

# A Rule-Based Approach to the Identification of Spanish Zero Pronouns

Luz Rello\* and Iustina Ilisei  
Research Group in Computational Linguistics  
University of Wolverhampton, Stafford Street  
WV1 1SB, United Kingdom  
*luzrello, iustina.ilisei@gmail.com*

## Abstract

This paper presents a new rule-based method to identify Spanish zero pronouns. The paper describes the comparative evaluation of a baseline method for the identification of zero pronouns with an approach that supplements the baseline by adding a set of restrictions treating impersonal sentences and other zero subject expressions. The identification rules have been tested on a new corpus in which zero pronouns have been manually annotated (the Z-Corpus). The comparative evaluation shows that this rule-based method outperforms the baseline.

## Keywords

zero pronoun identification; pronominal zero anaphora; subject ellipsis

## 1 Introduction

The identification of zero pronouns in Spanish is the first step in the development of pre-processing tools useful in NLP fields where zero anaphora resolution is necessary, *inter alia*, automatic summarisation, machine translation, question answering and the generation of multiple choice tests.

The identification and resolution of the Spanish zero pronoun is also relevant because this type of zero anaphora is fairly frequent in Spanish. Our previous research [29] on the distribution of zero pronouns reveals their ubiquity in three different genres (legal, instructional and encyclopaedic), and shows that the distribution of zero anaphors is more uniform in encyclopaedic and instructional genres than in legal texts.

This paper presents a method for the identification of zero pronouns in Spanish that distinguishes between omitted subjects which can be lexically retrieved and those that cannot. This step is essential for its resolution as the latter have to be discarded in the resolution process.

This study required the use of corpus analysis and annotation as well as deep dependency parsing techniques for the creation of a new corpus in order to test the experiments and validate the rule-based algorithm for zero pronoun identification.

\*This work was supported by Obra Social Fundación “la Caixa”.

Section 2 presents a description of the types of subject ellipsis that occur in Spanish, a delimitation of the zero anaphors taken into account in this research and some terminological and conceptual explanations about how zero pronouns and zero subjects have been treated throughout the previous literature, considering both linguistic and computational approaches.

The remainder of the paper is structured as follows: In Section 3 related work on the zero pronoun identification is provided while Section 4 describes the compilation and annotation of the corpora. Section 5 is devoted to the detailed description of the rule-based method which is evaluated in Section 6. Finally, in Section 7 we draw conclusions and discuss the future work.

## 2 Zero Pronouns, Zero Subjects

“In spite of the widespread and fruitful use of zero signs in linguistic theory, there is no universally accepted definition of the concept of the zero linguistic sign itself. [...] Nevertheless, a maximally general definition which would cover all possible types of zero signs is a sign whose signifier is *empty*” [22].

Following this idea, two kinds of elliptic subjects are found in Spanish: implicit subjects and zero subjects. The distinction lies in the fact that while the former can be lexically retrieved (a), the latter cannot (b) [18].

(a) *zp[Vosotros]*<sup>1</sup> no tenéis que preocuparos.  
*[You] don't have to worry.*

(b)  $\emptyset$  Llueve.  
*[It] is raining.*

In Spanish, clauses with zero subject (b) are syntactically impersonal whereas omitted or implicit subjects (a), which are not phonetically realised, can be lexically retrieved [18].

When the phenomenon of nominal ellipsis and, specifically, the Spanish subject omission is described in the literature, both zero subjects and implicit subjects are considered cases of subject ellipsis [5]. However a consensus has emerged in which four kinds of Spanish subject ellipsis [4] are distinguished.

<sup>1</sup> In what follows, the sing  $\emptyset$  will represent zero subjects, while implicit subjects will be presented between brackets and will be preceded by *zp* (zero pronoun).

1. Implicit Subject in a clause containing a finite verb<sup>2</sup>:

- (c) *zp[Ellos] no vendrán.*  
*[They] won't come.*

2. Argumental impersonal subject:

- (d) *En este estudio Ø se trabaja bien.*  
*In this room [one] can work properly.*

3. Non-argumental impersonal subject:

- (e) *Ø Nieva.*  
*[It] is snowing.*

4. Omitted subject in a non-finite verb clause:

- (f) *Juan intentaba (Ø decírselo a María.)*  
*John tried ([John] to tell Mary.)*

Moreover, elision of the subject can affect not only the noun phrase head (g), but also the entire noun phrase (h) [5].

- (g) *Miguel dice que los zp[familiares, amigos...] de María no vendrán.*

*Michael says that Maria's zp[friends, family...] won't come.*

- (h) *Miguel dice que zp[ellos] no vendrán.*

*Michael says that zp[they] wont come.*

Furthermore, the noun phrases affected by ellipsis in Spanish can syntactically function either as subjects (g, h) or as objects which are datives in most of the cases (i) [4].

- (i) *Eso zp[nos] induce a pensar que la noticia es falsa.*

*This lead zp[us] to think that the news are false.*

In computational approaches to anaphora resolution [25] and specifically those focused on the resolution of pronominal zero anaphora [17], the term *zero pronoun* is used.

Some linguistic approaches make also use of the term zero pronoun which is not equivalent to the computational concept. The *Meaning Text Theory* considers a zero pronoun in subject position the same as a zero subject (Ø Llueve, *[It] is raining*) [22], while in the *Zero Hypothesis* [21] a zero pronoun can have phonetic content (full pronoun) or not (null pronoun). In this last theory, the concept of zero pronoun has to do only with its lack of lexical content in opposition to *lexical pronouns* [1].

By contrast, in NLP approaches, a zero pronoun is the zero anaphor or the resultant “gap”, where pronominal zero anaphora or ellipsis occurs [25].

As in the linguistic approach, two different views about the distribution of zero pronouns have emerged. Whereas some consider that Spanish zero pronouns

only appear in subject position [27], others contend that zero pronouns can occur in the object position as well [1].

As observed in our corpus, a zero pronoun can be either anaphoric (j), when it points back to its antecedent, or non-anaphoric (k), when there is no linguistic entity to which it refers. Both are examples of pronominal zero anaphora [25] and a distinction has been made between the two types during the annotation process.

- (j) *La costumbre<sub>i</sub> sólo regirá en defecto de ley aplicable, siempre que zp[ella]<sub>i</sub> no sea contraria a la moral o al orden público y que zp[ella]<sub>i</sub> resulte probada.*<sup>3</sup>

*The custom<sub>i</sub> will only be valid by default on the applicable law, whenever zp[it]<sub>i</sub> is not opposite to the moral a the public order and zp[it]<sub>i</sub> is passed.*

- (k) *Los sistemas químicos que zp[nosotros] podemos estudiar por vía experimental son más complejos.*

*The chemistry systems that zp[we] can study through experimental way are more complex.*

It has been noted in previous work that both phenomena, anaphoric or non-anaphoric ellipsis, are closely related [14] and occur in ambiguous clauses especially when omitting the subject (l).

- (l) *La serpiente estaba detrás de Pedro y sin embargo, zp[¿él/ella?] no se asustó [26].*

*The snake was behind Peter, nevertheless, zp[he or it?] was not afraid.*

In conclusion, the goal of this study is to identify zero pronouns in text, *i.e.* implicit subjects—not zero subjects—which are the zero anaphors, when pronominal zero anaphora happens affecting the entire noun phrase in clauses containing finite verbs. Both types of zero pronouns, anaphoric and non-anaphoric, are taken into consideration.

### 3 Related Work

In addition to the studies on this topic in other languages such as Japanese, Chinese, Korean or Turkish, two approaches can be distinguished in the Spanish case: either the zero pronoun identification is considered as the first step to zero anaphora resolution [17], or the identification of zero pronouns has been useful in itself in the investigation of the convergence universal [8, 9].

The most influential work on this topic is the Ferrández and Peral algorithm for zero pronoun resolution [17] together with their previous works [15, 16, 27]. In this successful method, the location of the omitted subject is based on clause identification and the detection of noun phrases which appear before the verb. The mentioned authors use partial parsing while in this study deep dependency parsing and the check of whether each verb has a dependent previous or subsequent (m) subject is used.

<sup>2</sup> While some linguists consider these examples as a type of elision of the subject [5], it has been claimed that the subject is not elided as it is present in the verb morphology [28].

<sup>3</sup> Unless otherwise specified, all the examples provided from now on are taken from the Z-corpora.

- (m) Ø Se prohíben las asociaciones secretas y las de carácter paramilitar

*\*[It] is forbidden secret and paramilitary associations.*

The detection of verbs whose omitted subject appears before the verb in the Ferrández and Peral method [17] is very successful, however their lowest success rate (80%) corresponds with the location of verbs with zero subject. Since our set of rules are precisely focused on the treatment of this kind of verbs, they could complement this method.

Secondly, the detection of zero pronouns has been developed for the investigation of the convergence universal [8], since the number of zero pronouns a text has is correlated to its degree of explicitation, which is referred to as one of the universal features of translation [3].

The method created [8] to measure the number of zero pronouns in original and translated texts is very similar to the baseline of our algorithm, although our methodology is improved by the set of constraints which reduce the false positives introduced by the baseline.

Finally, the genres under observation in this paper are different from the ones studied before, since in [17] a technical handbook (Blue Book, 15,571 words) and journalistic texts (Lexesp 9,745 words) are used, and non-annotated medical and technical texts (six corpora containing a total of 8,127,416 tokens) are employed in the studies of translation universals [9].

## 4 Corpora

The experiments make use of a new corpus created for this purpose named the Z-corpus. The corpus is based on selected texts from three genres: legal, instructional and encyclopaedic. It has been enlarged from its previous version [29] and some annotation errors have been corrected. The Z-corpus is under development and it is currently being annotated by a second annotator. However, this data set (a total of 1202) is relatively large compared to those used in previous work [17] whose evaluation data consisted of only 734 zero pronouns.

The legal section of the Z-corpus contains the Spanish Constitution (from the beginning up to article 110), the first book of the Civil Law Code (up to article 10 in Chapter III), the Highway Code (up to Chapter II, article 27), the Penal Code (up to Chapter II) and the Gender Equality Code (up to article 6). The instructional part is composed of four handbooks taken from the open source Wikibooks: Chemistry, Sewage Engineering, Relativity Theory, Artificial Intelligence and Time Management. Thirdly, the encyclopaedic section is made up of 152 Wikipedia articles about mammals, medicine, linguistics and countries. Samples containing roughly equivalent numbers of zero pronouns were collected for each genre.

The Z-corpus was parsed with the dependency parser for Spanish Connexor's Machinese<sup>4</sup>, which does not identify zero pronouns [7], as is the case for the rest of the Spanish parsers examined.

<sup>4</sup> <http://www.connexor.eu/technology/machinese/>.

Each zero pronoun was annotated manually by adding an xml tag containing informative attributes. The tag was included preferably at the beginning of the clause unless this position produced an ungrammatical result. The information contained in its attributes is the following:

1. The position of the zero pronoun in the sentence.
2. The antecedent of the zero pronoun.
3. The dependency head (the clause verb) on which the zero pronoun depends.
4. The kind of sentence in which the zero pronoun appears.
5. The kind of clause where the zero pronoun stands.
6. The inter-annotator agreement score for the instance.
7. The zero pronoun is cataphoric (yes or no).
8. The antecedent of the zero pronoun corresponds with the "title" of an encyclopaedic entry or legal article (yes or no).

Point 8 turned out to be fairly significant: 168 out of 1202 zero pronouns find their antecedent in the title<sup>5</sup>. One example of a zero pronoun tag would be:

```
<ZERO_PRONOUN id='w2440.5' ant='w2419'  
depend_head='w2441' agreement='high'  
sentence_type='sub' title='yes'>
```

## 5 Rule-based Method

The algorithm for identification of Spanish zero pronouns is composed of a baseline and a set of rules. The baseline is an absolute filter which detects all the potential zero pronoun candidates (positions between tokens). The set of ordered rules are exceptions to this baseline and are applied to reduce the size of the set of candidates by discarding the false positives provided by the baseline. The candidates left over are the set of zero pronouns identified. When a zero pronoun is detected, the clause is marked as having a zero pronoun.

The objective of this study is to determine the extent to which this set of exceptions improves the baseline. The rules combine different sources of linguistic evidence: morphological, syntactic, structural and lexical. The lexical, syntactic and morphological information is provided by the parser; the semantic information is supplied by word lists which share one or more semantic features; and the structural information is contained in the rules themselves.

### Baseline:

If a clause contains a finite verb  
and it has no subject depending on this verb  
there is a potential zero pronoun.

<sup>5</sup> More information about the annotation criteria and the possible values of each xml attribute was detailed in previous work [29].

There is a total of 27 exceptions for the baseline. These exceptions are subsequently conditioned on a set of constraints for the sake of catching the impersonal constructions without zero pronouns included in the first set of candidates provided by the baseline.

Some studies consider different levels of Spanish impersonality (semantic and syntactic impersonality [19]) or distinguish several semantic grades in impersonality [23]. The impersonal clauses considered are the ones with zero subjects referred to in the literature as “natural impersonal clauses” [2] or “syntactic impersonal clauses” [18].

The restrictions placed upon the set of candidates catch a number of impersonal examples which are grouped in the following set of patterns:

1. Natural phenomena [6]:

- (n) Ø Hace mucho calor. Ø Es primavera. Ø Está nublado [13]. Ø Llovió cerca de 1.500ml por metro cuadrado  
*[It] is very warm. [It] is Spring. [It] is cloudy. [It] rained almost 1.500ml per square metre.*

2. Temporal expressions with verbs such as “ser” [13] and “haber” [12] among others:

- (o) Ø Es pronto [13].  
*[It] is early.*
- (p) La película de Ø hace dos años.  
*The film [that was] released two years ago.*

3. Existential use of verb “haber” [11]:

- (q) En un kilogramo de gas Ø hay tanta materia como en un kilogramo de sólido [13].  
*In a kilogram of gas [there] is the same amount of mass as in a kilogram of solid.*

4. Impersonal constructions with modal verbs [18, 30]. More specific examples of this rule would be “haber” + “que” or “poder” + “que”:

- (r) Para determinar una fórmula Ø hay que tener en cuenta la fórmula empírica y el peso molecular.  
*To determine a formula, [it] is needed to consider the empirical formula and the molecular weight.*

5. Impersonal constructions with auxiliary verbs such as “ser” and “estar” [13, 18]:

- (s) Ø Son las dos de la tarde. [13].  
*[It] is two o'clock in the afternoon.*

6. Impersonal expressions with locative and the type of verbs such as “sobrar”, “bastar” or “faltar”. [10, 31]:

- (t) Ø Basta con tres sesiones [13].  
*[It] is enough with three sessions.*

7. Pronominal unipersonal verbs with subject zero such as “tratarse de” [20, 13]:

- (u) Deberán adoptar las precauciones necesarias para su seguridad, especialmente cuando Ø se trate de niños.  
*Necessary measures should be taken, specially when [it] concerns children.*

8. Spanish impersonal constructions of the type “es suficiente”, “es bastante” or “esta bien” [20, 13, 18]:

- (v) Ø Ya es suficiente.  
*[It] is enough.*

9. Fixed constructions such as “ir para” + temporal expression, “es que”, “es para” or “es decir”:

- (w) El peso es una fuerza, Ø es decir, una cantidad vectorial.  
*Weight is a force, [that] is, a vectorial quantity.*

10. Spanish impersonal constructions with “se” [24, 23, 32]:

- (x) Ø Se estará a lo que establece el apartado siguiente.  
*[It] will be follow what is established in the next section.*

Three rules from our system are stated below. Each rule is part of a set of rules which corresponds with patterns 1, 3 and 9 respectively.

— RULE from set 1; Temporal expressions:

FOR every clause with no subject:

IF [[there is a verb that conjugated in third person singular contains a lemma 'ser'] OR  
 [a lemma 'parecer'] OR  
 [one dependent auxiliary verb pointing at one of these verbs]] AND  
 [[it is followed by 0 to 3 tokens] AND  
 [the tokens belong to the same clause]] AND  
 [it is followed by a member IN Temporal expressions list\*]:  
 THEN The clause has no zero pronoun

\* Temporal expressions list

lemma 'tarde',	lemma 'pronto',
lemma 'temprano',	lemma 'primavera',
lemma 'verano',	lemma 'otono',
lemma 'invierno',	lemma 'enero',
lemma 'febrero',	lemma 'marzo',
lemma 'abril',	lemma 'mayo',
lemma 'junio',	lemma 'julio',
lemma 'agosto',	lemma 'septiembre',
lemma 'octubre',	lemma 'noviembre',
lemma 'diciembre',	lemma 'lunes',
lemma 'martes',	lemma 'miercoles',
lemma 'jueves',	lemma 'viernes',
lemma 'sabado',	lemma 'domingo',
lemma 'mediodia',	
lemma 'de' + lemma 'dia' in next line,	
lemma 'de' + lemma 'noche' in next line,	
lemma 'de' + lemma 'tarde' in next line,	
text 'la' + morpho 'card. number' in next line,	
text 'las' + morpho 'card. number' in next line	

— RULE from set 9; unipersonal verbs:

```
FOR every clause with no subject:
IF [[there is a 'se'] OR
   ['Se' in text]] AND
[[this is followed by a verb that conjugated
in third person singular contains a lemma
'tratar'] OR
 [one dependent auxiliary verb pointing
at this verb]] AND
[[this is followed by 0 to 3 tokens] AND
 [the tokens belong to the same
clause]]AND
[this is followed by text 'de']
  THEN The clause has no zero pronoun

ELIF [[there is a 'de'] OR
      ['De' in text]] AND
[[this is followed by text 'que se'] AND
 [[this is followed by a verb which conjugated
in third person singular and this verb has a
lemma 'tratar'] OR
 [one dependent auxiliary verb pointing
this textitverb]]:
  THEN The clause has no zero pronoun
```

— RULE from set 10; Impersonal “se”:

```
FOR every clause with no subject:
IF[there is a token in nominative] AND
[[this token contains a 'se'] OR
 ['Se' in text]]:
  THEN The clause has no zero pronoun
```

## 6 Evaluation

Evaluation of the identification method was carried out separately for the baseline as well as for the combination of the baseline together with the proposed rules.

It should be noted from the outset that the results are influenced by the parser’s accuracy. For example, Connexor’s Machine parser does not detect every verb correctly (y):

- (y) The noun “mejoras” (*improvements*) is parsed as a verb in the following:

También se introducen mejoras en el actual permiso de maternidad.

*There are also added some improvements in the current maternity leave.*

```
<token id="w2589"><text>mejoras</text>
<lemma>mejorar</lemma>
<tags><syntax>@MAIN</syntax>
<morpho>V IND PRES SG P2</morpho></tags>
</token>
```

Moreover, the parser does not tag every subject in the clause (z) successfully:

- (z) The subject is not detected in:

Las unidades de significado en la semántica léxica se denominan unidades léxicas.

*The meaningful units in lexical semantics are called lexical units.*

Standard evaluation measurements in the identification of zero pronouns are recall and precision rates. Evaluation was carried out for the verbs considered to have a zero pronoun. All the zero pronouns are annotated in the Z-corpus and therefore the study benefits of a gold standard for a reliable evaluation of the methods.

Baseline method	Precision	Recall	F-measure
Instructional	0.54	0.88	0.65
Legal	0.39	0.79	0.52
Encyclopaedic	0.39	0.74	0.50
Z-corpus	0.44	0.80	0.56

**Table 1:** *Baseline evaluation*

Table 1 shows the results for the baseline methodology. A higher recall rate was expected given the fact that the baseline takes into consideration a larger data set of verbs with zero pronoun candidates.

Throughout the Z-corpus the baseline has a 0.56 f-measure rate, having its highest rate for the instructional domain (0.65) with a precision of 0.54 and a recall of 0.88.

The rule-based method improves these rates to a small degree, as it is shown in Table 2. The precision and f-measure values register better results compared to the baseline, reaching up to 0.46 and 0.57 respectively. More positive results were found in the instructional genre with a precision of 0.56 and a recall rate of 0.87. However, the recall remains the same.

Rule-based method	Precision	Recall	F-measure
Instructional	0.56	0.87	0.67
Legal	0.40	0.79	0.53
Encyclopaedic	0.40	0.74	0.51
Z-corpus	0.46	0.80	0.57

**Table 2:** *Rule-based evaluation*

To sum up, the zero pronoun identification method studied in this paper has high recall and low precision. This is valuable since it is possible to filter out the unwanted cases retrieved. The rules described improve the system and may help in the identification of more general restrictions in order to obtain a better rate of precision and recall.

## 7 Conclusions and Future Work

This paper presents an improved version of a corpus annotated for zero pronouns. It has clear advantages over others of its type with regard to the number of zero pronouns annotated and the variety of genres. Additionally, the Z-corpus could be applied for the investigation of other research topics once it is available online. The next version of Z-corpus will present an inter-annotator agreement score.

Moreover, the paper describes and evaluates a new methodology in the identification of zero pronouns. This method complements previous methodologies since a large part of our method is dedicated to the detection of impersonal structures and clauses.

Future work will be focused in overcoming parsers’ errors and improving the methodology adding not only

new restrictions but also preferences in order to catch the undetected examples.

In addition, a machine learning approach to tackle this topic will be applied as soon as the Z-corpus is able to offer enough training data. Furthermore, the methodology presented in this study can be used in conjunction with the findings of Spanish zero pronouns distribution [29] towards a zero pronoun resolution algorithm.

## Acknowledgments

We would like to express our gratitude to Prof. Ruslan Mitkov for his supervision throughout this study. We are also especially thankful to Prof. Igor Mel'čuk for his feedback on our previous research by providing invaluable suggestions. Ana Suárez Fernández, PhD student on Spanish syntax, pointed us important studies in the domain and helped us in the analysis of the impersonal sentences. Finally, many thanks are also due to our grammatical and mathematical conscience, Pablo Suárez García, who combed the text for errors of expression and exposition.

## References

- [1] L. Alonso-Ovalle and F. D'Introno. *Hispanic Linguistics at the Turn of the Millennium*, chapter "Full and Null Pronouns in Spanish: the Zero Pronoun Hypothesis", pages 189–210. Somerville, MA: Cascadia Press, 2001.
- [2] D. Antas. *El análisis gramatical*. Barcelona: Ediciones octaedro, 2007.
- [3] M. Baker. *Text and Technology: In Honour of John Sinclair*, chapter "Corpus Linguistics and Translation Studies: Implications and Applications", pages 233–250. Amsterdam and Philadelphia: John Benjamins, 1993.
- [4] J. M. Brucart. *La elisión sintáctica en español*. Barcelona: Universitat Autònoma de Barcelona, 1987.
- [5] J. M. Brucart. *Gramática descriptiva de la lengua española*, 2, chapter "La elipsis", pages 2787–2863. Madrid: Espasa-Calpe, 1999.
- [6] A. Calzado Roldán. La impersonalidad de los verbos metereológicos: una explicación pragmático-discursiva. *Dicenda*, 18:85–108, 2000.
- [7] Connexor. *Machine Language Model*. 1997–2006.
- [8] G. Corpas Pastor. *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation Vol. 49. Frankfurt am Main: Peter Lang, 2008.
- [9] G. Corpas Pastor, R. Mitkov, N. Afzal, and V. Pekar. Translation universals: do they exist? a corpus-based nlp study of convergence and simplification. In *8th AMTA conference*, pages 75–81, 2008.
- [10] O. Fernández Soriano. On impersonal sentences in spanish: Locative and dative subjects. *Cuadernos de lingüística del Instituto Ortega y Gasset*, V:43–68, 1998.
- [11] O. Fernández Soriano. Two types of impersonal sentences in spanish: Locative and dative subjects. *Syntax, Blackwell Publishing*, 2-2:101–140, August 1999.
- [12] O. Fernández Soriano and G. Rigau i Oliver. Construcciones temporales no impersonales en español. In *Estudios de lingüística del Español (ELiEs), Actas del II Congreso de la Región Noroeste de Europa de la Asociación de lingüística y Filología de América Latina (ALFAL)*, volume 22, 2005.
- [13] O. Fernández Soriano and S. Táboas Baylín. *Gramática descriptiva de la lengua española*, 2, chapter "Construcciones impersonales no reflejas", pages 1631–1722. Madrid: Espasa-Calpe, 1999.
- [14] A. Ferrández, A. Palomar, and L. Moreno. El problema del núcleo del sintagma nominal: ¿elipsis o anáfora? *Procesamiento del lenguaje natural*, 20:13–26, 1997.
- [15] A. Ferrández, A. Palomar, and L. Moreno. Anaphor resolution in unrestricted texts with partial parsing. In *Annual Meeting of the ACL, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 385–391, 1998.
- [16] A. Ferrández, A. Palomar, and L. Moreno. An empirical approach to spanish anaphora resolution. *Machine Translation*, 14-3/4:191–216, 1999.
- [17] A. Ferrández and J. Peral. A computational approach to zero-pronouns in spanish. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL2000)*, pages 166–172, 2000.
- [18] L. Gómez Torrego. *La impersonalidad gramatical: descripción y norma*. Arco Libros, Madrid, 1992.
- [19] L. Gómez Torrego. *La enseñanza de la lengua en la década de los noventa*, chapter "Impersonalidades, generalizaciones, encubrimientos y metonimias en la gramática", pages 37–58. Madrid: Universidad Autónoma de Madrid, 1994.
- [20] L. Hernando Cuadrado. Sobre la expresión de la impersonalidad. In *ASELE. Actas IV*, Marrakech, Morocco, 1994. Asociación de Profesores de Español como Lengua Extranjera (ASELE).
- [21] A. Kratzer. More structural analogies between pronouns and tenses. In *The Proceedings of Semantics and Linguistic Theory VIII (SALT VIII)*. Cornell University: CLC Publications, 1998.
- [22] I. Mel'čuk. *Aspects of the Theory of Morphology*. Mouton the Gruyter, Berlin, New York, 2006.
- [23] A. Mendikoetxea. *Las Construcciones con "se"*, chapter "La Semántica de la Impersonalidad", pages 239–267. Madrid: Visor, 1994.
- [24] A. Mendikoetxea. *Gramática descriptiva de la lengua española*, 2, chapter "Construcciones con se: Medias, pasivas e impersonales", pages 1575–1630. Madrid: Espasa-Calpe, 1999.
- [25] R. Mitkov. *Anaphora resolution*. London: Longman, 2002.
- [26] J. C. Moreno Cabrera. Tipología de la catáfora paratáctica: entre la sintaxis del discurso y la sintaxis de la oración. *Estudios de lingüística*, 3:165–192, 1985.
- [27] J. Peral and A. Ferrández. *Lecture Notes In Computer Science*, volume 1835, chapter "Generation of Spanish Zero-Pronouns into English", pages 252–260. London, UK: Springer-Verlag, 2000.
- [28] RAE. *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa Calpe, 1977.
- [29] L. Rello and I. Illisei. A comparative study of spanish zero pronoun distribution. In *Bulag 33: International Symposium on Data and Sense Mining Machine Translation and Controlled Languages, and their application to emergencies and safety critical domains*, pages 209–214. Presses universitaires de Franche-Comté (Presses-ufc), July 2009.
- [30] G. Rigau i Oliver. Los predicados impersonales relativos en las lenguas románicas. *Revista española de lingüística*, 29-2:317–356, 1999.
- [31] M. L. Rivero. Spanish quirky subjects, person restrictions, and the person-case constraint. *Linguistic Inquiry, Massachusetts Institute of Technology*, 2004.
- [32] M. Suñer. Demythologizing the impersonal "se" in spanish. *Hispania*, 59-2:268–275, 1976.