Extend English grammar to test for Subject verb number agreement and Noun adjective gender and number Agreement for Spanish

Carlos Perez
07/02/2017

Problem

For this report I wanted to explore the problem of extending a simple english grammar so it can detect number and gender agreement for sentences in Spanish without creating a new one. Some considerations must be taken into account the grammar of Spanish is inflectionally rich, it allows sentences with null subjects and the allows that subjects go after the verbs. . In the same manner nouns and adjectives have specific genders that must agree with each other. Also as there is an important population of speakers of spanglish in the united states and as of 2009 there were no tools that used a spanglish grammar it becomes interesting to explore the subject of exploring the extension of a spanish grammar from an english one, so in the future it is possible to combine the rules of both and create an spanglish grammar.

Relevant Studies

VIGLIOCCO and BUTTERWORTH claim that subject-verb agreement is one of the most widespread phenomenons of agreement across languages and that there are two possible ways of understanding the problem one is unification the other is feature selection. Also they propose the differences of the two languages as the following tables show.

Some features of spanish

Rich verbal inflectional system

Spanish	English
Yo vivo	I live
Tu vives	You live
El vive	He lives
Nosotros vivimos	We live
Ustedes viven	You live
Ellos viven	They live

Null subjects

Tengo hambre	(I) am hungry
Ojalá vengan	Let's hope (they) come
Parece una jovencita	(She) looks like a young lady

Postverbal subjects

Andaba yo caminando	Walking I was
Jugaban ustedes	Played you

Gender agreement

EL árbol es viejo	The tree is old

La rocas es vieja	The trees are old

Franco and Solorio describe Spanglish as the simultaneous use or alternating of both Spanish and English, commonly used in U.S. populations. As of 2009 there were no automated voice systems that recognized spanglish dialect.

As described in the paper by Bermejo and Lara the pronoun Ustedes is expanding in certain regions of Spain and is the standard way of speaking in most spanish-speaking countries of America(Continent)

Claim

The claim is that it is simple to implement spanish number and gender agreement from an english grammar that detects number agreement. By having the same identifiers for both grammars we could in the future implement a system that detects Spanglish.

Evidence

In order to be able to detect agreement in a spanish grammar for a particular set of sentences we must take into account the rules described above. In the case of the Spanish inflectional system it is worth noting that the easiest way to transform a given english grammar into an spanish grammar is to expand the existing rules to cover all the subjects. For example

For the English grammar that has the following rules to signify number agreement.

Given a regular grammar with the following set of rules to deal with agreement in English.

```
s --> np(Number), vp(Number).
np(Number) --> det(Number), n(Number).
vp(Number) --> v(Number).
n(singular) --> [dog].
n(plural) --> [dogs].
v(singular)-->[grumbles].
v(plural)-->[grumble].
```

In order to adapt the grammar for Spanish subject verb agreement we must expand it by adding a new non terminal pr that correspond to pronouns and add extra rules to deal with the six different persons that exist in the spanish language.

```
s--> pr(Number),vs(Number).
np(Number) --> pr(Number).
```

pr(spasingone) --> [yo]. pr(spasingtwo) --> [tu]. pr(spasingthree) --> [el]. pr(spaplurone) --> [nosotros].	vs(spasingone) --> [como]. vs(spasingtwo) --> [comes]. vs(spasingthree) --> [come]. vs(spaplurone) --> [comemos].

pr(spaplurtwo) --> [ustedes]. pr(spaplurthree) --> [ellos].	vs(spaplurtwo) --> [comen]. vs(spaplurthree) --> [comen].

With these implementation the grammar is able now to detect that the conjugation is correct in spanish and given that we extended the existing grammar it is still able to detect number agreement in english as the rules as separate.

```
1 ?- s([yo,como],[]).
true .

2 ?- s([ustedes,comen],[]).
true .

3 ?- s([yo,comen],[]).
false.
```

The case of the null subject is different as in the English language it is mandatory. To allow the detection of these sentences in spanish given the initial grammar we must add a rule that expresses that a sentence can be only composed of a verb in spanish and a noun in spanish. We must make at this point the distinction between verbs and nouns in the two languages to avoid sentences with verbs in english and objects in spanish to be identified as correct.

s--> vs(Number),ns.
ns--> [papa].
vs(spasingone) --> [como].
…

With the addition of these rules the program now detects sentences in spanish that have only the conjugated verb and a noun. Additionally it detects as incorrect sentences in english with no subject.

```
1 ?- s([eats,papa],[]).
false.

2 ?- s([comen,papa],[]).
true .

3 ?- s([como,papa],[]).
true .

4 ?- s([grumbles,papa],[]).
false.

5 ?- s([likes,dog],[]).
false.
```

For the case of gender agreement between noun and adjectives noun phrase for spanish identifier was created as the determinant and adjective rules for spanish. We must note that this gender agreement is only between the nouns and and adjectives and is completely different from subject verb agreement as we can have singular subject-verb agreement and plural masculine determinants noun adjective agreement.

s--> nps(Number),vs(Number).
s--> vs(Number),nps(Number).
nps(Number) --> pr(Number).
nps(Number) --> ns(Number).
nps(Number) --> dets(Num),ns(Num),adjs(Num).

dets(spasing masc)--> [el]. dets(spasingf	ns(spasingmas c) --> [color]. ns(spasingfem)	adjs(spasi ngmasc) -->

em)--> [la]. dets(spaplur masc)--> [los]. dets(spaplurf em)--> [las].	--> [marea]. ns(spaplurmas c) --> [colores]. ns(spaplurfem) --> [mesas].	[oscuro]. adjs(spasi ngfem) --> [rosada]. adjs(spapl urmasc) --> [oscuros]. adjs(spapl urfem) --> [rosadas].

```
1 ?- s([comemos,las,mareas,rosada],[]).
false.

2 ?- s([comemos,las,mareas,rosadas],[]).
false.

3 ?- s([comemos,la,marea,rosada],[]).
true .

4 ?- s([come,la,marea,rosada],[]).
true .

5 ?- s([come,el,color,rosada],[]).
false.

6 ?- s([come,el,color,oscuro],[]).
true
```

Discussion

The original grammar was taken from a laboratory exercise for the class of Natural Artificial Intelligence of Professor Jean Louis Dessalles at ParisTech Telecom. The program written in Prolog was modified to allow for Spanish number and gender agreement detection. Given the resulting implementation we can see that it is possible to extend an english grammar to detect number and gender agreement in spanish. However several new terminals and non terminals had to be created to make the distinction between this two grammars which may one wonder if the effort of extending a grammar is similar to the one from creating a new grammar from scratch. It is interesting to note that if we don't differentiate clearly the rules for the two languages we can create a grammar that detects sentences in spanglish that do not belong to either of these. For example "Let's go de rumba" or "Me olvidé las llaves en el fridge". This can be interesting in developing applications that deal with immigrants, expatriates and bilingual speakers.

Bibliography

VIGLIOCCO, Gabriella; BUTTERWORTH, Brian; GARRETT, Merrill F. Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 1996, vol. 61, no 3, p. 261-298.

BERMEJO, Víctor Lara. El avance de" ustedes" y las concordancias de tercera persona en el español hablado en la provincia de Málaga. *Analecta Malacitana (AnMal electrónica)*, 2013, no 34, p. 185-197.

FRANCO, Juan Carlos; SOLORIO, Thamar. Baby-steps towards building a Spanglish language model. En *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer Berlin Heidelberg, 2007. p. 75-84.