Annotating the Interaction between Focus and Modality: the case of exclusive particles

Amália Mendes[*], Iris Hendrickx[*†], Agostinho Salgueiro[*], Luciana Ávila[*‡]

* Centro de Linguística da Universidade de Lisboa
† Center for Language Studies, Radboud University Nijmegen
‡ PosLin-Universidade Federal de Minas Gerais / Capes
{amalia.mendes,iris}@clul.ul.pt, agostinhomms@gmail.com,
lucianabeatrizavila@gmail.com

Abstract

We discuss in this paper a proposal to integrate the annotation of contexts with focus-sensitive expressions (namely the Portuguese exclusive adverb *só* 'only') in a modality scheme. We describe some properties of contexts involving both exclusive particles and modal triggers and discuss how to integrate this with an existing annotation scheme implemented for European Portuguese. We present the results of the application of this annotation scheme to a sample of 100 sentences.

1 Introduction

Modality in language has been studied extensively (see Portner (2009) for an overview). In recent years, the study of modality has been associated with a trend in Information Extraction applications that aim to identify personal opinions in sentiment analysis and opinion mining (Wiebe et al., 2005), to identify events which are factual, probable or uncertain, as well as speculation and negation. This trend has lead to the development of several practical annotation schemes for modality (Sauri et al., 2006; Szarvas et al., 2008; Baker et al., 2010, Matsuyoshi et al., 2010).

Most of these annotation schemes focus on the annotation of modal elements like modal verbs or adverbs, but in the present study we go one step deeper and discuss the complex interaction between modality and focus in Portuguese. Our notion of modality focuses on the expression of the opinion and attitude of the speaker or the agent towards the proposition (Palmer, 1986). This attitude or opinion towards a state or event can assume diverse values. For example, the speaker (or subject) may consider something to be possible, probable or certain (epistemic mo-

dality), he might be obliged or allowed to do it (deontic modality), or he wants or fears it (volitive modality). Frequently, several modal expressions interact to compose the overall modal meaning of the sentence. Non modal elements can also directly influence the modality type and alter the meaning of the sentence. One such element, rather well studied, is the negation marker (Morante, 2010; Morante and Sporleder, 2012). In this paper however we discuss the element *focus*, taken as a means to "give prominence to meaning-bearing elements in an expression." (Krifka, 1995:240). The prominent constituent is called the *focus*, while the complement notion is called the *background*. We are especially concerned with the interaction between modality and a subtype of focus-sensitive expressions named exclusive particles (Beaver and Clark, 2008), and, for the purposes of this paper, we will center our discussion on the adverb *só* 'only'.

Our goal is to study closely how exclusive particles affect and alter the modal meaning of the sentence. By performing a systematic annotation of these interactions in examples drawn from a large corpus we better comprehend the role that these particles play and the different type of effects that exclusive particles can have.

Most annotation schemes for modality focus on English but resources are now being developed for other languages including Portuguese. Hendrickx et al. (2012b) have previously developed an annotation scheme for European Portuguese and applied it to a corpus of 2000 sentences. Ávila & Mello (2013) presented an annotation scheme for Brazilian Portuguese speech data, applied to information units. Here we look at the interaction between focus-sensitive adverbs and modality and discuss how to integrate these findings in the annotation scheme of Hendrickx et al. (2012a).

The structure of the paper is as follows. In section 2, we review related work in the field of modality, its annotation in texts, focus-sensitive expressions and the semantics of the exclusives. The discussion of the specific contexts with adverb *só* and a modality trigger is presented in section 3. We analyze the interaction between triggers and this specific adverb, focusing on the scope of the adverb and its influence over the modal value of the sentence. In section 4.1, we briefly summarize the annotation scheme for modality in Portuguese developed by Hendrickx et al. (2012a). We then demonstrate the implementation of our findings about adverb *só* in this annotation scheme in section 4.2. We discuss the results of the annotation of a sample of 100 sentences in 4.3 and conclude in section 5.

2   Related work

The literature on modality proposes different typologies. In linguistics, most modal systems are based on the contrast between epistemic and deontic modality. While the epistemic value is stable across typologies, the other values that are contrasted with it vary considerably. Some proposals distinguish between epistemic, participant-internal and participant-external modality (Van der Auwera and Plungian, 1998), or between epistemic, speaker-oriented modality and agent-oriented modality (Bybee et al., 1994). Other values generally considered are, for example, volition, related to the notions of will, hope and wish; evaluation, concerning the speaker's evaluation of facts and situations; and commissives, used by the speaker to express his commitment to make something happen (Palmer, 1986). Although most of the literature is centered on verbal expressions of modality (mostly semi-auxiliary verbs like *may*, *should*, *can*), studies on adverbs and modality have also been carried out for English (cf. Hoye, 1997).

In the literature on practical corpus annotation of modality, the attention focuses on the distinction between factual and non-factual information, as many NLP applications need to know what is presented as factual and certain and what is presented as non-factual or probable. Opposed to the theoretical typologies of modality, these schemes describe in detail which elements in the text are actually involved in the expression of modality and their roles. These are the subject of the modality (source) and the elements in its scope (target/scope/focus). Other schemes (Baker et al., 2010; Matsuyoshi et al., 2010; Sauri et al., 2006) also determine the relation between sentences in text, identifying temporal and conditional relations between events or the evaluation of the degree of relevance of some information within a text, rather than classifying modal values.

Rooth (1992) claims that the effects of focus on semantics can be said to be the introduction of a set of alternatives that contrasts with the ordinary semantic meaning of a sentence and that there are lexical items and construction specific-rules that refer directly to the notion of focus. The phenomenon of focalization is taken to be a grammatical feature that semantically conveys (i) newness/information update; (ii) answering the 'current question'; (iii) contrast; (iv) invocation of alternatives. In terms of semantic annotation, Matsuyoshi et al. (2010) propose an annotation scheme for representing extended modality of event mentions in Japanese. This scheme includes seven components among them the Focus, which represents the focus of negation, inference or interrogative.

There are no works on the annotation of focus and its relation with modality in Portuguese, in any of its variants. This is an attempt to put the two notions together and propose a scheme that describes the scope of exclusive particles and its impact on the meaning of the expressed modality.

3   Interaction between adverbs and modal value

In this section, we discuss in detail the possible interactions between the adverb *só* and modal expressions in the text. We extracted our examples from the online search platform of the Corpus de Referência do Português Contemporâneo (CRPC)[1], a highly diverse corpus of 312 million words covering a large variety of textual genres and Portuguese varieties (Généreux et al., 2012).

The adverb *só* is considered a focus-sensitive particle (Beaver & Clark, 2008; Aloni et al., 1999), defined as a word which semantics "involves essential reference to the information structure of the sentence containing it" (Aloni et al., 1999:1).

The meaning of *only* consists of asserting that no proposition from the set of relevant contrasts other than the one expressed is true (von Fintel, 1994). The standard views on exclusive particles consider that "the position of focal accent identi-

---

[1]http://alfclul.clul.ul.pt/CQPweb/crpcweb23/index.php

fies the constituent associated with *only*" (Dryer, 1994:2). Dryer (1994) and Schwarzschild (1997), on the other hand, assume that general principles of discourse could explain focus-sensitivity.

Exclusives can also downtone, by underlining the fact that this proposition "is not the strongest that in principle might have been the case", a function called Mirative (Beaver & Clark 2008: 250).

Constructions with exclusives involve a positive and a negative component: the positive is called the *prejacent* and, in sentence (1a), it is equivalent to 'he wants to go home'; the negative is called the *universal* and corresponds in (1a) to 'he does not want to do anything else'.

(1) a. Ele *só quer* ir para casa.
         'He only wants to go home.'

     b. As actividades de campanha eleitoral *só podem* ser financiadas por subvenção estatal.
         'The activities of the election campaign can only be financed by public funding.'

We will discuss in the following subsections some properties of constructions with exclusives and modal triggers.

3.1 Scope of the exclusive particle

Exclusives give prominence to a constituent in the sentence, called the focus. In sentence (1a), *só* has scope over the modal trigger (*quer* 'wants') and its target (*ir para casa* 'to go home'). The scope of *só* can also be a smaller constituent inside the target. In (1b), *só* has scope over the by-phrase (*por subvenção estatal* 'by public funding'), and in (2), over a temporal adjunct. The adverb *só* could also occur immediately before the temporal adjunct keeping the same focus reading as in (2) (*só depois de construído o novo palácio da justiça*).

(2) O presidente respondeu que tal só deverá acontecer *depois de construído o novo palácio da justiça*.
     'The president answered that it should/can only happen after the new courthouse is built.'

Two other possibilities are illustrated in (3) and (4): in (3), the focus is the subject *tu* 'you' and in (4) is the quantifier *7* 'seven':

(3) Só *tu* eras capaz de fazer juntar tanta gente.
     'Only you could bring together so many people.'

(4) Claro que só podem estar *7* jogadores em campo. 'Obviously there can only be 7 players in the field.'

As these examples show, the exclusive particle is not necessarily contiguous to its focus. The analysis of a sample of the occurrences of the exclusive *só* with a modal trigger in the CRPC corpus shows that in most cases *só* has scope over a specific constituent, rather than over the full target of the modal trigger.

There can be ambiguity in the scope of the exclusive particle *só* in certain contexts. This is the case when the focus can be interpreted as the full target of the modal trigger or as a specific constituent inside the target. We illustrate such cases in sentence (5): this sentence can be interpreted as 'the only thing I'm capable of doing is to ask a metaphysical question' or 'the only question I'm capable of asking is a metaphysical one'.

(5) Só sou capaz de colocar ao Sr. Ministro uma questão metafísica.
     'I'm only capable of asking a metaphysical question to Mr. Minister.'

However, in most contexts, there seems to be one preferential interpretation, in spite of the underlying ambiguity.

3.2 From possibility to necessity

In contexts where the verb *poder* has an epistemic reading, the exclusive can restrict the set of possibilities to the one presented in the sentence (x and only x), as illustrated in (6).

(6) Isto só pode ter sido um acidente.
     'This can only have been an accident'

By restricting the set of possible situations to one, the adverb leads to an overall reading of the sentence as expressing epistemic necessity. Sentence (6) has indeed an equivalent modal meaning to (7) and to a double negative polarity (over the modal verb and over its target), as in (8).

(7) Isto tem de ter sido um acidente.
     'It had to be an accident'

(8) Isto não pode não ter sido um acidente.
     'It could not not have been an accident'

The scope of the focus-sensitive particle plays an important role on whether an epistemic trigger may be interpreted as having a necessity reading or not (cf. 3.2). Contrary to (6), the interpretation of (9) is one of epistemic possibility, although the particle *só* is present. In this case, the particle has scope over a specific constituent, the tem-

poral adjunct, which establishes a condition over the modal trigger. But in (9b), without the temporal adjunct, the scope of *só* is coincident with the target of the modal trigger and the interpretation is one of necessity, as paraphrased by 'the member of parliament MFL has to be right'.

(9) a. Ora, a Sr.ª Deputada MFL *só pode* ter razão *quando acertar nalguma previsão*.
'Well, the member of parliament MFL may only be right when at least one of her forecasts turn out correct.'

 b. Ora, a Sr.ª Deputada MFL *só pode* ter razão.
'Well, the member of parliament MFL can only be right.'

This restriction holds for stative targets, as in (9a), and also for eventive targets, although in this case the possibility or necessity reading is also determined by the tense of the verbal predicate. The necessity reading is only associated to eventive targets temporally located in the past, and is not available, for example, in the sentence *ele só pode ir ao cinema* 'he may only go to the cinema', where the target is temporally located in the future.

It seems that when the target of the modal trigger is a state or a past event, the exclusive particle leads to a necessity reading instead of a possibility reading, as long as the scope of *só* is the full target of the modal trigger and not a different constituent. However, we need to assess these factors against more corpus data.

If we compare (6) with a related declarative sentence (cf. *isto foi um acidente* 'it was an accident'), we see that the declarative has an assertive value over the situation it denotes, while (6) establishes a set of possibilities and strongly asserts a single one, in what is considered by Moreira (2005) as a case of overmodalization.

The verb *dever* 'have to' can occur in contexts similar to (6), as exemplified in (10). However, the sentence with *dever* expresses the most probable event and does not entail an epistemic necessity reading, contrary to (6) with verb *poder*.

(10) Isso só deve ter sido um acidente.
'This was probably only an accident'

In (10), the sentence merely states that this is probably what happened. The fact that the necessity reading does not arise from (10), contrary to (6), follows from the differences that exist between the possibility and the probability reading. The possibility reading in (6) denotes one partic-ular event out of a set of possible ones and (6) singles out one possibility as the only valid one, affecting the truth-value of the set of alternatives considered. Sentence (10) denotes that this particular event is more probable to be true than other alternatives. So, in this sentence the exclusive strengthens the value of this probability but does not establish it as a single one. It is consequently a scalar use of the exclusive, in the sense that there is an ordering of propositions from weaker to stronger.

3.3 Contexts where *só* is required

Contrary to sentences (6), where the adverb *só* can be present or not (with effects on the interpretation), in sentences like (11), with the same modal verb, the adverb is required.

(11) Sr. Deputado, só pode estar a brincar!
'Congressman, you must be kidding!'

These are discursive contexts with a distinctive prosody consisting of a rising tone, marked in writing by the punctuation. The modal interpretation of (11) is one of epistemic necessity, as in (6). However, the equivalent sentence without *só* is not acceptable (**Sr. Deputado, pode estar a brincar*! 'Congressman, you can be kidding'). Contrary to (6), the speaker does not consider that a set of possibilities exist, from which one is singled out, but rather takes only into consideration the situation that the sentence denotes (to be kidding) and emphasizes it.

3.4 *Só* in ambiguous modal contexts

The presence of *só* can resolve ambiguity at the modal value level. For example, sentence (11a) might be interpreted as expressing a possibility or an internal capacity of the law itself. However, in sentence (12b), the presence of the adverb *só* blocks the participant internal reading. Sentence (12b) does not mean that this is the only property of the law but rather that it is inevitable that it reduces injustice. It has the same necessity value as (6).

(12) a. A nova lei pode reduzir a injustiça.
'The new legislation can reduce injustice'

 b. A nova lei só pode reduzir a injustiça.
'The new legislation can only reduce injustice'

3.5 Weak alternative

Besides highlighting one alternative, the exclusive particle can also mark this alternative as

weaker than expected. This is frequently the case with deontic modality, as illustrated in (13): the process to participate is presented as surprisingly easy.

(13) Para participar só tem de contactar a organização através dos telefones 96... ou 91...
'To participate, you only have to contact the organization through the phone numbers…'

3.6 Contrastive value

The epistemic subvalues belief and knowledge are expressed by main verbs like *achar* 'to believe' and *saber* 'to know'. When the adverb *só* occurs in these contexts, it has mainly a discursive function: it establishes a contrast with something that was previously said in the conversation. We exemplify such conversational contexts in (14):

(14) A: Eu não acho que ele seja corrupto.
'I don't think he is corrupted'

   B: Eu *só sei* que ele fez grandes depósitos em offshores.
'I only know that he made big deposits in offshores'

The different contexts discussed in this section show that the interpretation of *só* with modal trigger is complex and varies according to the lexical trigger and its value, but also to the linguistic context and to pragmatic factors.

4 Corpus Annotation

In this section, we first report on the annotation scheme previously implemented for Portuguese, in 4.1. We then discuss how to integrate our findings regarding the adverb *só* 'only', in 4.2, and report on the results of the annotation of a sample corpus in 4.3.

4.1 Annotation scheme for Portuguese

The annotation scheme for Portuguese presented in Hendrickx et al. (2012a) follow a theoretically-oriented perspective, but also addresses certain modal values that are important for practical applications in Information Extraction. The annotation is not restricted to modal verbs and instead covers several parts of speech with modal value: nouns, adjectives and adverbs. Tense, however, is not included, although it has an important part in the modal interpretation of sentences. Also, only modal events are annotated, not entities. The approach is very similar to the approach taken in the OntoSem (Mcshane et al.,

2005) annotation scheme for modality (Nirenburg and McShane, 2008).

Seven main modal values are considered (epistemic, deontic, participant-internal, volition, evaluation, effort and success), and several sub-values, based on the modality litterature, but also on studies focused on corpus annotation and information extraction (e.g. (Palmer, 1986; van der Auwera and Plungian, 1998; Baker et al., 2010)). There are five sub-values for epistemic modality: knowledge, belief, doubt, possibility and interrogative. Contexts traditionally considered of the modal type "evidentials" (i.e, supported by evidence) are annotated as epistemic belief. Two subvalues are identified for deontic modality: deontic obligation and deontic permission (this includes what is sometimes considered participant-external modality, as in van der Auwera and Plungian (1998)). Participant-internal modality is subdivided into necessity and capacity. Four other values are included: evaluation, volition and, following Baker et al. (2010), effort and success.

The annotation scheme comprises several components to be tagged: (a) the trigger, i.e, the lexical element conveying the modal value – we choose to tag the smallest possible unit (noun, verb, etc.); (b) the target, expressed typically by a clause and tagged maximally to include all relevant parts; (c) the source of the event mention (speaker or writer) and (d) the source of the modality (agent or experiencer), to distinguish between the person who is producing the sentence with modal value and the person who is 'undergoing' the modality. The trigger receives two attributes: Modal value (selection out of 13 possible values); and Polarity (positive or negative). The polarity attribute regards the value of the trigger and not of the full sentence.

This scheme has been applied to the manual annotation of a corpus sample of approximately 2000 sentences using the MMAX2 annotation software tool[2] (Müller and Strube, 2006). Sentences were extracted from the online search platform of written corpus CRPC.

4.2 Annotation of contexts with the adverb *só*

We will discuss in this subsection how to integrate our findings regarding the exclusive adverb *só* in modal contexts into an annotation scheme. For this purpose, we revised the modality scheme of Hendrickx et al. (2012b) to address the annotation of focus-sensitive particles in

---

[2] http://mmax2.sourceforge.net/

modal contexts. Instead of considering focus as an independent scheme, we treat it inside modality, inspired by the approach taken regarding polarity. The existence of a focus-sensitive particle is marked with an attribute of the trigger called "focus". This attribute has, for now, three possible values: none, exclusive, additive (for particles like *também* 'also'). The list can be enlarged in the future to address other categories of focus-sensitive particles. The focus particle does not typically have scope over the modal trigger, but rather over other components of the modal scheme (like the target or the source of modality). However, we decided to mark focus information in the trigger component, inspired by the approach of Miwa et al. (2012), since we are considering it as the main element that subsumes the total information regarding the modal event.

The component "focus cue" was added to the modal scheme to identify the focus-sensitive particle in the text. The scope of the focus-sensitive particle is an important aspect to consider in the annotation (cf. 3.1) and we decided to mark the scope of the particle with an extra component named "focus scope". The "focus cue" and the "focus scope" markables are linked to the trigger and, consequently, to the modal event. We illustrate in (15) the focus scope component of the annotation, as well as the features in the trigger component that are associated to the focus-sensitive particle.

(15) Há quem defenda que os medicamentos *só devem* ser usados *numa primeira fase do tratamento*.
'Some people argue that medical drugs should only be used in the first stage of the treatment.'

Trigger: devem
    Modal value: deontic_obligation
    Focus: exclusive
Focus cue: só
Focus scope: numa primeira fase do tratamento

There may be ambiguity regarding the scope of the focus particle (cf. 3.4) and a feature "ambiguity" is attributed to the focus scope component to deal with such cases. We mark the scope constituent according to the most natural interpretation and fill in the ambiguity feature if more than one interpretation is possible, as illustrated in (16).

(16) Portanto, *só temos de* votar a proposta 525-C, do PSD.
'So, we only have to vote proposal 525-C, of PSD.'

Trigger: temos de
    Modal value: deontic_obligation
    Focus: exclusive
Focus cue: só
Focus scope: a proposta 525-C, do PSD
    Ambiguity: votar a proposta 525-C, do PSD

When there are two consecutive modal triggers, we only give information on the focus-sensitive particle in the annotation of the first trigger. For example, in (17), the first trigger (*deverá*) is annotated with features "focus" and "focus cue", and the modal set includes the "focus scope" component. The second modal trigger (*poder*) is part of the target component of the first trigger and is consequently under the scope of its focus related features.

(17) O plantel do Estrela da Amadora *só deverá poder* voltar a contar com o guardião Tiago durante a próxima semana.
'The team of Estrela da Amadora shall only be able to count again on the goalkeeper Tiago during next week.'

In what concerns the necessity reading with *poder*, we believe that the regularities that we discussed in 3.2 allow us to recover the adequate modal value without the need of any special feature but the annotation discussed in the next section will prove if this is indeed the case or if a special feature has to be devised to handle these cases.

The non-optional nature of *só* in contexts as the one illustrated in (10) can be dealt with by selecting both *só* and the modal verb as a composite trigger. This solution would handle the fact that *só* is required in these contexts and would help identifying constructions which have a specific prosodic pattern. The modal value epistemic_necessity would be, in this case, attributed to both elements. We do not propose this solution, however, for cases like (6) and (18b) because *só* is optional in those contexts and the necessity reading follows from the compositional nature of the exclusive, the modal trigger and the target.

In contexts like (13), the exclusive singles out one alternative and also comments on the fact that it is weaker than expected (for example, easier in (13)). However, there is no change in the modal value and the annotation scheme can be

applied. To cope with these cases, we added the attribute "focus value" in the trigger component, and consider for now 3 possible values: none, mirative (Beaver and Clark, 2008) and contrastive (as in sentence (14)).

4.3 Results of the annotation

This scheme has been applied, using MMAX2 software, to the manual annotation of a corpus sample of 100 sentences extracted from the online search platform of written corpus CRPC. The 100 sentences all contain the focus particle *só* in the context of a verbal modal trigger, and are not syntactically annotated. We considered, for this purpose, 5 modal verbs: *poder* 'can/may', *dever* 'must', *ter de* 'have to', *ser capaz de* 'be able to', *querer* 'want', most of them covering more than one modal value. We selected a higher number of sentences with *poder*, *dever* and *ter de* because these modal verbs have proved to be more complex and would therefore provide a good test for our annotation scheme. The sentences were selected from a randomly ordered list, and cover different text types. Table 1 presents the distribution of modal values in our sample, taking into consideration only modal events that include the focus-sensitive particle *só*: we observed that deontic obligation and epistemic possibility are the most frequent values.

Modal value	Freq.
Deontic obligation	37
Epistemic possibility	30
Participant-internal capacity	15
Volition	13
Deontic permission	5
Total	100

Table 1: Frequency information about the modal values encountered in the corpus sample.

All ambiguous cases regarding modal value involve the verb *poder* 'can/may', which can denote readings of deontic permission, epistemic possibility and participant-internal capacity. The other four modal verbs are never marked as having more than one modal reading in the context. The most frequent ambiguity in this sample involves the two modal values: epistemic possibility and deontic permission. In three cases, the annotator marked the trigger as having a deontic permission reading, and considered it ambiguous (ambiguity feature of the trigger) with an epis-

temic possibility value. In two other contexts, the opposite choice was made: epistemic possibility was the marked value and deontic permission was annotated as a possible alternative value. The other four cases of modal ambiguity involve epistemic possibility and participant-internal capacity: in two cases, the former was selected as the most salient value, while the opposite choice was made in the other two cases.

The most frequent constituents in the scope of the exclusive in our annotation are temporal adjuncts, with a total of 27 cases. The verb *dever* stands out, with 14 occurrences, out of the total of 27. The second most frequent type of focus (freq. 23) corresponds to cases where the exclusive has scope over the whole target of the modal trigger. The two most frequent verbs with this type of focus are *querer* and *ter de* (in fact, all but two occurrences of *querer* are of this kind). The two most paradigmatic modal verbs, *poder* and *dever*, never occur with the whole target as focus, but rather favour cases where *só* has scope over different constituents of the sentence. There is a large set of possible constituents which receive the focus of the exclusives: subjects (7), objects (9), quantifiers (5), predicative adjectives (1), by-phrases in passive constructions (3), temporal adjuncts (27), locative adjuncts (1), conditional clauses (5), prepositional phrases (7), and adverbial phrases (3). While *dever* shows a preference for the construction with a temporal adjunct, the verb *poder* presents low frequencies of a large set of these possibilities.

There are 5 cases of ambiguity in the scope of the focus particle, 2 with *ter de* 'have to', 2 with *ser capaz de* 'to be able to' and one with *poder* 'can'. This is perhaps a surprisingly low number compared to our comments in subsection 3.1. Although focus scope is potentially extremely ambiguous, it turns out that the linguistic context seems to lead to one specific interpretation regarding the constituent under focus. In the 5 ambiguous cases, the scope of the focus particle can be understood as a specific constituent included in the target of the modal trigger, or it can be the whole event denoted by the target.

The interpretation of the exclusive and the verb *poder* as a case of epistemic necessity occurs a single time in our annotation, with a stative target in the future tense. No case of contrastive value (cf. (14)) was encountered, but this is certainly due to the fact that we annotated single sentences out of context, and also to the fact that we didn't select knowledge verbs, which typically allow this value. Also, no case of

non-optional exclusive particle was found in our set of sentences. We did, however, identify 7 contexts with *ter de* and 3 contexts with *querer,* which denote a weaker alternative than expected and were marked with the value "mirative".

Overall, the proposed solution for handling the complex annotation of the interaction between exclusive particles and modality captured all cases we encountered in our small sample of 100 sentences. However, the annotation of more data is required to evaluate if our modal scheme can deal with the discursive values assumed by the exclusive in certain contexts.

Two different human annotators performed the annotation of a subset of 50 sentences with *só* independently. We conducted a small study to measure the inter-annotator agreement (IAA) for this annotation task. Such a study gives us information about the feasibility of the annotation scheme and about the level of detail of our guidelines. We computed IAA using the kappa-statistic (Cohen, 1960) for each field in the annotation[3]. The trigger achieved a kappa value of 0.85, while the modal value attained a value of 0.83. Although the task involved a higher level of complexity due to the annotation of both modal and focus information, the results are in line with the ones reported in Hendrickx et al. (2012b) and, for English, in Matsuyoshi et al. (2010). We also measured the kappa value for the target component, which attained 0.64. For the focus scope an inter-annotator agreement of 0.63 is achieved. These are lower scores than the ones achieved for trigger and modal value, which is due to small differences in the delimitation of the constituents between the two annotators (for example, the inclusion or not of an adjunct).

5   Conclusion

In this paper, we have presented a detailed analysis of the interaction between the exclusive *só* and modal expressions occurring in texts.

As Portner puts it "It seems that modality is not something that one simply observes, but rather something that one discovers, perhaps only after careful work." (Portner, 2009:1) and this is what we have attempted in this study.

We presented the extension of a modality scheme developed for Portuguese to account for focus-sensitive particles in modal contexts and our experience in annotating a sample of 100 sentences with this extended scheme. Data show

that this is a complex issue that needs to consider the modal value, the linguistic context and each modal trigger. The annotation confirms the dual nature of exclusives, due to the fact that in certain contexts they both signal one of the possible alternatives and describe it as weaker that would be expected by the participants. The scope of the focus particle plays an important role in the meaning of the sentence since it adds a condition to the modal value and can affect the global meaning of the sentence. Discursive aspects have also to be taken into consideration and evaluated against our annotation scheme.

As a next step we aim to study a context larger than the sentence for the annotation of the interaction between modals and exclusives. We plan to proceed with the analysis of the interaction of *só* in a larger number of modal contexts and also to enlarge the analysis to other adverbs of the same type, like *apenas* and *unicamente*. Another objective is to explore the combined effects of polarity, modality and this type of adverbs, and to later contrast the results with other Romance languages, as well as English.

References

Maria Aloni, David Beaver, and Brady Clark. 1999. Focus and Topic Sensitive Operators. In Paul Dekker (ed.), *Proceedings of the Twelfth Amsterdam Colloquium*, ILLC, University of Amsterdam, Amsterdam, The Netherlands.

Luciana Beatriz Ávila, and Heliana Mello. 2013. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In *Proceedings of WAMM-IWCS2013*, Potsdam, Germany.

Kathrin Baker, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine Piatko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of LREC'10*, Valletta, Malta. ELRA, 1402-1407.

Kathryn Baker, Bonnie Dorr, Michael Bloodgood, Chris Callison-Burch, Nathaniel Filardo, Christine Piatko, Lori Levin, and Scott Miller. 2012. Use of Modality and Negation in Semantically-Informed

---

[3] Note that we are very strict in the computation, only full string matches are counted as agreement.

Syntactic MT. *Computational Linguistics*. 38(2): 411-438

David Beaver and Brady Z. Clark. 2008. *Sense and Sensitivity. How Focus Determines Meaning*. Wyley-Blackwell, Oxford, UK.

Joan L. Bybee, Revere Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. University of Chicago Press, Chicago, USA.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20:37-46.

Philip W. Davis. 2009. The constitution of focus. Available at: http://www.philipwdavis.com/sands03.pdf. Accessed: March 28, 2013.

Matthew S. Dryer. 1994. The pragmatics of association with only. Paper presented at the 1994 Winter Meeting of the L.S.A. Boston, Massachussets, USA.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The conll-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, ACL, 1-12.

Michel Généreux, Iris Hendrickx, and Amália Mendes. 2012. Introducing the Reference Corpus of Contemporary Portuguese On-Line". In *Proceedings of the Eighth International Conference on Language Resources and Evaluation* - LREC 2012, Istanbul, May 21-27, 2012, pp. 2237-2244.

Iris Hendrickx, Amália Mendes, Silvia Mencarelli, and Agostinho Salgueiro. 2012a. *Modality Annotation Manual*, version 1.0. Centro de Linguística da Universidade de Lisboa, Lisboa, Portugal.

Iris Hendrickx, Amália Mendes, and Silvia Mencarelli. 2012b. Modality in Text: a Proposal for Corpus Annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation* - LREC 2012, Istanbul, May 21-27, 2012, pp. 1805-1812.

Leo Hoye. 1997. *Adverbs and Modality in English*. Longman, London, UK.

Manfred Krifka. 1995. Focus and the Interpretation of Generic Sentences. In Gregory N. Carlson and Francis Jeffry Pelletier (eds). *The Generic Book*. The University of Chicago Press, Chicago, USA, 238-264.

Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. In *Proceedings of LREC'10*, Valletta, Malta. ELRA.

Marjorie McShane, Sergei Nirenburg, Stephen Beale, and Thomas O'Hara. 2005. Semantically rich human-aided machine annotation. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II*: Pie in the Sky, 68-75. ACL.

Makoto Miwa, Paul Thompson, John McNaught, Douglas B Kell and Sophia Ananiadou. 2012. Extracting semantically enriched events from biomedical literature. *BMC Bioinformatics* 13:108.

Roser Morante. 2010. Descriptive analysis of negation cues in biomedical texts. In *Proceedings of LREC'10*, Valletta, Malta. ELRA, 1429-1436.

Roser Morante and Caroline Sporleder. 2012. Modality and Negation: An Introduction to the Special Issue. *Computational Linguistics*, 38(2):223-260.

Benjamim Moreira. 2005. *Estudo de alguns marcadores enunciativos do português*. PhD Dissertation. Universidade de Santiago de Compostela, Faculdade de Filologia, Santiago de Compostela, Spain.

Christoph Müller and Michael Strube. 2006. Multilevel annotation of linguistic data with MMAX2. In *Corpus Technology and Language Pedagogy*: New Resources, New Tools, New Methods, 197-214. Peter Lang.

Sergei Nirenburg and Marjorie McShane. 2008. Annotating modality. Technical report, University of Maryland, Baltimore County, March 19, 2008.

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

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

Paul Portner. 2009. *Modality*. Oxford University Press, Oxford, UK.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75-116.

Roser Saurí, Marc Verhagen, and James Pustejovsky. 2006. Annotating and recognizing event modality in text. In *Proceedings of the 19th International FLAIRS Conference*.

Roger Schwarzschild. 1997. Why some foci must associate. Unpublished ms. Rutgers University.

György Szarvas, Veronika Vincze, Ricárd Farkas, and János Csirik. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. *BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, Columbus, Ohio, USA, 38-45.

Paul Thompson, Giulia Venturi, John Mcnaught, Simonetta Montemagni, and Sophia Ananiadou. 2008. Categorising modality in biomedical texts. In *Proceedings of the LREC 2008 Workshop on*

Building and Evaluating Resources for Biomedical Text Mining, Marrakech, Morocco, 27-34.

Johan Van der Auwera and Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology*, 2(1): 79-124.

Veronika Vincze, György Szarvas, Móra György, Tomoko Ohta, and Richárd Farkas. 2011. Linguistic scope-based and biological event-based speculation and negation annotations in the bioscope and genia event corpora. *Journal of Biomedical Semantics*, 2(5).

Kai-Uwe Von Fintel, 1994. *Restriction on Quantifier Domains*. UMass Amherst dissertation, Amherst, Massachussets, USA.

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