *Corpus de Referência do Português*

Contemporâneo, composed of samples of real written and oral discourse. The annotation of the corpus with our scheme allowed us to understand if the scheme was an efficient mean for capturing the linguistic phenomenon of modality in Portuguese. In fact, along the annotation of the corpus, we have further developed and fine-tuned the annotation scheme to capture as many aspects of modality in Portuguese as possible. For example, at the beginning we did not have two separate labels for the annotation of the producer of the event mention (Source of the event mention) and for the holder of the modality (Source of the modality). The necessity for the two components was felt as soon as we faced sentences in which the holder of the modality was not the same as the producer of the event mention, the speaker or writer.

The annotation of modality in the corpus allowed us to explore how modality is conveyed in the real use of Portuguese, instead of analysing the expression of modality in a set of sentences built specifically for the purpose. We annotated the corpus sentence per sentence and avoided, in this initial phase of the research, the annotation of entire texts. The annotation of the corpus at the sentence level was very useful to test the scheme: to each sentence containing a modality trigger, we applied the full annotation scheme and annotated for each component the linguistic expression conveying it.

Through the application of our scheme to the corpus, we identified the most expressed modalities in the corpus, the most used expressions to convey them, and the relations between modal values and between the modal values and the expressions used to convey them. This was done through the annotation of the components *Trigger*, used as a place where to annotate the linguistic expression with modal meaning, and *Modal value*, used as a place where to annotate the modal value expressed by the trigger. In our corpus, the most frequent triggers for modality were verbs, however also adverbs, nouns and adjectives can carry some modal meaning. The annotation of verbal triggers allowed us to discover new interesting verbs which are not generally considered as modal but which in many sentences could be annotated as triggers for modality. As concerns modal values, as we expected, the most frequent modalities expressed were epistemic and deontic modality. However, as we have seen in the Results chapter, also the value *volition* was very frequent.

Not always the identification of the modal value conveyed by the trigger is immediate: in some cases, in fact, an expression can be interpreted according to different modal values. Since during the annotation of our corpus, we found several expressions that could be associated to more modal values at the same time and in the same context, we decided to report this aspect in the annotation scheme in two ways: in the component *Modal value*, we write all the modal values the trigger can be associated to in that context, and in the component *Ambiguity*, we mark that the trigger is ambiguous through the word *Yes*. Annotating this component allows: a) to understand the efficiency of our list of modal values; b) to observe if there are recurrent ambiguities between modal values; c) to observe if there

are expressions that are always ambiguous: d) to catch the linguistic modal behaviour of some expressions. In our corpus, the most frequent ambiguity found was the one between the values *epistemic possibility* and *deontic permission*, while the less frequent ambiguity was the one between the values *participant-internal capacity* and *epistemic possibility*. Comparing the frequency of the ambiguities to the frequency of occurrence of the modal values alone, we got to the conclusion that our system of modal values is effective, since the number of ambiguities is much lower than that of the modal values alone.

Through the annotation of the other components of the scheme (*Source of the event mention*, *Source of the modality*, *Target*), we also observed all the elements involved in the expression of modality and affected by the modal value conveyed in the sentence. We especially found out that the *Source of the event mention* is generally the speaker or writer and in most cases is not lexically expressed in the sentence, and that the *Source of the modality* is frequently but not necessarily conveyed by a nominal phrase. There are, in fact, sentences in which the source of the modality is not lexically expressed at all and therefore we have to tag the verbal trigger also in the component *Source of the modality*, as it is the only linguistic element in the sentence carrying the subject information. Moreover, we expected the source of the modality to be always expressed by an animate element. However, we have found several examples in which the source of the modality was an object. As concerns the annotation of the component *Target*, we saw that most frequently the target is either a nominal phrase or a subordinate clause. However, targets can also be main clauses, in the case of interrogative sentences and when the subordinate clause containing the trigger precedes the main clause. We also observed that in some cases we tag as target and as source of the modality the same linguistic elements; it occurs especially with the grammatical subjects of the verbs *falir* and *falhar* 'to fail'.

Moreover, the annotation of the component *Polarity* allowed us to analyse how the negation of modality is expressed in Portuguese: we have observed that the lexically marked polarity in Portuguese is the negative one, which generally is expressed by the negative particle *não*, even though also other elements can be used to convey it.

As we can see, the annotation of all these components allows us to give a deeper insight in the phenomenon of modality in Portuguese, since through their annotation we can capture all the elements in the sentence affected by the modal value expressed.

The annotation scheme we have created is an important instrument for the study of modality in Portuguese. First of all, our scheme can be used by linguists who want to explore the expression of modality in specific texts or for those who want to explore the meaning of some modal expressions in Portuguese. Second, the existence of a corpus annotated with modal information can be useful for different investigation fields. For example, in the field of computer science, some NLP applications may profit of having modality annotated in corpora

in order to improve their efficiency. It is, for example, the case of the application Automatic Translation, used to automatically translate text from a language to another, which may be advantaged of having modal information annotated in the text in order to catch all the little hues of a linguistic expression in a specific context (Baker at al., 2010). Also the Sentiment Analysis application would profit of having text annotated with modal information. This application extracts subjective information from a set of documents in order to identify trends of public opinion in the social media, for the purpose of marketing: having text annotated with modal information would help in distinguishing personal opinions, beliefs and desires of the participant from factual information and from obligations and duties (Wiebe et al., 2005).

As we have seen, the annotation of modality is an important process that can be used both in Linguistics and in other investigation fields for different purposes. In the next section, we will present the possible future applications of our annotation scheme and some possible further study to be done in the topic.

Future work

The research we have done on the annotation of modality in this work has given important information on the expression of modality in Portuguese. However, there are still some tasks to be fulfilled in further research.

As we have already said, the scheme we have proposed is the first existing proposal to annotate modality in corpora in Portuguese. Till now, our annotation scheme has been tested specifically in a corpus of Portuguese sentences extracted from the CRPC. However, it would be interesting to use the scheme to annotate entire texts. The annotation of entire texts would imply the observation of more elements in the text, since we could treat coreferential relations between words and recover elements which are not expressed in the sentence but are expressed in the whole text (ellipsis). We also believe that the annotation of full text would permit the discovery of other interesting issues about modality: we could, for example, find interesting non-verbal triggers for modality and study in a deeper way the relations between them and their targets and sources.

We are also planning to test the annotation scheme on spoken material, to see if it can be used to annotate modality in oral texts and to understand if the strategies used to express modality are the same as in written texts. We have already done a preliminary application of our annotation scheme to some spoken material extracted from the Portuguese C-ORAL-ROM corpus (Bacelar de Nascimento et al., 2005), and have observed that in spoken texts there are new elements carrying modal information. While in written language, modality is conveyed especially at the lexical and syntactic level, in oral language, modality can also be conveyed through extra-linguistic elements, such as silent pauses or pauses filled by some

extra-linguistic marker (*hum*), which may be used by the hearer to indicate some doubt about what the speaker is saying. Moreover, in spoken language there are some lexical expressions used specifically to convey the value *epistemic doubt*, such as *vamos lá ver* 'let's see' (315), and the value *epistemic belief*, such as *ora bem* 'well' (316)[35].

(315) *A: eu acho que o cortar com a língua francesa / implica uma viragem também nesse continuum <cultural> //$*

'A: I think that cutting off French / implies a change also in that cultural continuum //$'

B: [<] <pois> //$ ***vamos lá ver*** *//$ porque / o francês já está / desde há uns anos para cá / numa situação de inferioridade em relação ao inglês //$*

'B: [<]<yes> //$ let's see //$ because French is already / since some years ago / in a situation of inferiority with respect to English //$'

(316) *A:* ***ora bem*** */ &ah / esta é uma exposição / de / digamos / um bocadinho ligada aquela preocupação / de mostrar / os trabalhos do cinema de animação português /*

'A: well / &ah / this is an exhibition / of / let's say / a bit linked to that concern / of showing / the Portuguese animation films /'

Other expressions related to the value *epistemic knowledge* and used specifically in the oral to mark agreement or understanding are *sim, eu sei* 'yes, I know' (317), *é é* 'ya ya' (318) and *pois* 'yes' (319).

(317) *A: olha / &eh / daqui mãe //$*

'A: look / &eh / here is mom //$'

B: ***sim / eu sei*** */ já <percebi / hhh> //$*

'B: yes / I know / I <understood / hhh> //$'

---

[35] The annotation guidelines are described in Bacelar do Nascimento, M. F., J. Bettencourt Gonçalves, R. Veloso, S. Antunes, F. Barreto, R. Amaro "The Portuguese Corpus", in Cresti, E. & M. Moneglia (eds.), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, John Benjamins Publishing Company, Studies in Corpus Linguistics nº 15, Amsterdam/Philadelphia, pp. 163-207 (with DVD).

(318) *A: / [<] <é uma questão> de insegurança //$*

'A: / [<] <it is a problem> of insecurity //$'

*B: é //$ <**é / é**> //$*

'B: yes //$ <ya / ya> //$'

(319) *A: [<] <mas isso depende> de como as vais usar //$ com saias tem que ser altas //$*

'A: [<] <but that depends> on how you'll use them //$ with skirts they must be high //$'

*B: <**pois**> //$*

'B: <yes> //$'

In addition, the scheme, even though it has been created for Portuguese, can be used to annotate modality in other texts or corpora for other languages in order to explore how modality is conveyed in those languages. We, in fact, truly believe that the scheme could be used to annotate modality in other romance languages, for example Italian. Through the annotation of an Italian corpus with our scheme, we could, first, discover which are the expressions used in Italian to convey modal information; and, second, we could compare how modality is expressed in Portuguese to how modality is expressed in Italian, to see if the two languages use the same expressions to convey the same modal meaning.

Further on, there is still work to do as concerns the component *Polarity*. As we have seen, with the current annotation scheme, we only annotate the polarity of the modal value, but in some sentences the polarity of the modal value is different from the polarity of the whole sentence (Saurí et al., 2006). Therefore, it would be interesting to create a further component in the scheme to annotate the overall polarity of the clause. In this case, there would be a component *Polarity of the modal value* and another component *Polarity of the clause*. In sentence (320), the polarity of the modal value *participant-internal necessity* conveyed by *é necessário* 'it is necessary' in the subordinate clause is positive, but the polarity of the whole subordinate clause is negative because of the negation of the verb in the target of the modality trigger (*não se esforçar* 'not to push').

(320) *Para acabar uma prova de natação sem dores, nem cansaço, é necessário não se esforçar muito no dia anterior.*

'In order to end a swimming competition with no pain or fatigue, it is necessary not to push too much the day before'.

In the Results chapter, we also have pointed out how our annotation scheme is not able to catch ambiguities at the structural level. In fact, sentence (321) can have two different interpretations and each interpretation implies a different annotation: we can consider as target only the expression *ganhar com justiça* 'win with justice', in which case the expression *o campeão do mundo* 'the world's champion' is tagged as source of the modality and the modal value is participant-internal necessity (annotation shown in the left column); or we can consider as a discontinuous target both the expressions *o campeão do mundo* 'the world's champion' and *ganhar com justiça* 'win with justice', in which case the only source is the speaker or writer and the modal value is deontic obligation (annotation shown in the right column).

(321) *Até o campeão do Mundo tem que ganhar com justiça.*
'Even the World's Champion must win with justice'.

Trigger: tem que Target: ganhar com justiça Source of the event mention: sp/wr Source of the modality: o campeão do mundo Modal value: participant-internal_necessity Polarity: positive	Trigger: tem que Target: o campeão do mundo@ganhar com justiça Source of the event mention: sp/wr Source of the modality: sp/wr Modal value: deontic_obligation Polarity: positive

Moreover, from the Inter-annotator Agreement (IAA) computed on the annotation of the same 50 sentences done by two different linguists, we have understood that we should better specify what to consider as Trigger for modality, since, as we have seen, the Kappa value for the Trigger (.65) was quite lower than the one for the Modal value (.85). This means that a further analysis of modal expressions in Portuguese is needed in order to precisely define which expressions have to be considered as Trigger and which are their boundaries.

Another topic that still has to be investigated in a deeper way is the use of nouns, adjectives and adverbs to convey modality, to see if the annotation scheme can be applied to these elements too, to analyse if they convey the same modal values as verbal triggers and especially to observe how they combine with other linguistic elements in the sentence when carrying some modal meaning.

As we can see, the investigation in the annotation of the linguistic expression of modality is still in an initial phase and we really hope that the interest in this topic keeps on growing in order to provide this area of Linguistics with more material on the modality.

References

Reference Corpus of Contemporary Portuguese (CRPC) of the Centro de Linguística da Universidade de Lisboa - CLUL (version 2.0, 2010, using CQPWeb in the period [February/2011 – November/2011]).

BACELAR DO NASCIMENTO, M. F. (2000), *Corpus de Référence du Portugais Contemporain*, in BILGER, M. (ed.) Corpus, Méthodologie et Applications Linguistiques, Paris, H. Champion et Presses Universitaires de Perpignan (2000), pp. 25-30.

BAKER, K., BLOODGOOD, M., DORR, B. J., FILARDO, N. W., LEVIN, L., PIATKO, C., (2010), *A Modality Lexicon and its use in Automatic Tagging*, in Proceedings of the Seventh Language Resources and Evaluation Conference (LREC'10).

BYBEE, J., PERKINS, R., PAGLIUCA, W., (1994), *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

COSTA CAMPOS, M. H., (1989), *Abordagem Enunciativa de um subsistema do sistema modal do português: os verbos DEVER e PODER*, Tese de Doutoramento apresentada na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.

COSTA CAMPOS, M. H., (1997), *Tempo, Aspecto e Modalidade. Estudos de Linguística Portuguesa*, Porto editora.

GÉNÉREUX, M., HENDRICKX, I., MENDES, A., *Introducing the Reference Corpus of Contemporary Portuguese*, Proceedings of LREC 2012, to appear.

HUDDLESTON, R., PULLUM, G. K., (2002), *Mood and modality*, in The Cambridge Grammar of the English Language, Cambridge University Press, pp. 172-208.

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

MATTHEWS, G. H., (1965), *Hidatsa syntax*, The Hague: Mouton, pp. 99 – 100.

MC ENERY, T., WILSON, A., (1996), *Corpus Linguistics*, Edinburgh University Press.

OLIVEIRA, F., (1988), *Para uma semântica e pragmática de DEVER e PODER*, Dissertação de Doutoramento em Linguística Portuguesa apresentada à Universidade do Porto, Faculdade de Letras.

OLIVEIRA, F., (1990), *Sobre Condicionais*, in Actas do 6º Encontro da APL*,* Porto, pp.239-258.

OLIVEIRA, F., (1993), *Questões sobre Modalidade em Português*, in Cadernos de Semântica nº 15, FLUL.

OLIVEIRA, F., (2003), *Modalidade e modo*, in MIRA MATEUS, M. H., BRITO, A. M., DUARTE, I., HUB FARIA, I., and FROTA, S., MATOS, G., OLIVEIRA, F., VIGÁRIO, M., VILLALVA, A., Gramática da Língua Portuguesa, Lisboa, Editorial Caminho, pp. 243-272.

PALMER, F. R., (1986), *Mood and Modality*, Cambridge textbooks in linguistics, Cambridge University Press, Cambridge.

SAURÍ, R., VERHAGEN, M., PUSTEJOVSKY, J., (2006), *Annotating and Recognizing Event Modality in Text*, in Proceedings of the 19th International FLAIRS Conference, FLAIRS 2006.

SAURÍ, R., PUSTEJOVSKY, J., (2007), *Determining Modality and Factuality for Text Entailment*, Internationa Conference on Semantic Computing, ICSC 2007.

SEARLE, J., (1979), *Expression and Meaning*, Cambridge University Press.

SEARLE, J., (1983), *Intentionality: An essay in the philosophy of mind* (Vol. 9), Cambridge, England: Cambridge University Press.

SZARVAS, G., VINCZE, V., FARKAS, R., CSIRIK, J., (2008), *The BioScope corpus: annotation for negation, uncertainty, and their scope in biomedical texts*, in Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, Association for Computational Linguistics, pp. 38-45.

VAN DER AUWERA, J., PLUNGIAN, V., (1998), *Modality's semantic map*, in Linguistic Typology 2, pp. 79-124.

VAN DER AUWERA, J., DOBRUSHINA, N., GOUSSEV, V., (2004), *A semantic map for Imperative-Hortatives*, in WILLEMS, D., COLLEMAN, T., DEFRANCQ, B., NOËl, D., *Contrastive Analysis in Language, Identifying Linguistic Units of Comparison*, Palgrave Macmillan, Basingstoke, pp. 1-21.

WIEBE, J., WILSON, T., CARDIE, C., (2005), *Annotating Expressions of Opinions and Emotions in Language*, in Kluwer Academic Publishers, pp. 1-54.