Analysis and Interpretation of Negative Queries in Spanish to a Geographical Database

Rogelio Dávila Pérez

Computer Science Department, The University of Texas at El Paso,
El Paso, Texas, U.S.A., 79968-0518
rdav@cs.utep.edu
Centro de Investigación CENTIA, Universidad de las Américas, Puebla,
Cholula, Pue., México
Laboratorio Nacional de Informática Avanzada LANIA,
Jalapa, Ver., México

Contact Author: Computer Science Department, University of Texas at El Paso, El Paso, Texas 79968-0518, phone: (915) 747-6950, FAX: (915) 747-5030, E-mail: rdav@cs.utep.edu, rdav9@hotmail.com

Keywords: Computational Semantics, Natural Language Processing, Database Systems, Knowledge Base Systems, Formal Semantics.

Abstract: Negative sentences play an important role in communication; their implementation in natural language query systems deserve some discussion, as there are many forms in which negation can be introduced in natural language sentences, e.g. at the noun level, the verb phrase level or the sentence level. The contribution of this paper is an analysis and interpretation of negative queries in Spanish to a geographical database. The negative forms considered are actually implemented in the system NATLIN [3] in the form shown in the present document.

Analysis and Interpretation of Negative Queries in Spanish to a Geographical Database

Rogelio Dávila Pérez

Computer Science Department, The University of Texas at El Paso,
El Paso, Texas, U.S.A., 79968-0518
rdav@cs.utep.edu
Centro de Investigación CENTIA, Universidad de las Américas, Puebla,
Cholula, Pue., México
Laboratorio Nacional de Informática Avanzada LANIA,
Jalapa, Ver., México

Abstract. Negative sentences play an important role in communication; their implementation in natural language query systems deserve some discussion, as there are many forms in which negation can be introduced in natural language sentences, e.g. at the noun level, the verb phrase level or the sentence level. The contribution of this paper is an analysis and interpretation of negative queries in Spanish to a geographical database. The negative forms considered are actually implemented in the system NATLIN [3] in the form shown in the present document.

1 Introduction

The development of natural language query systems for retrieving information from databases and knowledge bases is an important goal of Artificial Intelligence and natural language processing. Several projects have been developed for English, but only few papers have considered Spanish as the main language for retrieving information [1, 2, 3]. In this paper an implementation of a query system to a geographical database, in which information is accessed in Spanish is presented. Emphasis is on the analysis and implementation of negative queries. In the following section, a brief introduction to a grammar for a subset of Spanish is presented. In Section 3, the main forms of negative queries are discussed and some proposals are introduced for implementing a solution. Section 4 shows some conclusions.

2 Grammar for a subset of Spanish

Following the tradition initiated by Montague [4], we define a categorical grammar for our subset of Spanish. In the system, syntax and semantics are stated in a one-to-one correspondence, so every syntactic rule is reinterpreted semantically providing for

a compositional analysis. On this particular case I will use the notation of a generative grammar which is more traditional in the linguistics context, just a small subset taken from the system NATLIN [3] is presented.

Syntactic Rules	Semantic Rules
1. S → NP VP	S' = NP' (VP')
2. NP → PN	NP' = PN'
3. NP → Det Noun	NP' = Det ' (Noun')
4. NP → IP Noun	NP' = IP' (Noun')
5. VP → Vt NP	VP' = λz.((NP' (z)) (VP'))
6. Noun → Noun Adj	Noun' = λw.(Noun' (w) & Adj' (w))
7. Noun → Noun RP	Noun'= λz.((Noun' (z)) & (RP' (z)))
8. RP → que VP	RP'= VP'
9. Det → algún(os)	Det'= some' = = λP.(λQ.(exists(x, P(x) & Q(x))))
10. IP' → Qué	IP'= which' = λP.(λQ.(wh(x, P(x) & Q(x))))
11. Adj → Europeo	Adj'= european' = λu.(europeo(u))
12. PN → Australia	PN' = [Australia]'= λP.(P(australia))
13. PN → El_Mediterráneo	PN' = [Mediterranean] '= λP.(P(mediterráneo))
14. Noun → país(es)	Noun'= country'= λy.(país(y))
15. Vt → colinda(n)	Vt'= borders' λx.(λy.(colinda(x,y)))

3 Analysis of negation in Spanish

Spanish is a language rich in constructions and consequently it has various forms of negation. In this section we consider some of the more common and basic ones, even though their inclusion will increase considerably the expressiveness of the subset of language we are choosing to access the geographical database.

1. Negation at the noun phrase level. An important form of negation is introduced by the negative determiner 'ningún' (no, none) which expresses the emptiness in the set of objects defined by the noun that follows it. This determiner is the opposite of the 'todos' (forall) determiner and combines with nouns and adjectives in the noun phrase, e.g.

a). ¿**Ningún** país colinda con Australia?
(**No** country borders with Australia?)

b). ¿**Ningún** país asiático colinda con Australia?
(**No** Asian country borders with Australia?)

In order to provide a semantic analysis for the 'ningún' determiner we consider the interpretation of 'algún' (some) and 'todos' (forall) which are expressed as follows,

Algún ==> λP.(λQ.(exists(x, P(x) & Q(x))))
Todos ==> λP.(λQ.(forall(x, P(x) → Q(x)))) ≡ λP.(λQ.(~exists(x, P(x) & ~Q(x))))

The interpretation of 'ningún' can be paraphrased as 'there are no x such that P(x) and Q(x) hold' and is expressed as follows,

Ningún ==> λP.(λQ.(~exists(x, P(x) & Q(x))))

Hence, for the two noun phrases in the sample sentences presented above, we will get the following interpretations respectively:

a). λQ.(~exists(x, país(x) & Q(x)))
b). λQ.(~exists(x, (país(x) & asiático(x)) & Q(x)))

This interpretation will be achieved if we add the following rule to the grammar presented on Section 3:

9a. Det → ningún Det'= none'= λP.(λQ.(~exists(x, P(x) & Q(x))))

2. Negation at the verb phrase level. Another form of negation is generated through the introduction of the word 'no' (not) which traditionally is placed before the verb in the sentence and generally applies to the whole verb phrase. Some examples follow,

a). ¿Que países **no** colindan con el Mediterráneo?
(Which countries do **not** border with the Mediterranean?)

b). ¿Que países europeos **no** colindan con aquellos países que colindan con el Mediterráneo?
(Which European countries do **not** border with any country, which borders with the Mediterranean?)

To interpret properly the previous sentences, we propose the introduction of the following rules to the grammar presented on Section 3:

5a. VP → Neg VP VP'= Neg' (VP')
14a. Neg → no Neg'= no'= λP. (λz.(~P(z)))

Applying these rules to the analysis of the verb phrases of the sample sentences introduced above, we obtain the interpretations shown bellow:

a). λz.(~colinda(z, mediterráneo))
b). λz.(~exists(y, país(y) & colinda(y, mediterráneo) & colinda(z, y)))

3. Negation at the adjectival level. The word 'no' which modifies the verb phrase negating its contents, may also be introduced at a position just before an adjective in the sentence, negating then the property expressed by the adjective, e.g.

¿Que países **no** europeos colindan con el Mediterráneo?
(Which (**not**) European countries border with the Mediterranean?)

In order to deal with this construction we add the following rule to the grammar presented in Section 3:

11a. Adj → Neg Adj Adj'= Neg' (Adj')

Applying this rule we get the interpretation of the sentence presented above as:

wh(x, país(x) & ~europeo(x) & …)

4. Negation at the level of the sentence. The negation of a whole sentence is generally achieved by introducing the words: 'No es verdad que …' (it is not true that …) just before the sentence, e.g.

No es verdad que Australia colinda con El Mediterráneo.
(**It is not true that** Australia borders with the Mediterranean.)

In this case what we need is to negate the whole formula generated by the semantic interpretation of the sentence. We accomplished this introducing by the following rule to the grammar in Section 3:

2a. S → no_es_verdad_que S S'= it_is_not_true_that' S'= ~(S')

For the sentence introduced above we have the following interpretation:

~(colinda(australia, mediterráneo))

5. The problem with Spanish double negation. Spanish inherited double negation from Latin. Formally speaking double negation presents a problem to compositionality as shown bellow. Take as an example the following sentence,

¿Que países **no** colindan con **ningún** país?
(Which countries do **not** border with (**no**) country?)

If we apply to this sentence the rules we introduced before in this section, we get its interpretation as follows:

Wh(x, país(x) & ~(~exists(y, país(y) & colinda(x,y))))

which, after simplifying the double negation, we get

Wh(x, país(x) & exists(y, país(y) & colinda(x,y)))

The problem is that this is not the meaning intended by the speaker of the sentence. A native speaker of the language interprets the sentence as meaning 'which are the countries that do not border any country' whose interpretation is presented bellow.

Wh(x, país(x) & ~exists(y, país(y) & colinda(x,y)))

The only proposal we have to solve the problem is to notice that this kind of construction appears only at the level of the verb phrase, hence it is possible to introduce a restriction on the application of Rule 5a for preventing its use if the verb phrase interpretation contains a negation as the first symbol of the formula.

4 Conclusions

In the present paper an analysis of negation in some sentential forms of Spanish is presented and a proposal for implementing a solution is discussed within the context of a geographical database system. Further work on this area should involve other forms of negation, such as the one introduced by the adverb 'solamente' (just, only) in sentences as: 'solamente España colinda con Portugal' (only Spain borders with Portugal) which means something like 'no country but Spain borders with Portugal.'

References

1. V. Dahl: Translating Spanish into Logic through Logic. *American Journal of Computational Linguistics*, 13:149-164, (1981).
2. R. Dávila-Pérez: Access in Spanish to a Predicate Calculus Database: M. Sc. Dissertation: University of Essex, England (1986).
3. R. Dávila, Osorio M., Quintero M.: Interfaz en Español para Recuperación de Información en una Base de Datos Geográfica: To appear in the Proceedings of the CONIELECOMP 2002, Acapulco, Mexico, February (2002).
4. R. Montague: The Proper Treatment of Quantification in Ordinary English: In: R. H. Thomason, editor: *Formal Philosophy. Selected Papers of Richard Montague*: Yale University Press, New Haven (1974).