Automatic Detection of Gender and Number Agreement Errors in Spanish Texts Written by Japanese Learners

Maria del Pilar Valverde Ibañez
Faculty of Foreign Languages
Aichi Prefectural University
1522-3 Ibaragasama, Nagakute-shi
Aichi, 480-1198, Japan
valverde@for.aichi-pu.ac.jp

Akira Ohtani
Faculty of Informatics
Osaka Gakuin University
2-36-1 Kishibe-minami, Suita-shi
Osaka, 564-8511, Japan
ohtani@ogu.ac.jp

Abstract

This paper describes the creation of a grammar to automatically detect agreement errors (gender and number) in Spanish texts written by Japanese learners. The grammar has been written using the Constraint Grammar formalism (Karlsson et al., 1995), and uses as input the morphosyntactic analysis provided by the Spanish parser HISPAL (Bick, 2006). For developing and testing the grammar, a learner corpus of 25,000 words has been manually annotated with agreement error tags. Both the grammar and the data from the corpus serve us to draw some conclusions about the characteristics of agreement errors in Japanese learners' Spanish.

1 Introduction

In this paper we describe the creation of a grammar to automatically detect agreement errors -in gender and number- in Spanish texts written by Japanese learners.

Automatic detection of grammatical learner errors can be used for the automatic annotation of learner corpora and for the creation of intelligent computer-assisted language learning systems (Heift and Schulze, 2007). Such tools can benefit both teachers -who will be able to study learner errors and the language acquisition process more systematically- and learners -who can foster their language learning with the help of automatic tools and improved traditional language materials-.

There are two reasons why we focus on agreement errors. First, for Japanese students, agreement is a problematic aspect for learning Spanish and indeed agreement errors are significantly more frequent among Japanese learners than among speakers of other languages (Fernández, 1997). Second, agreement errors in texts can be identified and corrected straightforwardly by a native speaker, unlike other type of errors like article and preposition usage, for example, where annotator agreement may be problematic.

While there is a substantial research on detecting grammatical errors in Learners' English, Spanish has received little attention, probably because of the lack of freely available large learner corpora (Lozano, 2009). For the construction of the grammar, we have manually annotated with agreement error tags a fragment of 25,000 words from the CORANE learner corpus (Mancera et al., 2001) and to control false positives of the grammar, we have also used native corpora: 22,000 words for development amd 12,000 words for test.

The paper is organized as follows. Section 2 deals with the characteristics of gender and number agreement in Spanish and the coverage of the grammar, section 3 deals with the development phase (the corpus, grammar formalism and design principles), section 4 gives the results and analysis of the evaluation, section 5 studies the data in the learner corpus and section 6 presents the conclusions.

2 Gender and number agreement in Spanish

Agreement, defined as the condition of having the same number or gender, serves to relate and identify

26th Pacific Asia Conference on Language,Information and Computation pages 299–307

lexically and syntactically the agreeing words.[1] In Spanish, for a structure to be grammatically correct, the inflecting words involved in a head-dependent syntactic relation must agree in gender and number.

Nouns can be classified into two categories, masculine and feminine, and the gender of the noun determines the gender of its dependendents. Here follow some examples of agreement between a noun and an adjective (1 to 4).

(1) Coche pequeño.
car.MASC small.MASC.SING
'small car'

(2) Coches pequeños.
cars.MASC small.MASC.PLUR
'small cars'

(3) Bicicleta pequeña.
bicycle.FEM small.FEM.SING
'small bicycle'

(4) Bicicletas pequeñas.
bicycles.FEM small.FEM.PLUR
'small bicycles'

(5) chiisai kuruma.
small car
'small car'.

(6) chiisai jitensha.
small bicycle
'small bicycle'.

Examples 1 and 2 show gender and number agreement with a masculine noun, while 2 and 3 show agreement with a feminine noun. Since Japanese does not have number agreement, 1 and 2 correspond to 5 in Japanese, and 3 and 4 correspond to 6. As for gender agreement, the Spanish adjective "pequeño" ('small') changes its ending to agree with a masculine noun in 1 and 2 and a feminine noun in 3 and 4, while in Japanese the noun nor the adjective have gender (the adjective "chiisai" is the same in 5 and 6).

Our grammar contains rules to detect gender and/or number errors. With regard to gender, in Spanish the following word classes have gender: determiners, nouns, pronouns, adjectives and participle verbs. Our grammar checks the following gender agreement cases:

1. Agreement within the noun phrase: between the head (noun or pronoun) and its dependents (the determiner, the adjective and the past participle).
2. Agreement within the clause:
   (a) Between the subject and the subject complement (adjective or past participle) in attributive clauses.
   (b) Between the subject and the past participle verb in passive clauses.

As for number, the previous word classes in addition to the verb have number. Our grammar checks the following number agreement cases:

1. Agreement within the noun phrase: between the head (noun or pronoun) and its dependents (the determiner, the adjective and the past participle).
2. Agreement within the clause:
   (a) Between the subject and the verb.
   (b) Between the subject and the subject complement (adjective or participle) in attributive clauses.
   (c) Between the verb and the subject complement (adjective or participle) in attributive clauses.
   (d) Between the subject and the past participle verb in passive clauses.
   (e) Between the indirect object (prepositional phrase) and the dative pronoun.

3 Development

3.1 The learner corpus

Given the lack of annotated learner corpus, for the development and test of the grammar we have created a manually annotated 25,000 words corpus, extracted from the CORANE learner corpus (Mancera et al., 2001). Our corpus contains 133 Spanish

[1]This relation could be achieved by other linguistic means, specially by the fixed order of words. For example, in Spanish the systematic anteposition and contiguity of the article with the noun in the noun phrase makes agreement between them redundant.

texts written by 47 Japanese native speakers studying Spanish, and it has been divided into two parts, as shown in table 1: 15,000 words for development, corresponding to learners with a level A2 to B1;[2] and 10,000 words for testing, corresponding to learners with a level B2 to C1.

Language level	Learners	Texts	Words
Development			
A2	2	7	1,105
B1	19	90	13,947
Total	21	97	15,052
Test			
B2	9	18	4,758
C1	17	18	5,321
Total	26	36	10,079
Corpus	47	133	25,131

Table 1: Learner corpus: development and test. Language level, number of learners, texts and words.

The annotation/evaluation process has been carried out by one native speaker. Although it would have been desirable to involve more than one annotator in order to report inter-annotator agreement, we believe that the error type treated here shows very high reliability (inter-annotator disagreement may be limited to lapses in concentration), unlike other type of errors like article or preposition usage, which are likely to be much less reliable (Tetreault and Chodorow, 2008).

The error tag appended to the word not only identifies the error but also provides a straightforward correction; since gender and number have only two possible values, there is no possibility of confusion -a masculine token with a gender error tag should be feminine, a singular token with a number error tag should be plural, and so on-.

To control false positives of the grammar, in addition to learner corpora we have also used native corpora: 22,000 words for development and 12,000 words for test, extracted from the Spanish section of the Europarl Parallel Corpus (Koehn, 2005).

[2]A level = basic user, B = independent user and C = Proficient user, according to the Common European Framework of Reference for Languages

3.2 Grammar formalism

Different techniques have been used in the literature to detect different error types (made by learners of English) (Leacock et al., 2010): for errors that require large amounts of contextual information, like preposition and article errors, statistical approaches seem particularly advantageous, while for more local errors, like over-regularized inflection, a rule-based approach seems to work quite well. Error detection systems for learners of languages other than English are scarce, as in the case of learner corpora.

To write our grammar we have use the Constraint Grammar (CG) formalism (Karlsson et al., 1995), which has already been used to detect grammatical errors in other languages: Swedish (Arppe, 2000; Birnn, 2000), Norwegian (Johannessen et al., 2002) Catalan (Badia et al., 2004) and Basque (Uria et al., 2009).

To be able to detect agreement errors, a variable amount of linguistic information is needed: since agreement can occur both at the clause-level and at the phrase-level (as seen in section 2), more syntactic information is needed to resolve the former than the latter. Our grammar uses as input the morphosyntactic analysis provided by the Spanish parser HISPAL (Bick, 2006), which provides us with a full syntactic analysis of sentences (in constituents and syntactic functions) and is error-tolerant, that is, it is capable of parsing (correctly or not) sentences containing grammatical errors.

CG is basically a disambiguation and information mapping methodology designed to operate on token-based grammatical tags that can be added, removed or changed in an incremental and context-sensitive fashion. In a CG rule, a context condition (in parenthesis) contains an obligatory position marker, consisting of a number indicating relative distance in tokens. The default (positive number) is a right context, while a negative number indicates a left context. For example, the following rule adds a plural tag (%agr-p) to a singular noun (NP-HEAD-S) if it is immediately preceded (-1) by a definite determiner (DET-DEF), which is immediately preceded by preposition ”de” (PRP-DE), which is immediately preceded by the word ”uno”).

```
ADD (%agr-p) TARGET NP-HEAD-S
(-1 DET-DEF LINK -1 PRP-DE LINK -1 ("uno"));
```

This rule will assign the tag ”%agr-p” (plural agreement) to the noun ”hombre” (man) in the following fragment that contains an agreement error:

(7) Uno de los hombre
One of the man
’One of the men’

3.3 Design principles

In the design of our grammar our main aim is to achieve a high precision, keeping false positives to a minimum, even at a noticeable loss of recall, following the common practice in grammatical error detection applications.

Even though we can use the full syntactic analysis provided by the HISPAL parser as input, we have written our rules using as low-level information as possible, that is, morphological information, instead of higher-level information like syntactic function, whenever possible. The reason for this is an interesting problem found during the construction of rules: on the one hand, it is necessary to have as much grammatical information as possible about the text we are going to analyse; on the other hand, it is difficult to have such information because even though the parser always provides a syntactic analysis, it is hard to parse a text with grammatical errors correctly, and the errors of the parser may cause our grammar to fail -even for a native speaker it can be hard to parse and understand some fragments of learner language-.

Another decision that has to be made, both during the manual annotation of the corpus and the design of the grammar is, given two (or more) words syntactically related, which word determines the correct gender and number. That is, if we find for example a masculine determiner followed by a feminine noun (or a singular determiner followed by a plural noun), we know that there is a disagreement but we also want to know which is the correct gender (or number) from the native point of view.

For gender, we consider that the syntactic head determines the gender of the dependents (whether the gender of the head is correct or not).[3] Therefore, within the noun phrase, the noun or pronoun determines the gender of the other words. Within the clause, the subject determines the gender of the subject complement.

For number, in most cases the head of the syntactic dependency determines the number of the dependent. However, this is not as straightforward as with gender. In the following cases, the right number is not given by the head of the syntactic dependency, instead:

1. The subject ”gives” the number to the verb.
2. In copulative sentences, the subject gives the number to the verb and to the subject complement. (When there is no subject, the verb determines the number of the subject complement.)
3. Inherently plural determiners (e.g. numerals) give the number to the noun.
4. The indirect object gives the number to the dative clitic.

3.4 Construction of the rules

For the manual construction and refinement of the rules we have looked at example sentences with errors from the annotated learner corpus (a fragment of 15,000 words, corresponding to texts written by learners with a level A1 or B1, as seen in table 1). In addition to that, to control false positives we have also used a 22,000 words native corpus, extracted from the Spanish section of the Europarl Parallel Corpus (**?**). Finally, our grammar contains 31 rules to detect gender agreement errors and 50 rules to detect number agreement errors.

4 Evaluation

We have evaluated the performance of the grammar with a fragment of 10,000 words from the learner corpus (approximately 5,000 words from level B2 and 5,000 words from level C1, as seen in table 1). The results are shown in tables 2 for gender (64.52% precision and 71.43% recall) and 3 for number (58.62% precision and 31.48% recall).[4]

[3] We do not treat here the wrong assignment of gender to words but only the wrong agreement. If the learner makes a mistake choosing the gender of the noun but its complements agree with it, there is no agreement error.

[4] Precision is calculated by dividing the number of true positives (tp) by the number of (true (tp) or false (fp)) positives (Precision = tp / (tp + fn).

Level	tp	fp	fn	Precision	Recall
B2	25	14	13	64.10%	65.79%
C1	15	8	3	65.22%	83.33%
Total	40	22	16	64.52%	71.43%

Table 2: Gender agreement. Grammar results in the test part of the learner corpus.

Level	tp	fp	fn	Precision	Recall
B2	10	9	13	52.63%	43.48%
C1	7	3	24	70.00%	22.58%
Total	17	12	37	58.62%	31.48%

Table 3: Number agreement. Grammar results in the test part of the learner corpus.

To analyse false alarms (that is, false positives and false negatives) with more detail, following (Uria et al., 2009), we have classified them into 4 types:

1. Spelling errors: the text contains a spelling error (which may cause the parser to provide an erroneous input to our grammar).
2. Structural errors: the words in the text are correctly written but the structure contains some error –different from an agreement error-.
3. Parser errors: the words and structure do not contain a learner error -different from an agreement error- but the parser provides a wrong analysis (usually wrong word class).
4. ”Real” errors: None of the above, the grammar fails detecting or non detecting an agreement error.

4.1 Gender

The main source of false alarms in detecting gender disagreement are parser errors: because of the fact that the words do not agree in gender, or simply because of the limitations of the parser, sometimes the words do not receive the correct morphosyntactic interpretation, which makes the grammar fail.

Table 4 shows the frequency of the causes that make the grammar fail, and its precision and recall taking into account only ”real” false alarms.

Recall is calculated by dividing the number of true positives (tp) by the number of true positives (tp) plus false negatives (fn) (Recall = tp / (tp+fn).

False positive	B2	C1	Total
Spelling	0	6	6
Structure	2	0	2
Parser	6	2	8
Real	6	0	6
Total	14	8	22
False negative			
Spelling	1	0	1
Structure	0	0	0
Parser	9	2	11
Real	3	1	4
Total	13	3	16
”Real” precision	80.65%	100.00%	86.96%
”Real” recall	89.29%	93.75%	90.91%

Table 4: Gender agreement. Grammar results in the test part of the learner corpus taking into account only ”real” false alarms.

4.1.1 Recall: false negatives

As for false negatives, the parser provides a wrong analysis of the word in 11 cases. In 7 of them, an article -followed by a noun with different gender values- is analysed instead as a pronoun by the parser.

False negatives also inform us about some phenomena that were not treated by our grammar, and should be addressed in the future:

1. Agreement between the subject and the subject complement -in attributive clauses- when the subject complement is a noun.
2. Agreement between the subject and the subject complement in non-attributive clauses.
3. Agreement between the object and object complement.
4. Agreement between the accusative clitic and the object.
5. Agreement between the relative pronoun and its antecedent.
6. Agreement between the noun and its coordinated dependents (when the noun is comple-

mented by two coordinated adjectives, such adjectives should both agree in gender with the noun.)

7. Agreement across the prepositional phrase boundaries (which requires solving pp-attachment): when a noun is complemented by a prepositional phrase and an adjective, we need to know which noun the adjective depends on (the noun inside the prepositional phrase or the head noun) to determine its correct gender.

4.1.2 Precision: false positives

As for tagger errors, for example, in B2 texts the complex word ”carne=picada” (”minced meat”) appears 4 times analysed as a masculine noun instead of a feminine noun, and thus our grammar detects a (false) disagreement with the article.

With regard to the behaviour of the grammar in the 12,000 words native corpus, our grammar has flagged only 10 false positives (and no true positive).

4.2 Number

The main source of false alarms in the detection of number agreement errors are not learner or parser errors but the design of the grammar itself. Table 5 shows the frequency of the causes that make the grammar fail, and its precision and recall taking into account only ”real” false alarms. As we can see, recall is still considerably low, so our grammar needs to be improved to detect more disagreement contexts.

4.2.1 Recall: false negatives

As we see in table 5, there is a clear difference in the performance of the grammar depending on the language level: in C1 level texts, recall is specially low. This is due to the fact that the higher the language level, the more syntactically elaborated the learner errors, and thus the more difficult for our grammar to detect them safely. As we can see in table 6,[5] the percentage of errors that occur at the clause-level (as opposed to the phrase-level) increases with the language level.

[5]Level A2 texts are excluded because of their low frequency (they contain only 9 errors, 6 at the clause-level and 3 at the phrase-level).

False Positive	B2	C1	Total
Spelling	3	0	3
Structure	1	0	1
Parser	2	2	4
Real	3	1	4
Total	9	3	12
False negative			
Spelling	0	0	0
Structure	0	0	0
Parser	4	0	4
Real	9	24	33
Total	13	24	37
Precision	76.92%	87.50%	80.95%
Recall	52.63%	22.58%	34.00%

Table 5: Number agreement. Grammar results in the test part of the learner corpus taking into account only ”real” false alarms.

Phrase-level errors are those in which the head and the dependent are within the same constituent (the determiner and the noun in the noun phrase, the noun and the adjective in the noun phrase, and so on.), while clause-level errors are those in which the head and the dependent are in different constituents (the subject and the verb, the subject and the subject complement, the object and the object complement, and so on.). Phrase-level errors are easier to detect automatically than clause-level errors, because the latter require a full syntactic analysis or even more to be detected.

	B1	B2	C1
Phrase-level	45.65%	45.45%	32.26%
Clause-level	54.35%	54.55%	67.74%

Table 6: Number agreement. Percentage (and frequency) of phrase-level and clause-level errors by language level.

Therefore, among false negatives, there is still room for improvement for our grammar. Table 7 shows the frequency of some syntactic phenomena in the test part of the corpus where false negatives occurred.

Number 3, 4, 5 and 6 type of agreement have in common the fact that there is a distance between the words involved in the agreement; to identify such agreement errors we need a full sentential analysis with syntactic function information (4, 5) or even the reference of pronouns within or between sentences (3, 6), which is difficult due to the fact that agreement is one of the clues used to identify such relationships. We consider these kind of number agreement errors are specially difficult to detect.

Number 1, 2, 7 , 8 and 9 type of agreement have in common the fact that in those structures, the subject is confused with the direct object by the learner (because the subject occupies a non canonical position or works like a direct object from the semantic point of view) and because of that it is a assigned the wrong number feature. To detect these kind of errors, we need to solve the ambiguity between the subject and the object. Considering that in Spanish the subject can be (usually is) ellided, detecting such errors would require identifying the explicit subject or the referent of the ellided subject safely, which is considerably difficult to achieve, too.

To sum up, even though number agreement errors have a low recall, it is rather difficult to improve recall significantly because to detect such errors we would need a safe full sentential analysis, identifying the referent of the pronouns or the reference of the ellided subject.

Constituents that agree in number	B2	C1
1) Subject-Unaccusative verb	1	3
2) Impersonal verb-*Direct object	1	2
3) Relative Subject-Verb (not 3)	0	3
4) Subject/Verb-Subject complement	1	0
5) Object-Object complement	1	0
6) Indirect object – Clitic	1	0
7) Postposed subject - Verb (not 3.)	0	1
8) Subject-Verb in a clause with “se”	0	1
9) Subject-Verb with “gustar”-like verbs	0	1
Total	5	11

Table 7: Number agreement. Analysis of false negatives.

4.2.2 Precision: false positives

Out of the 29 flagged errors by the grammar, there have been 12 false positives. However, 3 of them were due to misspellings in the learner corpus, 1 to syntactic errors, 4 to parser errors, and 4 of them are ”real” false alarms due to the grammar.

In the 12,000 words native corpus, our grammar has flagged 15 agreement errors, from which 1 is a true positive, and the rest are false positives. Although agreement errors are typical in learner corpora, native corpora also contains such kind of errors because being an inflecting language, the last letters of the word reveal its gender or number value, so a spelling or typing mistake can easily lead to an ”agreement” error.[6]

5 Data from the learner corpus

We can use the data in the annotated corpus not only to develop and evaluate the grammar but also to draw some conclusions about gender and number agreement among Japanese learners.

The 25,000 words fragment contains 154 number agreement errors and 171 gender agreement errors, distributed by language level as table 8 shows. We can see that the frequency of gender errors per word decreases as language level increases, while the frequency of number errors does not show a clear pattern.

Tag	A2	B1	B2	C1	Total
Sing	1	20	4	8	33
Plur	8	72	18	23	121
Total	9	92	22	31	154
% word	0.81	0.66	0.46	0.65	0.61
Masc	5	35	20	9	69
Fem	11	67	17	7	102
Total	16	102	37	16	171
% word	1.45	0.73	0.78	0.34	0.68

Table 8: Learner corpus: frequency of error tags by language level. Number: Singular (Sing) or Plural (Plur). Gender: Masculine (Masc) or Feminine (Fem).

With the evaluation of the grammar and this data, it is clear that number errors need more attention. When dealing with agreement, teachers of Spanish as a foreign language and students usually focus on

[6] In (Bustamante and León, 1996)’s native Spanish 70,000 words error annotated corpus (errors including spelling, structural and non structural errors) 18.5% of errors consist on agreement errors in gender, number or person.

gender, considered the most difficult type of agreement to learn, probably because from the beginning, it requires much effort for the learner to know which is the inherent gender value of every noun than to choose the right number value depending on the context (although there are some morphological hints, gender is arbitrary and must be memorized). However, among Japanese learners, gender errors tend to decrease as the language level increases, while number agreement errors are considerably frequent even among advanced students. In the evaluation of the grammar and in the corpus we have confirmed that number agreement requires a higher level of syntactic analysis than gender agreement: while gender errors occur mainly within the noun phrase, number errors move from the phrase-level to the clause-level as students proficiency increases, affecting distant constituents of the clause or requiring the distinction between syntactic and semantic object.

6 Conclusions and future work

In this paper we have presented a grammar for the detection of agreement errors in Spanish learners texts. Gender error has a precision of 64.52% and recall of 71.43%, and number errors have a precision of 58.62% and recall of 31.48%.

The comparison with other work in the area is particularly difficult, since unlike other NLP areas, grammatical error detection systems do not have a shared corpus or task upon which to evaluate. Although work on different languages is hardly comparable, we can refer to other rule-based systems like (Fliedner, 2002) who detects noun phrase agreement errors in German with precision and recall scores of 67%, and (Gill and Lehal, 2008) error detection system for Punjabi with recall at 76.8% for modifier and noun agreement errors and 87.1% on subject-verb agreement errors.

During the construction of the rules we have tried to find a balance between the necessity of using an input text with as much syntactic information as possible and the fact that parser errors in learner texts are more frequent, which will make the error grammar fail. By writing safe rules we have given priority to precision over recall.

In the gender part of the grammar the main source of false alarms are parser errors, usually a wrong word class tag. If we only take into account the false alarms attributed to our grammar the precision would be 86.96% and recall 90.91%.

In the number part of the grammar the main source of false alarms is the design of the grammar itself, and not learners' or parser errors. Number errors can happen at the phrase level and at the clause level, but as the learners' proficiency progresses, they are more common at the clause level, and thus more difficult for the grammar to detect them because a full syntactic analysis (with information about syntactic functions, the pronoun referent, or the referent of the elided subject, for example) would be required. Therefore, although the grammar's recall is rather low, we consider it is very difficult to improve it without lowering its precision.

By examining the manually annotated learner corpus we have used for the development and test of the grammar, we have confirmed that teachers and learners should pay more attention to number errors: gender errors are more frequent in the beginning but, as students' proficiency increases, gender errors decrease and number errors increase. The type of number errors also change, from phrase-level errors to clause-level errors in the most advanced language group.

The automatic detection of learner errors can contribute to language teaching and learning in several ways: the automatic annotation of corpora with error information, automatic detection of errors in intelligent computer assisted language learning systems and the design of improved learning materials based on corpus data, among others.

The construction of the grammar has served us to confirm the validity of our approach and to gain expertise in the writing of the rules for error detection. As future lines of research, we would like to treat a more challenging error type like article usage, specially prevalent among Japanese students. That will require a more elaborated annotation of corpus and the use of more categories for the evaluation of the system.

Finally, after the evaluation of our grammar, we would like to evaluate the usefulness of the system for language learners in a real language learning context.

Acknowledgments

This work was supported by *kakenhi* (22820047), Grant-in-Aid for Scientific Research (Start-up) from the Japan Society for the Promotion of Science.

References

A. Arppe. 2000. Developing a grammar checker for swedish. In *Proceedings of the 12th Nordic Conference in Computational Linguistics*, pages 13–27.

T. Badia, A. Gil, M. Quixal, and O. Valentín. 2004. Nlp-enhanced error checking for catalan unrestricted text. In *Proceedings of the fourth international conference on Language Resources and Evaluation*, pages 1919–1922.

E. Bick. 2006. A constraint grammar-based parser for spanish. In *Proceedings of TIL 2006 - 4th Workshop on Information and Human Language Technology*, Ribeirão, Preto.

J. Birnn. 2000. Detecting grammar errors with lingsoft's swedish grammar checker. In *Proceedings of the 12th Nordic Conference in Computational Linguistics*, pages 28–40, Norway.

F. R. Bustamante and F. S. León. 1996. Gramcheck: A grammar and style checker. In *The proceedings of the 16th International Conference on Computational Linguistics*, pages 175–181.

S. Fernández. 1997. *Interlengua y análisis de errores en el aprendizaje del español como lengua extranjera*. Edelsa.

G. Fliedner. 2002. A system for checking np agreement in german texts. In *Proceedings of the Student Research Workshop at the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 12–17, Philadelphia.

M. S. Gill and G. S. Lehal. 2008. A grammar checking system for punjabi. In *Proceeding of the 22nd International Conference on Computational Linguistics (COLING)*, pages 149–152.

T. Heift and M. Schulze. 2007. *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge Studies in Computer Assisted Language Learning.

J. B. Johannessen, K. Hagen, and P. Lane. 2002. The performance of a grammar checker with deviant language input. In Association for Computational Linguistics, editor, *Proceedings of the 19th international conference on Computational Linguistics*, pages 1–8.

F. Karlsson, A. Voutilainen, J. Heikkilä, and A. Anttila. 1995. *Constraint grammar. A language-independent system for parsing unrestricted text*. Mouton de Gruyter, Berlin/New York.

P. Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit 200*.

C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault. 2010. Different approaches for different errors. In *Automated Grammatical Error Detection for Language Learners*. Morgan & Claypool.

C. Lozano. 2009. Cedel2: Corpus escrito del español l2. In *Applied Linguistics Now: Understanding Language and Mind*, pages 197–212.

A.M. Cestero Mancera, I. Penadés Martínez, A. Blanco Canales, L. Camargo Fernández, and J.F. Simón Granda. 2001. Corpus para el análisis de errores de aprendices de E/LE (CORANE). In Ana Gimeno Sanz, editor, *Actas del XII Congreso Internacional de ASELE: tecnologías de la información y de las comunicaciones en la enseñanza de la E/LE*, pages 527–534. Asociación para la Enseñanza del Español como Lengua Extranjera. Congreso Internacional.

J. Tetreault and M. Chodorow. 2008. Native judgments of non-native usage: Experiments in preposition error detection. In *Proceedings of the Workshop on Human Judgments in Computational Linguistics at the 22nd International Conference on Computational Linguistics (COLING)*, pages 24–32.

L. Uria, B. Arrieta, A. Díaz de Ilarraza, M. Maritxalar, and M. Oronoz. 2009. Determiner errors in basque: Analysis and automatic detection. *Procesamiento del Lenguaje Natural*, (43):41–48.