

Identifying Pronominal Verbs: Towards Automatic Disambiguation of the Clitic 'se' in Portuguese

Magali Sanches Duran¹

Carolina Evaristo Scarton¹

Sandra Maria Aluísio¹

Carlos Ramisch²

Goal

Through a survey of the 6 uses of the clitic pronoun *se* in a Portuguese corpus, we intend to provide a lexicon of pronominal verbs corresponding to the use of *se* as a CONSTITUTIVE PARTICLE

"SE"
(clitic pronoun)

Rules of thumb

- SUBJECT INDETERMINATION → no NP before, PP after the verb
- PASSIVE → no NP before the verb, NP after the verb
- RECIPROCAL → verb in plural inflection
- INCHOATIVE → NP before the verb, often no NP nor PP after the verb
- CONSTITUTIVE PARTICLE → NP before the verb, PP after the verb

Methodology

- 1) Parse with PALAVRAS the PLN-BR-FULL corpus (Brazilian newspapers, 29M words)
- 2) Search with the mwetoolkit for sentences with verbs in third person singular followed by the clitic *se*
- 3) For each verb, test the sentences returned and annotate the possible uses according to the criteria above.

Pronominal Verb - definition

Verb that takes the clitic *se* as part of its lemma. In this case, the clitic *se* is not argumental (has no syntactic function and no semantic role). The *se* of a pronominal verb is called "constitutive particle".

se uses	Unambiguous	Ambiguous	Total
SUBJECT INDETERMINATION	17	6	23
PASSIVE	467	630	1097
REFLEXIVE	25	333	358
RECIPROCAL	0	33	33
INCHOATIVE	190	64	254
CONSTITUTIVE PARTICLE	83	104	187
Total	782	1170	1952

Types of pronominal verbs

- 1) Verbs used exclusively in pronominal form:
 - abster-se* (to abstain)
 - esvair-se* (to go out, to extinguish)
 - referir-se* (to refer or to concern)
- 2) Verbs that coexist with a non-pronominal form:
 - realizar* (to carry out) *realizar-se* (to feel fulfilled)
 - desculpar* (to forgive) *desculpar-se* (to apologize)
 - Importar* (to import) *importar-se* (to care about)
- 3) Verbs that have a pronominal form, but accept clitic drop with no change of meaning:
 - esquecer* and *esquecer-se* (to forget)
 - lembrar* and *lembrar-se* (to remember)
 - sentar* and *sentar-se* (to sit)

Future work

The generated lexicon will be used to guide current SRL annotation. Moreover, we intend to project the lexicon on the corpus and use the annotated data to train a classifier. We would also like to perform a cross-lingual study to build bilingual dictionaries.

1: University of São Paulo (Brazil) 2: Joseph Fourier University (France)

magali.duran@uol.com.br carol.scarton@gmail.com sandra@icmc.usp.br carlosramisch@gmail.com
Financial support from FAPESP (process 2011/22337-1) and from the CAMELEON project (CAPES-COFECUB 707-11)

SUBJECT INDETERMINATION

☐ syntactic ☑ semantic

Falou-**SE** muito nesse assunto

*Spoke-**SE** much about this matter

One has already spoken a lot about this matter

PASSIVE

☑ syntactic ☑ semantic

Sugeriram-**SE** muitas alternativas

*Suggested-**SE** many alternatives

Many alternatives were suggested

REFLEXIVE

☑ syntactic ☑ semantic

Você deveria olhar-**SE** no espelho

*You should look-**SE** in the mirror

You should look yourself in the mirror

RECIPROCAL

☑ syntactic ☑ semantic

Eles cumprimentaram-**SE** com um aperto de mão

*They greeted-**SE** with a handshake

They greeted each other with a handshake

INCHOATIVE

☑ syntactic ☐ semantic

Esse esporte popularizou-**SE** no Brasil

*This sport popularED-**SE** in Brazil

This sport became popular in Brazil

CONSTITUTIVE PARTICLE

☐ syntactic ☐ semantic

Eles queixaram-**SE** de dor no joelho

*They complained-**SE** about knee pain

They complained about knee pain