

Compilation and Annotation of the Reference Corpus of Contemporary Portuguese

Amália Mendes

Centro de Linguística da Universidade de Lisboa



Outline

- The CRPC design and compilation
- Lemmatization, POS tagging and online access
- Some resources based on the CRPC
- Modality annotation

Presentation of the CRPC

- The CRPC is a written and spoken corpus of 311 million words

<http://www.clul.ul.pt/en/research-teams/183-reference-corpus-of-contemporary-portuguese-crpc>

- Project started in 1988
- Written subcorpus: 309,8 M
- Spoken subcorpus: 1,6 M

Team: Maria Fernanda Bacelar do Nascimento, Amália Mendes, Michel Génèreux, Luísa Alice Santos Pereira, Rita Veloso, Sandra Antunes, Antónia Estrela

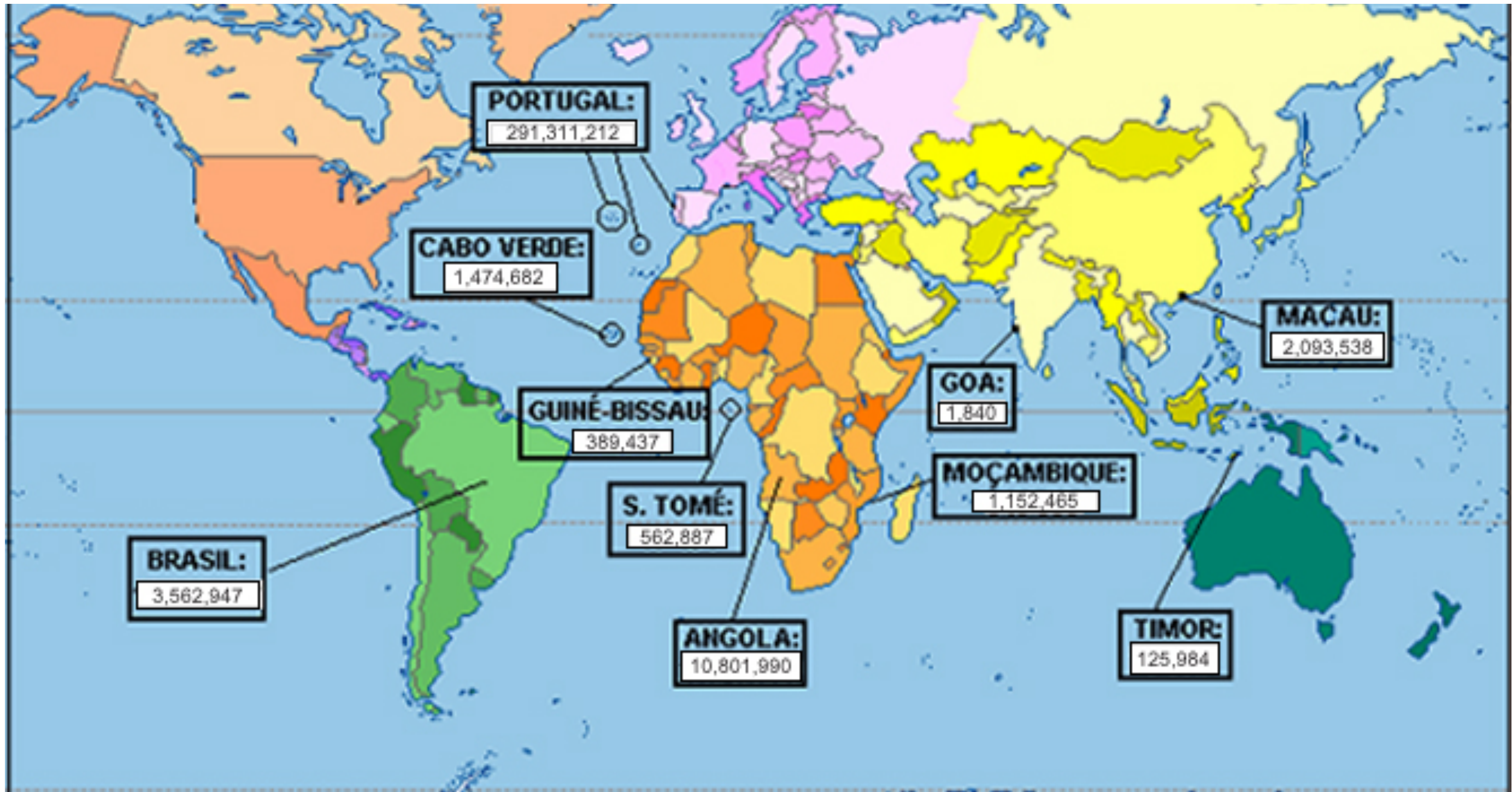
Balanced corpus vs. monitor corpus

- Initial goal: compilation of a balanced corpus, covering the diversity of genres in Portuguese
- The CRPC later evolved into a “monitor corpus” (“open corpus”)
- Constitution of subcorpora for specific goals:
 - The balanced corpus COMBINA (50 M) for MWE studies
 - The balanced corpus CORLEX (16 M) for the extraction of frequency information
 - Comparable corpora of Portuguese varieties in the world

Time Coverage

- Time coverage:
 - from the second half of the XIX century to 2008
 - most of the texts have been produced after 1970

Portuguese varieties



Compilation

- Inclusion of a wide set of genres: fiction, newspapers, technical and scientific texts, didactic texts, laws and other legal documents, Parliamentary speeches, letters, etc. (in preparation, a corpus of Computer Mediated Communication)
- Digitalization with OCR and manual revision
- Selection of samples from every text
- But also internet download (and file cleaning)
- Detailed metadata: 44 fields

Ex: variety of Portuguese, text type, country of publication, author's nationality

Written Corpus Design

Type	Texts	Tokens
Newspaper	50.8%	110,503,376
Politics	45.9%	163,267,089
Magazine	1.4%	7,581,850
Various	1.2%	4,806,176
Law	0.3%	2,927,953
Book	0.3%	20,557,296
Correspondence	0.03%	88,370
Brochure	0.01%	80,833
Total	100%	309,812,943

Spoken corpus design

- **Language register**

Informal vs. formal

- **Channel**

Face to face communication

Broadcasting

Telephone

- **Structure of the communication event**

Monologue / Dialogue / Conversation

- **Social context**

Family/private vs. Public

Annotation

- Tokenization: Lx-Tokenizer (Branco & Silva, 2004)
- Automatic annotation with 48 main PoS tags (+ 9 for contractions)
 - MBT tagger (Daelemans *et al.*, 1996)
 - Success rate: 95,5%
 - Next: automatic annotation with nominal and verbal inflection tags using a manually annotated training corpus
- Automatic lemmatization
 - MBLEM lemmatizer (Van den Bosch & Daelemans, 1999)
 - Success rate: 96,7%
- NP Chunking
 - YamCha chunker (Kudo and Matsumoto, 2003) trained on 1,000 random sentences

Access

- CRPC written subcorpus: CQPWeb (Hardie, A. 2012)
<http://alfclul.clul.ul.pt/CQPweb/>
- Spoken subcorpora with text-to-sound alignment and XML format (produced with the EXMARaLDA software) on the ELDA catalogue
 - C-ORAL-ROM
 - Fundamental Portuguese (free for research)
 - Spoken Portuguese – varieties of Portuguese in the world (free for research)

Menu

Corpus queries

Standard query

Restricted query

Word lookup

Frequency lists

Corpus info

View corpus metadata

Corpus documentation

About CQPweb

CQPweb main menu

CQPweb manual

Who did it?

Latest news

Reference Corpus of Contemporary Portuguese (CRPC) v2.3: powered by CQPweb

Restricted Query

Query mode:

Simple query (ignore case) ▾

[Simple query language syntax](#)

Number of hits per page:

50 ▾

Start Query

Reset Query

Select the text-type restrictions for your query:

Fonte	País do autor	
<input type="checkbox"/> correspond	<input type="checkbox"/> Angola	
<input type="checkbox"/> direito	<input type="checkbox"/> Brasil	
<input type="checkbox"/> folheto	<input type="checkbox"/> Cabo_Verde	
<input type="checkbox"/> jornal	<input type="checkbox"/> Guine_Bissau	
<input type="checkbox"/> livro_didactico	<input type="checkbox"/> Macau	
<input type="checkbox"/> livro_literario	<input type="checkbox"/> Mocambique	
<input type="checkbox"/> livro_tecnico	<input type="checkbox"/> Portugal	
<input type="checkbox"/> politica	<input type="checkbox"/> Sao_Tome_Principe	
<input type="checkbox"/> revista	<input type="checkbox"/> Timor	
<input type="checkbox"/> revista_tecnica		

Applications: Lexicography and Grammar

- Dictionary of the Sciences Academy of Lisbon: source of data for examples and dictionary entries
- Comprehensive Grammar of the Portuguese Language: one of the sources of linguistic contexts

Applications: Frequency Lexicon

Frequency lexicon

- 26.443 lemmas - 140.315 word forms (including compounds)
- Extracted from a subcorpus of the CRPC: CORLEX (16M)
- Lemmas with frequency ≥ 6
- PoS category and lemma
- Ex: most frequent verbs:

@ ser 'to be' (V) # 333747
@ ter 'to have' (V) # 117212
@ estar 'to be' (V) # 65520
@ fazer 'to do' (V) # 47273
@ ir 'to go' (V) # 45510
@ haver 'to have' (V) # 42560
@ poder 'can' (V) # 42138
@ dizer 'to say' (V) # 39890

@ dar 'to give' (V) # 27265
@ ver 'to see' (V) # 23316
@ dever 'to owe' (V) # 22034
@ saber 'to know' (V) # 21639
@ querer 'to want' (V) # 20593
@ ficar 'to stay' (V) # 18031
@ vir 'to come' (V) # 17884
@ passar 'to pass' (V) # 15936

Applications: MWE

- Lexicon of MWE
 - Expressions with more than one word
 - Syntactically well formed
 - With some degree of lexical or syntactic fixedness
 - May be idiomatic
 - ex: proverbs, idiomatic expressions, compounds, collocations
- Lexicon extracted from a subcorpus of CRPC with 50M
- 12.184 MWE (47.224 inflected forms)
221.847 concordances manually verified

Applications: varieties of Portuguese in the world

Lexicon of African varieties of Portuguese

- Extracted from 5 comparable corpora with 600.000 words each (written and spoken) for Angola, Cape Verde, Guinea-Bissau, Mozambique and Sao Tome and Príncipe.
- Contrastive values for lemmas and word forms, with distribution by subcorpus and genre.

	Angola	Cape Verde	Guinea-Bissau	Mozambique	S. Tome and Principe
Nouns	desatracção desinteriorização	desaculturação descrucificação	desfeitura	descamponês destriunfo	desarrazoavel
Verbs	desconseguir desestrelar	desbaralhar		desconseguir desconter	
Adjectives	descrispado	desapontador		desapetitoso	

Applications: lexical variation in diachrony

- Contrastive lexical analysis of Parliamentary Speeches before (1954-1974) and after (1974-94) the revolution of April 25, 1974
- 1 subcorpus for each of the four decades
- Frequencies in subcorpora and in the CRPC

Highly salient words in P1 and P2	Highly salient words in P3 and P4
Metrópole 'metropolis' Câmara Corporativa 'Corporate Board' Colonização 'colonization' espaço português 'portuguese space' Fomento Nacional 'National Development'	Democracia 'Democracy' Abstenções 'abstentions' salário mínimo 'minimum wage' pré-escolar 'preschool' partidos políticos 'political parties'

Annotation of modality in a subset of the CRPC

- Modality is defined as the expression of the speaker's (or the subject's) opinion and attitude towards the content of the sentence (Palmer, 1988; Oliveira, 1988; van der Auwera and Plungian, 1998)
- Modal typologies, in Linguistics, traditionally distinguish between epistemic and deontic modality
 - epistemic: degree of commitment to the truth of the proposition
 - deontic: command and permission
- Many other values are included but vary considerably across typologies, like volition, evaluation, commissive, internal capacity

Modality in Information Extraction

- recent trend in information extraction applications to go beyond the extraction of pure facts
 - to focus on personal opinions in sentiment analysis and opinion mining (Wiebe et al., 2005)
 - to distinguish between factual and non factual information (Saurí et al., 2006)
 - to detect uncertainty, speculation and negation in text mining (Szarvaz et al., 2008; Baker et al., 2010); Matsuyoshi et al., 2010)
- the result are some annotation schemes for English, where the concept of modality is less constrained (including, for example, temporal and conditional relations between events)

Annotation scheme

- Our approach is similar to the OntoSem annotation scheme (Nirenburg and McShane, 2008)
- We combine a practical annotation with a theoretically-oriented perspective focusing on modal values
- The scheme includes now 7 values and 9 sub-values (some initial values were removed during the annotation process)
- We annotate events, not entities
- We apply the scheme to different POS categories: verbs, nouns, adjectives, adverbs

Annotation scheme: modal values

Epistemic	knowledge	to know, knowledge
	belief	to believe, opinion, belief
	doubt	to doubt, doubt
	possibility	may, possibility
	interrogative	to ask
Deontic	obligation	to oblige, obligation
	permission	to allow, permission
Participant-internal	capacity	to be able, capacity, capable
	necessity	to need, necessity
Evaluation		to evaluate, good, bad, great
Volition		to want, to desire, to expect
Effort		to try, effort
Success		to achieve, success

Annotation scheme: components

Trigger		The lexical expression conveying the modal value
	Modal value	The modal value expressed by the trigger
	Ambiguity	When the triggers conveys more than one modal value in a context
	Polarity	To mark if the modal value expressed is positive or negative
Target	Polarity	The linguistic expression in the scope of the trigger
Source of the event mention		The producer of the sentence: speaker or writer
Source of the modality		The holder of the modality
Comment		Any aspect of the annotation which is not covered by the scheme

Annotation scheme: example

PCP quer esclarecimentos sobre Polis de Gondomar
'The PCP wants clarification fo the Gondomar Polis'

Trigger: quer

Modal value: volition

Polarity: positive

Target: esclarecimentos sobre Polis de Gondomar

Source of the event mention: speaker/writer

Source of the modality: PCP

Annotation scheme

Polarity

- We consider polarity only in relation to the trigger (and target)
- We keep polarity as a simple binary feature where the positive value expresses the unmarked cases

Trigger, Target and Sources

- We annotate triggers minimally keeping the smallest possible unit, for example the head noun in a noun phrase but we annotate targets and sources maximally, including all relevant parts
- Negation markers or auxiliary verbs are not considered part of the trigger
- We annotate discontinuous elements in the target if they are essential to the interpretation

Annotation scheme: example

No terreno das indústrias da cultura - cinema, livro, televisão -, arriscamo-nos a ser dominados pelo mercado americano.

‘On the terrain of cultural industry - film, books, tv -, we risk being dominated by the american market.’

Trigger: arriscamo

- Modal value: epistemic_possibility
- Polarity: positive

Target: No terreno das indústrias da cultura - cinema, livro, televisão -
@a ser dominados pelo mercado americano

Source of the modality: -nos

Source of the event mention: speaker/writer

Annotation software tool

- We use a free software annotation tool: MMAX2 (Müller and Strube, 2006)
<http://mmax2.sourceforge.net/>
- visual interface for marking up textual strings and creating links between the marked elements, allowing discontinuous elements and overlap
- Stand-off XML
- We consider modality as an event with several markables (marked elements)

Corpus sample

- We applied the annotation scheme for modality to a corpus sample of approximately 2000 sentences
- Sentences were extracted from the written part of the CRPC
- Sentences were gathered based on a list of 40 Portuguese verbs with modal meaning: we attempted to select equal sets of verbs for each modal meaning

Results (i)

- 1946 sentences annotated
- Triggers: 2377 (2509 modal values)
- Mainly verbal triggers, but also:
 - Nominal: *tentativa* ‘attempt’, *ambição* ‘ambition’
 - Adjectival: *difícil* ‘difficult’, *necessário* ‘necessary’
 - Adverbial: *obrigatoriamente* ‘obligatorily’ -> co-occurring with verbal triggers
- Source of the event mention: expressed in only 6%
- Source of the modality is present in 70%; in the other 30% of the cases source of the modality and source of the event mention is the same (speaker or writer)

Modal value	Freq	%
Deontic	740	28,8
obligation	581	22,7
permission	159	6,2
Epistemic	739	28,8
possibility	279	10,9
knowledge	183	7,1
belief	161	6,3
interrogative	87	3,4
doubt	29	1,1
Volition	396	15,4
Participant-internal	248	9,7
capacity	126	4,9
necessity	122	4,8
Evaluation	159	6,2
Success	119	4,6
Effort	110	4,3

Results (2)

- 450 events with negative polarity
- 84% triggered by the word *não* ‘no’
- other cases of negative polarity conveyed by the modal verb like *impedir* ‘to prevent’ or adjective *impossível* ‘impossible’
- Targets: nominal phrases, verbal phrases, subordinate clauses, main clauses when the trigger occurs in a parenthetical clause which has scope over the main one

Só que no futebol, como todos **sabem**, **tudo pode acontecer** (...).
‘It is just that in football, as you know, everything can happen’

Inter-annotator agreement

- Two linguists each annotated 50 sentences of the corpus
- Kappa-statistic (Cohen, 1960) computed for each field (including modal value and polarity associated to the trigger)

Field	Kappa value
Trigger	.65
Modal value	.85

- Similar results as for English in Matsuyoshi et al. (2010)

Particular cases: ambiguity

- Frequent ambiguity between epistemic possibility and deontic permission (34 cases), mostly with *permitir* 'to allow'

As condições climáticas **permitem** o desenvolvimento de árvores como abetos, pinheiros e outras plantas resinosas (coníferas).

'The climatic conditions permit the growth of trees such as spruce, pine and other coniferous plants (conifers).'

the climate makes it possible for trees to grow

the climate is a necessary condition for the growth of the trees

Particular cases: polarity (1)

- When the trigger in the scope of the negative particle scopes over another modality trigger
-> negative polarity of the particle affects both triggers

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!’

Trigger: conseguir (negative polarity)

Trigger: crer (positive polarity → inherited negative polarity)

Particular cases: polarity (2)

Negative polarity of two triggers produces a positive polarity of the second trigger which is not captured by the annotation

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’.

Trigger: **impossível** (negative polarity)

Trigger: **capacidade** (negative polarity)

-> capacity is interpreted as positive polarity despite the negative adverb

Particular cases: embedded triggers

The first trigger can have an influence on the modal value of the second, especially the values evaluation and deontic obligation

Se o aluno se perde, **pode** ser **difícil** voltar a apanhar.

‘If the student loses himself, it may be difficult for him to catch up again’.

Here, the modal value epistemic possibility (trigger *pode* ‘may’) influences the certainty of the value evaluation (trigger *difícil* ‘difficult’)

Modality and Focus

- Focus
a means to “give prominence to meaning-bearing elements in an expression.” (Krifka, 1995:240)
- Exclusive particles: *só*, *apenas* ‘only’

The effects of Focus on Modality (1)

- The exclusive **can restrict the set of possibilities** to the one presented (depending on the scope of the particle)

epistemic possibility → epistemic necessity

Isto **só** pode ter sido um acidente.

‘This can only have been an accident’

- **Weaker alternative** than expected (mirative reading)

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...’

The effects of Focus on Modality (2)

- In certain contexts, the exclusive particle is **not optional**
epistemic necessity + emphatic reading

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

- In contexts with modal ambiguity, **the exclusive can block** one of the modal readings

A nova lei (**só**) pode reduzir a injustiça.

'The new legislation can (only) reduce injustice'

w/o exclusive: epistemic possibility OR participant-internal capacity
with exclusive: epistemic possibility → epistemic necessity

New Features of the Annotation Scheme

- New attribute “focus” of the trigger component
Three possible values: none, exclusive, additive (for particles such as *também* ‘also’).
- New attribute “focus_value” of the trigger component to handle pragmatic values (see poster)
Three possible values: none, mirative and contrastive
- Two new components to mark the focus particle and its scope: “focus_cue” and “focus_scope”
- Ambiguity attribute of the “focus_scope” component

MMAX2

O plantel do Estrela da Amadora só deverá poder voltar a contar com o guardião Tiago durante a próxima semana , mas a data exacta para a recuperação clínica total do jogador ainda está por definir .

Fonte do Ministério da Saúde informa que está a aguardar a conclusão de um estudo a mortalidade causado pela vaga de calor de Agosto , da responsabilidade do Instituto Ricardo Jorge , *trabalho* que *só deverá ficar terminado dentro de dois ou três meses* .

Há quem defenda que *os medicamentos só devem ser usados numa primeira fase do tratamento* .

Por sua vez , *Wally Wonfor* , destacado para coordenar a tarefa dos apanha-bolas , estabeleceu as regras de tratamento relativamente a essa " diva " natural de Moscovo e residente na Florida : " A postura e a cortesia serão guardadas a todo o instante e *só se deve dizer obrigado quando tal for oportuno* " .

Também *os frutos só devem ser descascados , assim como os sumos preparados , na hora da refeição* , evitando tanto quanto possível *a exposição ao ar e ao calor* , que *podem alterar as vitaminas que contêm* . Cerca de 62 % do peso do nosso organismo é água .

MMAX2

One-click annotation

Panel

Settings

modal

sentences

< > Type

none target trigger source_modality source_event focus_cue scope_focus

modal_value

deontic_obligation

polarity

pos neg

polarity_cue

focus

none exclusive additive

focus_cue

só

extra_focus_value

none mirative contrastive

ambiguity

modal_class

set_28

comment

Suppress check Warn on extra attributes

Results of the annotation

- 100 sentences with the focus particle *só* ‘only’ and a modal verb (*poder* ‘can/may’, *dever* ‘must’, *ter de* ‘have to’, *ser capaz de* ‘be able to’, *querer* ‘want’)
- Deontic obligation and epistemic possibility are the most frequent modal values
- Temporal adjuncts are the most frequent scopes of the exclusive.
- Only 5 cases of ambiguous scope of the exclusive
- 7 contexts with *ter de* ‘have to’ and 3 contexts with *querer* ‘want’, which denote a weaker alternative than expected and were marked with the value “mirative”

Results of the annotation (2)

- The annotation task shows the dual nature of exclusives: in certain contexts, they both signal one of the possible alternatives and describe it as weaker than would be expected by the participants.
- The scope of the focus particle plays an important role in the meaning of the sentence.

Experiment in automatic annotation

- Collaborative work with U. of Evora
 - restrict our experiment to the most frequent (and most ambiguous) semi-auxiliary modal verbs: poder ‘may/can’, dever ‘shall/might’ and ter de ‘have to’
 - identification of modal semi-auxiliaries (which we call the modal trigger)
 - attribution of a modal value to this trigger
- average F-values: 51 - 67
average precision: 49 - 75
recall: 54 - 69

Further studies

- To look more closely at the evaluation value: refine its scope to better identify triggers
- Interaction between evaluation and polarity is more complex and seems to differ from the other values: negation does not affect the value itself, but rather the kind of evaluation
- Explore possible ways to integrate negation in the annotation scheme
- Apply the scheme to full texts, and also to the spoken corpus

Thank you!
Obrigada!

PoS tags

ADJ	Adjectives	bom, brilhante, eficaz, ...
ADV	Adverbs	hoje, já, sim, felizmente, ...
CARD	Cardinals	zero, dez, cem, mil, ...
CJ	Conjunctions	e, ou, tal como, ...
CL	Clitics	o, lhe, se, ...
CN	Common Nouns	computador, cidade, ideia, ...
DA	Definite Articles	o, os, ...
DEM	Demonstratives	este, esses, aquele, ...
DFR	Denominators of Fractions	meio, terço, décimo, %, ...
DGTR	Roman Numerals	VI, LX, MMIII, MCMXCIX, ...
DGT	Digits	0, 1, 42, 12345, 67890, ...
DM	Discourse Marker	olá, ...
EADR	Electronic Addresses	http://www.clul.ul.pt , ...
EOE	End of Enumeration	etc
EXC	Exclamatives	que, quanto, ...
GER	Gerunds	sendo, afirmando, vivendo, ...
GERAUX	Gerunds as auxiliary verbs	tendo, havendo, ...
IA	Indefinite Articles	uns, umas, ...
IND	Indefinites	tudo, alguém, ninguém ...
INF	Infinitive	ser, afirmar, viver, ...
INFAUX	Infinitive auxiliary verb	ter, haveremos, ...
INT	Interrogatives	quem, como, quando, ...
ITJ	Interjection	bolas, caramba, ...
LTR	Letters	a, b, c, ...

PoS tags

MGT	Magnitude Classes	unidade, dezena, dúzia,
MTH	Months	Janeiro, Dezembro, etc.
NP	Noun Phrases	idem, ...
ORD	Ordinals	primeiro, centésimo, penúltimo, ...
PADR	Part of Address	Rua, av., rot., ...
PNM	Part of Name	Lisboa, Manuel, ...
PNT	Punctuation Marks	., ?, (, ...
POSS	Possessives	meu, teu, seu, ...
PPA	Past Participles not in compound tenses	livros <u>lidos</u>
PP	Prepositional Phrases	algures, ...
PPT	Past Participle in compound tenses	tinha <u>afirmado</u> , tinha <u>vivido</u> , ...
PREP	Prepositions	de, para, em redor de, ...
PRS	Personals	eu, tu, ele, ...
QNT	Quantifiers	todos, muitos, nenhum, ...
REL	Relatives	que, cujo, ...
STT	Social Titles	Presidente, Dr., Prof., ...
SYB	Symbols	@, #, &, ...
TERMIN	Optional Terminations	(s), (as), ...
UM	“um” or “uma”	uma, uma, uns, umas
UNIT	Measurement units in abbreviated form	Kg, h, seg, Hz, Mbytes,...
VAUX	Finite "ter" or "haver" in compound tenses	temos, haveriam, ...
V	Verbs (other than PPA, PPT, INF or GER)	falou, falaria, ...
WD	Weak Days	segunda, quarta, ...
LADV1... LADVn	Multi-Word Adverbs	de facto, em suma, ...

PoS tags

CONTRACTIONS

PREP+ADV	Preposition and Adverb	dali, daí, daqui, ...
PREP+DA	Preposition and Definite Articles	aos, na, nos, da, dos
PREP+DEM	Preposition and Demonstratives	desse,deste, naquela
PREP+IND	Preposition and Indefinite	noutra, noutros, doutra, ...
PREP+INT	Preposition and Interrogative	aonde
PREP+PRS	Preposition and Personal pronoun	comigo, conosco, dela, nele, ...
PREP+QNT	Preposition and Quantifier	nalguns, noutro,noutras, ...
PREP+REL	Preposition and Relative	donde, aonde
PREP+UM	Preposition and "um" or "uma"	dum, duma

Referências

- Branco, A. & Silva, J. (2004). “Evaluating Solutions for the Rapid Development of State-of-the-Art POS Taggers for Portuguese”, in *Proceedings of the 4th International Conference of LREC*, pp.507-510, Lisboa, Portugal.
- Daelemans, W., Zavrel, J., Berck, P. & Gillis, S. (1996). “MBT: A memory-based part of speech tagger-generator”, in *Proceedings of the 4th ACL/SIGDAT Workshop on Very Large Corpora*, pp. 14-27.
- Hardie, A. (2012). “CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17:3, pp. 380-409.
- Van den Bosch, A. & Daelemans, W. (1999). “Memory-based morphological analysis“, in *Proceedings of the 37th Annual of ACL*, pp. 285-292.
- T. Kudo and Y. Matsumoto. (2003). Fast methods for kernel-based text analysis. In *Proceedings of ACL 2003*

Particular cases: structural ambiguity

- Ambiguity between deontic obligation and participant-internal necessity -> require different components and is not captured by our annotation

A tem de ser feito. 'A must be done'

Ambiguity	Option 1	Option 2
Trigger	tem de	tem de
Modal value	Deontic obligation	Part-internal necessity
Target	A @ ser feito	ser feito
Source of the modality	speaker	A

Experiment in automatic annotation: results

poder	Count	sentences			window parse tree		
		precision	recall	F	precision	recall	F
Total/weighted average	236	49	55.9	51.1	49.2	53.8	51.2
deontic permission	42	23.1	7.1	10.9	18.8	14.3	16.2
epistemic possibility	154	64.6	80.5	71.7	65.5	75.3	70.1
participant-internal capacity	40	16.1	12.5	14.1	18.5	12.5	14.9

dever	count	sentences			window parse tree		
		precision	recall	F	precision	recall	F
Total/weighted average	108	68.3	69.4	68.6	68.6	69.4	68.9
deontic obligation	71	74.4	81.7	77.9	75.0	80.3	77.6
epistemic possibility	37	56.7	45.9	50.7	56.3	48.6	52.2