

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

Magali Sanches Duran<sup>♡</sup>, Carolina Evaristo Scarton<sup>♡</sup>,  
Sandra Maria Aluísio<sup>♡</sup>, Carlos Ramisch<sup>♣</sup>

<sup>♡</sup> University of São Paulo (Brazil)

<sup>♣</sup> Joseph Fourier University (France)

magali.duran@uol.com.br, carol.scarton@gmail.com

sandra@icmc.usp.br, carlosramisch@gmail.com

## Abstract

A challenging topic in Portuguese language processing is the multifunctional and ambiguous use of the clitic pronoun *se*, which impacts NLP tasks such as syntactic parsing, semantic role labeling and machine translation. Aiming to give a step forward towards the automatic disambiguation of *se*, our study focuses on the identification of pronominal verbs, which correspond to one of the six uses of *se* as a clitic pronoun, when *se* is considered a CONSTITUTIVE PARTICLE of the verb lemma to which it is bound, as a multiword unit. Our strategy to identify such verbs is to analyze the results of a corpus search and to rule out all the other possible uses of *se*. This process evidenced the features needed in a computational lexicon to automatically perform the disambiguation task. The availability of the resulting lexicon of pronominal verbs on the web enables their inclusion in broader lexical resources, such as the Portuguese versions of Wordnet, Propbank and VerbNet. Moreover, it will allow the revision of parsers and dictionaries already in use.

## 1 Introduction

In Portuguese, the word *se* is multifunctional. POS taggers have succeeded in distinguishing between *se* as a conjunction (meaning *if* or *whether*) and *se* as a pronoun (see Martins et al. (1999) for more details on the complexity of such task). As a clitic<sup>1</sup> pro-

<sup>1</sup>A clitic is a bound form, phonologically unstressed, attached to a word from an open class (noun, verb, adjective, adverbial). It belongs to closed classes, that is, classes that have grammatical rather than lexical meaning (pronouns, auxiliary verbs, determiners, conjunctions, prepositions, numerals).

noun, however, *se* has six uses:

1. marker of SUBJECT INDETERMINATION:  
*Já se falou muito nesse assunto.*  
*\*Has-SE already spoken a lot about this matter.*  
*One has already spoken a lot about this matter.*
2. marker of pronominal PASSIVE voice (synthetic passive voice):  
*Sugeriram-se muitas alternativas.*  
*\*Have-SE suggested many alternatives.*  
*Many alternatives have been suggested.*
3. REFLEXIVE pronoun (-self pronouns):  
*Você deveria se olhar no espelho.*  
*\*You should look-SE on the mirror.*  
*You should look at yourself on the mirror.*
4. RECIPROCAL pronoun (each other):  
*Eles se cumprimentaram com um aperto de mão.*  
*\*They greeted-SE with a handshake.*  
*They greeted each other with a handshake.*
5. marker of causative-INCHOATIVE alternation<sup>2</sup>:  
*Esse esporte popularizou-se no Brasil.*  
*\*This sport popularED-SE in Brazil.*  
*This sport became popular in Brazil.*
6. CONSTITUTIVE PARTICLE of the verb lexical item (pronominal verb):  
*Eles se queixaram de dor no joelho.*  
*\*They complained-SE about knee pain.*  
*They complained about knee pain.*

<sup>2</sup>Causative-inchoative alternation: a same verb can be used two different ways, one transitive, in which the subject position is occupied by the argument which causes the action or process described by the verb (causative use), and one intransitive, in which the subject position is occupied by the argument affected by the action or process (inchoative use).

| Clitic <i>se</i> uses   | Syntactic function | Semantic function |
|-------------------------|--------------------|-------------------|
| SUBJECT INDETERMINATION | NO                 | YES <sup>3</sup>  |
| PASSIVE                 | YES                | YES <sup>3</sup>  |
| REFLEXIVE               | YES                | YES               |
| RECIPROCAL              | YES                | YES               |
| INCHOATIVE              | YES                | NO                |
| CONSTITUTIVE PARTICLE   | NO                 | NO                |

Table 1: Uses of the clitic *se* from the point of view of syntax and semantics.

The identification of these uses is very important for Portuguese language processing, notably for syntactic parsing, semantic role labeling (SRL) and machine translation. Table 1 shows which of these six uses support syntactic and/or semantic functions.

Since superficial syntactic features seem not sufficient to disambiguate the uses of the pronoun *se*, we propose the use of a computational lexicon to contribute to this task. To give a step forward to solve this problem, we decided to survey the verbs undergoing *se* as an integral part of their lexical form (item 6), called herein *pronominal verbs*, but also known as *inherent reflexive verbs* (Rosário Ribeiro, 2011). Grammars usually mention this kind of verbs and give two classical examples: *queixar-se* (to complain) and *arrepender-se* (to repent). For the best of our knowledge, a comprehensive list of these multiword verbs is not available in electronic format for NLP uses, and not even in a paper-based format, such as a printed dictionary.

An example of the relevance of pronominal verbs is that, in spite of not being argumental, that is, not being eligible for a semantic role label, the use of *se* as a CONSTITUTIVE PARTICLE should integrate the verb that evokes the argumental structure, as may be seen in Figure 1.

The identification of pronominal verbs is not a trivial task because a pronominal verb has a nega-

<sup>3</sup>In these cases, the clitic may support the semantic role label of the suppressed external argument (agent).

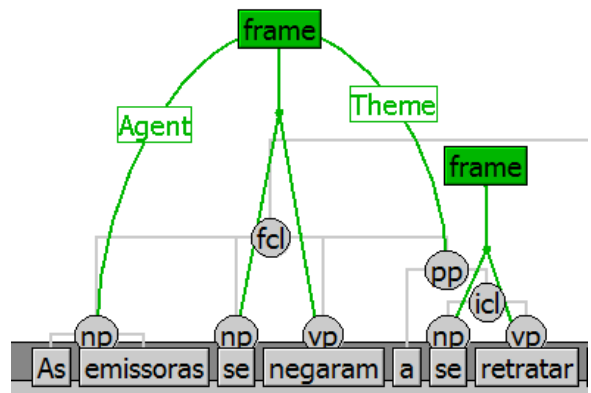


Figure 1: Sentence *The broadcasters refused to apologize* includes pronominal verbs *negar-se* (refuse) and *retratar-se* (apologize) that evoke frames in SRL.

tive definition: if *se* does not match the restrictions imposed by the other five uses, so it is a CONSTITUTIVE PARTICLE of the verb, that is, it composes a multiword. Therefore, the identification of pronominal verbs requires linguistic knowledge to distinguish *se* as a CONSTITUTIVE PARTICLE from the other uses of the the pronoun *se* (SUBJECT INDETERMINATION, PASSIVE, REFLEXIVE, RECIPROCAL and INCHOATIVE.)

There are several theoretical linguistic studies about the clitic pronoun *se* in Portuguese. Some of these studies present an overview of the *se* pronoun uses, but none of them prioritized the identification of pronominal verbs. The study we report in this paper is intended to fill this gap.

## 2 Related Work

From a linguistic perspective, the clitic pronoun *se* has been the subject of studies focusing on: SUBJECT INDETERMINATION and PASSIVE uses (Morais Nunes, 1990; Cyrino, 2007; Pereira-Santos, 2010); REFLEXIVE use (Godoy, 2012), and INCHOATIVE use (Fonseca, 2010; Nunes-Ribeiro, 2010; Rosário Ribeiro, 2011). Despite none of these works concerning specifically pronominal verbs, they provided us an important theoretical basis for the analysis undertaken herein.

The problem of the multifunctional use of clitic pronouns is not restricted to Portuguese. Romance languages, Hebrew, Russian, Bulgarian and others also have similar constructions. There are

crosslinguistic studies regarding this matter reported in Siloni (2001) and Slavcheva (2006), showing that there are partial coincidence of verbs taking clitic pronouns to produce alternations and reflexive voice.

From an NLP perspective, the problem of the ambiguity of the clitic pronoun *se* was studied by Martins et al. (1999) to solve a problem of categorization, that is, to decide which part-of-speech tag should be assigned to *se*. However, we have not found studies regarding pronominal verbs aiming at Portuguese automatic language processing.

Even though in Portuguese all the uses of the clitic pronoun *se* share the same realization at the surface form level, the use as a CONSTITUTIVE PARTICLE of pronominal verbs is the only one in which the verb and the clitic form a multiword lexical unit on its own. In the other uses, the clitic keeps a separate syntactic and/or semantic function, as presented in Table 1.

The particle *se* is an integral part of pronominal verbs in the same way as the particles of English phrasal verbs. As future work, we would like to investigate possible semantic contributions of the *se* particle to the meaning of pronominal verbs, as done by Cook and Stevenson (2006), for example, who try to automatically classify the uses of the particle *up* in verb-particle constructions. Like in the present paper, they estimate a set of linguistic features which are in turn used to train a Support Vector Machine (SVM) classifier citecook:2006:mwe.

### 3 Methodology

For the automatic identification of multiword verb+*se* occurrences, we performed corpus searches on the PLN-BR-FULL corpus (Muniz et al., 2007), which consists of news texts extracted from a major Brazilian newspaper, *Folha de São Paulo*, from 1994 to 2005, with 29,014,089 tokens. The corpus was first preprocessed for sentence splitting, case homogenization, lemmatization, morphological analysis and POS tagging using the PALAVRAS parser (Bick, 2000). Then, we executed the corpus searches using the *mwetoolkit* (Ramisch et al., 2010). The tool allowed us to define two multilevel word patterns, for proclitic and enclitic cases, based on surface forms, morphology and POS. The pat-

terns covered all the verbs in third person singular (POS=*V\**, morphology=*3S*) followed/preceded by the clitic pronoun *se* (surface form=*se*, POS=*PERS*). The patterns returned a set of *se* occurrences, that is, for each verb, a set of sentences in the corpus in which this verb is followed/preceded by the clitic *se*.

In our analysis, we looked at all the verbs taking an enclitic *se*, that is, where the clitic *se* is attached after the verb. We could as well have included the occurrences of verbs with a proclitic *se* (clitic attached before the verb). However, we suspected that this would increase the number of occurrences (sentences) to analyze without a proportional increase in verb lemmas. Indeed, our search for proclitic *se* occurrences returned 40% more verb lemmas and 264% more sentences than for the enclitic *se* (59,874 sentences), thus confirming our hypothesis. Moreover, as we could see at a first glance, proclitic *se* results included *se* conjunctions erroneously tagged as pronouns (when the parser fails the categorial disambiguation). This error does not occur when the pronoun is enclitic because Portuguese orthographic rules require a hyphen between the verb and the clitic when *se* is enclitic, but never when it is proclitic.

We decided to look at sentences as opposed to looking only at candidate verb lemmas, because we did not trust that our intuition as native speakers would be sufficient to identify all the uses of the clitic *se* for a given verb, specially as some verbs allow more than one of the six uses we listed herein.

For performing the annotation, we used a table with the verb lemmas in the lines and a column for each one of the six uses of *se* as a clitic pronoun. Working with two screens (one for the table and the other for the sentences), we read the sentences and, once a new use was verified, we ticked the appropriate column. This annotation setup accelerated the analyses, as we only stopped the reading when we identified a new use. The annotation was performed manually by a linguist, expert in semantics of Portuguese verbs, and also an author of this paper.

After having summarized the results obtained from corpus analysis, we realized that some cliticized verb uses that we know as native speakers did not appear in the corpus (mainly reflexive and reciprocal uses). In these cases, we added a comment on our table which indicates the need to look for the use

in another corpus aiming to confirm it.

For example, the most frequent cliticized verb, *tratar-se* has no occurrence with the meaning of *to take medical treatment*. We checked this meaning in another corpus and found one example: *O senador se tratou com tecido embrionário...* (*\*The senator treated himself with embryonic tissue...*), proving that our intuition may help us to improve the results with specific corpus searches. A comparative multi-corpus extension of the present study is planned as future work.

The strategy we adopted to analyze the sentences in order to identify pronominal verbs was to make a series of questions to rule out the other possible *se* uses.

**Question 1** Does the *se* particle function as a marker of PASSIVE voice or SUBJECT INDETERMINATION?

In order to answer this question, it is important to know that both uses involve the suppression of the external argument of the verb. The difference is that, in the pronominal PASSIVE voice, the remaining NP (noun phrase) is shifted to the subject position (and the verb must then be inflected according to such subject), whereas in SUBJECT INDETERMINATION, the remaining argument, always a PP (prepositional phrase), remains as an indirect object. For example:

- Pronominal PASSIVE voice:  
*Fizeram-se várias tentativas.*  
*\*Made-SE several trials.*  
*Several trials were made.*
- SUBJECT INDETERMINATION:  
*Reclamou-se de falta de higiene.*  
*\*Complained-SE about the lack of hygiene.*  
*One has complained about the lack of hygiene.*

**Question 2** Is it possible to substitute *se* for *a si mesmo* (-self)?

If so, it is a case of REFLEXIVE use. A clue for this is that it is always possible to substitute *se* for another personal pronoun, creating a non-reflexive use keeping the same subject. For example:

- *Ele perguntou-se se aquilo era certo.*  
*He asked himself whether that was correct.*
- *Ele perguntou-me se aquilo era certo.*  
*He asked me whether that was correct.*

**Question 3** Is it possible to substitute *se* for *um ao outro* (each other)?

If so, it is a case of RECIPROCAL use. A clue for this interpretation is that, in this case, the verb is always in plural form as the subject refers to more than one person. RECIPROCAL uses were not included in the corpus searches, as we only looked for cliticized verbs in third person singular. However, aiming to gather data for future work, we have ticked the table every time we annotated sentences of a verb that admits reciprocal use. The reciprocal use of such verbs have been later verified in other corpora.

- *Eles se beijaram.*  
*They kissed each other.*

**Question 4** Has the verb, without *se*, a transitive use? If so, are the senses related to causative-inchoative alternation? In other words, is the meaning of the transitive use *to cause X become Y*?

If so, it is a case of INCHOATIVE use, for example:

- *A porta abriu-se.*  
*The door opened.*

Compare with the basic transitive use:

- *Ele abriu a porta.*  
*He opened the door.*

It is important to mention that verbs which allow causative-inchoative alternation in Portuguese may not have an equivalent in English that allows this alternation, and vice-versa. For example, the inchoative use of the verb *tornar* corresponds to the verb *to become* and the causative use corresponds to the verb *to make*:

- *Esse fato tornou-se conhecido em todo o mundo.*  
*This fact became known all around the world.*
- *A imprensa tornou o fato conhecido em todo o mundo.*  
*The press made the fact known all around the world.*

If the verb being analyzed failed the four tests, the clitic *se* has neither semantic nor syntactic function and is considered a CONSTITUTIVE PARTICLE of the verb, for example:

- *Ele vangloriou-se de seus talentos.*  
*He boasted of his talents.*

Therefore, we made the identification of pronominal verbs based on the negation of the other possibilities.

#### 4 Discussion

The corpus search resulted in 22,618 sentences of cliticized verbs, corresponding to 1,333 verb lemmas. Some verbs allow only one of the uses of the clitic *se* (unambiguous cliticized verbs), whereas others allow more than one use (ambiguous cliticized verbs), as shown in Table 2. Therefore, a lexicon can only disambiguate part of the cliticized verbs (others need additional features to be disambiguated).

The analysis of the verbs' distribution reveals that 10% of them (133) account for 73% of the sentences. Moreover, among the remaining 90% verb lemmas, there are 477 *hapax legomena*, that is, verbs that occur only once. Such distribution indicates that computational models which focus on very frequently cliticized verbs might significantly improve NLP applications.

Contrary to our expectations, very frequently cliticized verbs did not necessarily present high polysemy. For example, the most frequent verb of our corpus is *tratar*, with 2,130 occurrences. Although *tratar-se* has more than one possible use, only one appeared in the corpus, as a marker of SUBJECT INDETERMINATION, for example:

- *Trata-se de uma nova tendência.*  
*It is the case of a new tendency.*

Despite being very frequent, when we search for translations of *tratar-se de* in bilingual (parallel) Portuguese-English corpora and dictionaries available on the web,<sup>4,5,6</sup> we observed that there are several solutions to convey this idea in English (determining a subject, as English does not allow subject omission). Six examples extracted from the Compara corpus illustrate this fact:

<sup>4</sup><http://www.linguateca.pt/COMPARA/>

<sup>5</sup><http://www.linguee.com.br/portugues-ingles>

<sup>6</sup><http://pt.bab.la/dicionario/portugues-ingles>

| <i>se</i> uses          | Unamb.     | Amb.        | Total       |
|-------------------------|------------|-------------|-------------|
| SUBJECT INDETERMINATION | 17         | 6           | 23          |
| PASSIVE                 | 467        | 630         | 1097        |
| REFLEXIVE               | 25         | 333         | 358         |
| INCHOATIVE              | 190        | 64          | 254         |
| RECIPROCAL              | 0          | 33          | 33          |
| CONSTITUTIVE PARTICLE   | 83         | 104         | 187         |
| <b>Total</b>            | <b>782</b> | <b>1170</b> | <b>1952</b> |

Table 2: Proportion of unambiguous (Unamb.) and ambiguous (Amb.) verbs that allow each *se* use.

- **Trata-se de recriar o próprio passado.**  
**It's a question of re-creating your own past.**
- **Mas o assunto era curioso, trata-se do casamento, e a viúva interessa-me.**  
*But the subject was a curious one; it was about her marriage, and the widow interests me.*
- **Não há mais dúvidas, trata-se realmente de um louco.**  
*There's no longer any doubt; we're truly dealing with a maniac.*
- **Trata-se realmente de uma emergência, Sr. Hoffman.**  
*This really is a matter of some urgency, Mr Hoffman.*
- **Trata-se de um regime repousante e civilizado.**  
**It is a restful, civilized régime.**
- **Trata-se de um simples caso de confusão de identidades, dizem vocês.**  
*(??) Simple case of mistaken identity.*

In what concerns specifically pronominal verbs, our analysis of the data showed they are of three kinds:

1. Verbs that are used exclusively in pronominal form, as *abster-se* (*to abstain*). This does not mean that the pronominal form is unambiguous, as we found some pronominal verbs that present more than one sense, as for example the verb *referir-se*, which means *to refer* or *to concern*, depending on the subject's animacy status [+ human] or [- human], respectively;

2. Verbs that have a non-pronominal and a pronominal form, but both forms are not related, e.g.: *realizar* (to make or to carry on, which allows the passive alternation *realizar-se*); and the pronominal form *realizar-se* (to feel fulfilled);
3. Verbs that have pronominal form, but accept clitic drop in some varieties of Portuguese without change of meaning, as *esquecer-se* and *esquecer* (both mean to forget)

We did not study the clitic drop (3), but we uncovered several pronominal verbs of the second kind above (2). The ambiguity among the uses of *se* increases with such cases. The verb *desculpar* (to forgive), for example, allows the REFLEXIVE use *desculpar-se* (to forgive oneself), but also constitutes a pronominal verb: *desculpar-se* (to apologize). The verb *encontrar* (to find) allows the REFLEXIVE use (to find oneself, from a psychological point of view) and the PASSIVE use (to be found). The same verb also constitutes a pronominal verb which means to meet (1) or functions as a copula verb, as to be (2):

1. *Ele encontrou-se com o irmão.*  
*He met his brother.*
2. *Ele encontra-se doente.*  
*He is ill.*

In most sentences of cliticized verbs' occurrences, it is easy to observe that, as a rule of thumb:<sup>7</sup>

- SUBJECT INDETERMINATION uses of *se* do not present an NP before the verb, present a PP after the verb and the verb is always inflected in the third person singular;
- PASSIVE uses of *se* present an NP after the verb and no NP before the verb;
- INCHOATIVE uses of *se* present an NP before the verb and almost always neither a PP nor a NP after the verb;
- CONSTITUTIVE PARTICLE uses of *se* present an NP before the verb and a PP after the verb;

<sup>7</sup>Syntactic clues do not help to identify REFLEXIVE verbs. The distinction depends on the semantic level, as the reflexive use requires a [+ animate] subject to play simultaneously the roles of agent and patient.

- RECIPROCAL uses of *se* only occur with verbs taking a plural inflection.

Problems arise when a sentence follows none of these rules. For example, subjects in PASSIVE use of *se* usually come on the right of the verb. Thus, when the subject appears before the verb, it looks, at a first glance, to be an active sentence. For example:

- *O IDH baseia-se em dados sobre renda, escolaridade e expectativa de vida.*  
*\*The HDI bases-SE on income, education and life expectancy data.*  
*The HDI is based on income, education and life expectancy data.*

These cases usually occur with stative passives (see Rosário Ribeiro (2011, p. 196)) or with ditransitive action verbs<sup>8</sup> when a [- animate] NP takes the place usually occupied by a [+ animate] NP. Semantic features, again, help to disambiguate and to reveal a non-canonical passive.

The opposite also occurs, that is, the subject, usually placed on the left of the verb in active voice, appears on the right, giving to the sentence a false passive appearance:

- *Desesperaram-se todos os passageiros.*  
*\*Fell-SE into despair all the passengers.*  
*All the passengers fell into despair.*

Sometimes the meaning distinctions of a verb are very subtle, making the matter more complex. In the following sections, we comment two examples of difficult disambiguation.

#### 4.1 Distinguishing Pronominal PASSIVE Voice from Pronominal Verbs

The verb *seguir* (to follow) conveys the idea of *obeying* when it has a [+ human] subject in the active voice (an agent). The passive voice may be constructed using *se*, like in (2). Additionally, this verb has a pronominal active use, *seguir-se*, which means to occur after, as shown in (3):

1. Active voice:

- *[Eles]<sub>Agent</sub> seguem [uma série de convenções]<sub>Theme - thing followed</sub>.*  
*They follow a series of conventions.*

<sup>8</sup>Ditransitive verbs take two internal arguments: an NP as direct object and a PP as indirect object.

## 2. PASSIVE voice:

- *Segue-se [uma série de convenções]*<sub>Theme - thing followed</sub>  
*A series of conventions are followed.*

## 3. Pronominal verb – active voice:

- *[A queda]*<sub>Theme - thing occurring after</sub> *seguiu-se [à divulgação dos dados de desemprego em o país]*<sub>Theme - thing occurring before</sub>  
*The drop followed the announcement of unemployment figures in the country.*

The preposition *a* introducing one of the arguments in (3) distinguishes the two meanings, as the PASSIVE voice presents an NP and not a PP immediately after or before the verb.

### 4.2 Distinguishing REFLEXIVE, INCHOATIVE and PASSIVE Uses

The verb *transformar*, when cliticized, may be interpreted as a PASSIVE (*to be transformed*), as a REFLEXIVE (*to transform oneself*) or as an INCHOATIVE use (*to become transformed*). The PASSIVE voice is identified by the subject position, after the verb (1). The difference between the REFLEXIVE (2) and INCHOATIVE (3) uses, on its turn, is a semantic feature: only a [+ human] subject may act to become something (REFLEXIVE use):

#### 1. PASSIVE:

- Transformou-se o encontro em uma grande festa.*  
*The meeting was transformed into a big party.*

#### 2. REFLEXIVE:

- *A mulher jovem transformou-se em uma pessoa sofisticada.*  
*The young woman transformed herself into a sophisticated person.*

#### 3. INCHOATIVE:

- *O encontro transformou-se em uma grande festa.*  
*The meeting transformed into a big party.*

## 5 Conclusions and Future Work

The lexicon gathered through this research will partially enable disambiguating the uses of the clitic pronoun *se*, as there are several verbs that allow only

one of the *se* clitic uses. For the other verbs, whose polysemy entails more than one possible use of *se*, it is necessary to add further information on each verb sense.

The analysis we reported here evidenced the need for enriching Portuguese computational lexicons, encompassing (a) the semantic role labels assigned by each verb sense, (b) the selectional restrictions a verb imposes to its arguments, and (c) the alternations a verb (dis)allows. The semantic predicate decomposition used by Levin (1993) has proved to be worthy to formalize the use of *se* in reflexive constructions (Godoy, 2012) and we think it should be adopted to describe other uses of the pronoun *se*. Another alternative is to construct a detailed computational verb lexicon along the lines suggested by Gardent et al. (2005), based on Maurice Gross' lexicon-grammar.

The data generated by this study can also be used to automatically learn classifiers for ambiguous uses of the clitic *se*. On the one hand, the annotation of uses can be semi-automatically projected on the sentences extracted from the corpus. On the other hand, the findings of this work in terms of syntactic and semantic characteristics can be used to propose features for the classifier, trying to reproduce those that can be automatically obtained (e.g., subcategorization frame) and to simulate those that cannot be easily automated (e.g., whether the subject is animate). For these future experiments, we intend to compare different learning models, based on SVM and on sequence models like conditional random fields (Vincze, 2012).

As languages are different in what concerns allowed alternations, the use of clitic *se* in Portuguese becomes even more complex when approached from a bilingual point of view. Depending on how different the languages compared are, the classification of *se* adopted here may be of little use. For example, several verbs classified as reflexive in Portuguese, like *vestir-se* (*to dress*), *barbear-se* (*to shave*) and *demitir-se* (*to resign*) are not translated into a reflexive form in English (*\*to dress oneself*, *\*to shave oneself* and *\*to dismiss oneself*). Similarly, typical inchoative verb uses in Portuguese need to be translated into a periphrasis in English, like *surpreender-se* (*to be surprised at*), *orgulhar-se* (*to be proud of*) and *irritar-se* (*to get angry*). Such evidences lead

us to conclude that it would be useful to count on a bilingual description not only of pronominal, but also of the other *se* uses.

The results of this work are available at [www.nilc.icmc.usp.br/portlex](http://www.nilc.icmc.usp.br/portlex).

## Acknowledgments

This study was funded by FAPESP (process 2011/22337-1) and by the CAMELEON project (CAPES-COFECUB 707-11).

## References

- Eckhard Bick. 2000. *The parsing system Palavras*. Aarhus University Press. 411 p.
- Paul Cook and Suzanne Stevenson. 2006. Classifying particle semantics in English verb-particle constructions. In *Proceedings of MWE 2006*, pages 45–53, Sydney, Australia.
- Sonia Maria Lazzarino Cyrino. 2007. Construções com SE e promoção de argumento no português brasileiro: Uma investigação diacrônica. *Revista da ABRALIN*, 6:85–116.
- Paula Fonseca. 2010. *Os verbos pseudo-reflexos em Português Europeu*. Master's thesis, Universidade do Porto.
- Claire Gardent, Bruno Guillaume, Guy Perrier, and Ingrid Falk. 2005. Maurice gross' grammar lexicon and natural language processing. In *Proceedings of the 2nd Language and Technology Conference*, Poznań, Poland.
- Luisa Andrade Gomes Godoy. 2012. *A reflexivização no PB e a decomposição semântica de predicados*. Ph.D. thesis, Universidade Federal de Minas Gerais.
- Beth Levin. 1993. *English Verb Classes and Alternations: a preliminary investigation*. The University of Chicago Press, Chicago, USA.
- Ronaldo Teixeira Martins, Gisele Montilha, Lucia Helena Machado Rino, and Maria da Graça Volpe Nunes. 1999. Dos modelos de resolução da ambiguidade categorial: o problema do SE. In *Proceedings of IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR 1999)*, pages 115–128, Évora, Portugal, September.
- Jairo Morais Nunes. 1990. *O famigerado SE: uma análise sincrônica e diacrônica das construções com SE apassivador e indeterminador*. Master's thesis, Universidade Estadual de Campinas.
- Marcelo Muniz, Fernando V. Paulovich, Rosane Minghim, Kleber Infante, Fernando Muniz, Renata Vieira, and Sandra Aluísio. 2007. Taming the tiger topic: an XCES compliant corpus portal to generate subcorpus based on automatic text topic identification. In *Proceedings of The Corpus Linguistics Conference (CL 2007)*, Birmingham, UK.
- Pablo Nunes-Ribeiro. 2010. *A alternância causativa no Português do Brasil: a distribuição do clítico SE*. Master's thesis, Universidade Federal do Rio Grande do Sul.
- José Ricardo Pereira-Santos. 2010. *Alternância passiva com verbos transitivos indiretos do português do Brasil*. Master's thesis, Universidade de Brasília.
- Carlos Ramisch, Aline Villavicencio, and Christian Boitet. 2010. Multiword expressions in the wild? the mwetoolkit comes in handy. In *Proceedings of the 23rd COLING (COLING 2010) - Demonstrations*, pages 57–60, Beijing, China.
- Sílvia Isabel do Rosário Ribeiro. 2011. *Estruturas com "se" Anafórico, Impessoal e Decausativo em Português*. Ph.D. thesis, Faculdade de Letras da Universidade de Coimbra.
- Tal Siloni. 2001. Reciprocal verbs. In *Online Proceedings of IATL 17*, Jerusalem, Israel.
- Milena Slavcheva. 2006. Semantic descriptors: The case of reflexive verbs. In *Proceedings of LREC 2006*, pages 1009–1014, Genoa, Italy.
- Veronika Vincze. 2012. Light verb constructions in the szegedparalellFX English–Hungarian parallel corpus. In *Proceedings of LREC 2012*, Istanbul, Turkey.