

Phrase Structure Grammars for the Expression of Vague Concepts in Spanish

A. Sobrino^a & J.A. Olivas^b

^a Area de Lógica y Filosofía de la Ciencia. Fac. Filosofía.
Univ. Santiago de Compostela, Campus Universitario s/n
15706 - Santiago (Spain) *e-mail: lftgalex@usc.es*

^b Dpto. de Ingeniería Informática. ETSI Informática.
Universidad Antonio de Nebrija. Pirineos, 55.
28040 - Madrid (Spain) *e-mail: jolivas@dii.unnet.es*

Abstract

This paper characterizes grammars generating Spanish sentences of imprecise meaning. The basic grammar generating sentences conforming to the standard paradigm of fuzzy logic (Quantifier + Noun + Verb + Modifier + Adjective) is systematically deformed (by inversion, suppression or addition of nonterminal elements in its production rules) to produce a series of grammars generating grammatical, semantically acceptable semigrammatical or ungrammatical sentences.

1 Introduction

One of the characteristics of natural language is that it contains a large number of expressions of imprecise meaning, such as *few*, *most*, etc. Most logics exclude such expressions, or allow them to be included only if their meaning is made precise. Fuzzy logic is designed precisely to handle this aspect of natural language.

For Aristotle, logical analysis of ordinary language could be adequately based on the subject-predicate structure - 'subject' being understood to refer to the logical subject or theme. The rules of syllogism, for example, assume this structure. Fuzzy logic follows suit except that both subject and predicate are inherently complex entities. The subject usually consists of a noun and an imprecise quantifying adjective (e.g. *few* or *most*), while the predicate usually consists of a verb, an imprecise modifying adverb (e.g. *very* or *slightly*) and an adjective. This structure,

Quantifier + Noun + Verb + Modifier + Adjective

(QNVMA for short) is already to be found in Zadeh's earliest work on fuzzy semantics (2). It may be called the standard fuzzy sentence structure. A typical Spanish sentence susceptible of this kind of analysis is

Pocos suecos son muy bajos
(Few Swedes are very short)

(for simplicity, in what follows the only verb used is, as here, *ser*). Prepositions and definite or indefinite articles can also appear in typical fuzzy sentences, but since they make no contribution to fuzziness they will not be paid special attention in this paper. The standard fuzzy sentence structure, QNVMA, can be considered as an intermediate stage of the generation of *Pocos suecos son muy bajos* by the following canonical grammar:

\mathbf{G}_0

$R_S^0 : S \rightarrow NP + VP$

$R_{NP}^0 : NP \rightarrow Q + [] + N$

$R_{VP}^0 : VP \rightarrow V + M + [] + A$

$R_Q^0 : Q \rightarrow \text{pocos} \mid \text{la mayoría de}$ (few \mid most of)

$R_{[]}^0 : [] \rightarrow \text{los} \mid \lambda$ (the [plural])

$R_N^0 : N \rightarrow \text{suecos}$ (Swedes)

$R_V^0 : V \rightarrow \text{son}$ (are)

$R_M^0 : M \rightarrow \text{muy}$ (very)

$R_A^0 : A \rightarrow \text{bajos} \mid \text{altos}$ (short \mid tall)

where S stands for Sentence, NP for Noun phrase, VP for verb phrase, Q for Quantifier, N for Noun, V for Verb, M for Modifier, A for Adjective and λ for the null string (as usual, " \rightarrow " indicates substitution, and " \mid " separates alternative substituent strings). Note that quantifiers (and in subsequent examples, modifiers too) can be either simple, such as *pocos* (few), or complex, such as *la mayoría de* (most of).

In generative grammar it is customary [1] to distinguish between grammatical sentences (those which fully conform to the grammar of the language, such as *Pocos suecos son muy bajos*), semigrammatical sentences (those which are not strictly grammatical but are semantically acceptable in that they are intelligible to competent speakers, such as *Una idea surgió a mí* instead of the grammatically proper *Me surgió una idea* (One idea came to me)), and ungrammatical sentences (those which are so grammatically incorrect as to be unintelligible, such as *Verde la de muy sol* (Lit. Green the very sun)). In what follows, a number of grammars will be derived from \mathbf{G}_0 by inversion, suppression or addition of nonterminal elements in its production rules. Examination of the grammatical and semantic acceptability of the resulting sentences will show which of the alternative grammars reasonably extend the grammar of fuzzy sentences beyond the standard QNVMA format so as to allow fuzzy semantics to be applied to a wider field of fuzzy discourse than hitherto.

2 Variants of the canonical grammar G_0

2.1 Variants preserving grammaticality

G_1

G_1 is obtained by inversion of NP and VP in R_S^0 :

$R_S^1 : S \rightarrow VP + NP$
 $R_{NP}^1 : NP \rightarrow Q + \square + N$
 $R_{VP}^1 : VP \rightarrow V + M + \square + A$
 etc., as in G_0

Examples:

Son muy bajos pocos suecos
 (Lit. Are very short few Swedes)
Son muy altos la mayoría de los suecos
 (Lit. Are very tall most of Swedes)

It may be noted that the example with *pocos* is less natural than the example with *la mayoría de* because it is unusual to focus on negative concepts.

G_2

Inversion of Q and N in R_{NP}^0 affords G_2 :

$R_S^2 : S \rightarrow NP + VP$
 $R_{NP}^2 : NP \rightarrow \square + N + P + Q$
 $R_{VP}^2 : VP \rightarrow V + M + \square + A$

 $R_P^2 : P \rightarrow,$

Example:

Los suecos, pocos son muy bajos
 (Lit. (as for the) Swedes, few are very short)

G_3

Insertion of N (as an adjectival noun) or Q in R_{NP}^0 affords G_3 :

$R_S^3 : S \rightarrow NP + VP$
 $R_{NP}^3 : NP \rightarrow Q + \square + Q + N \rightarrow Q + N + N[?]$
 $R_{VP}^3 : VP \rightarrow V + M + A$

Example:

La mayoría de los pocos kurdos son muy perseguidos
 (Most of the few Kurds are very persecuted)
Pocos finlandeses lapones son muy morenos
 (Few Lapp Finns are very dark)

where *lapones* is an adjectival noun.

G_4

Replacement of $[\]$ by M or A in R_{VP}^0 affords G_4 :

$R_S^4 : S \rightarrow NP + VP$
 $R_{NP}^4 : NP \rightarrow Q + \square + N$
 $R_{VP}^4 : VP \rightarrow V + M + M + A \rightarrow V + M + A + A$

Examples:

La mayoría de los españoles son casi seguramente morenos

(Most of Spaniards are almost certainly dark)

La mayoría de los españoles son muy moreno oscuros

(Most of Spaniards are very deep dark)

G₅

Repetition of the *N* or *Q* in R_{NP}^0 (i.e. addition of an N_{echo} or Q_{echo} that must be realized by the same terminal strings as the *N* or *Q* that are already present) affords **G₅**:

$R_S^5 : S \rightarrow NP + VP$

$R_{NP}^5 : NP \rightarrow Q_r + Q_{echo} + N \rightarrow Q + [] + N + N_{echo}$

.....

where Q_r is a repeatable quantifier (see Section 3 below).

Examples:

Pocos pocos suecos son muy bajos

(Few few Swedes are very short)

Pocos suecos suecos son muy bajos

(Few Swedish Swedes are very short)

Note that *suecos suecos* has the sense of “true Swedes” and hence changes the meaning of the original sentence (whereas the repetition of *pocos* in the first example merely reinforces the meaning of the original sentence).

G₆

Repetition of the *M* or *A* in R_{VP}^0 in replacement of the $[]$ affords **G₆**:

$R_S^6 : S \rightarrow NP + VP$

$R_{VP}^6 : VP \rightarrow V + M + M_{echo} + A|V + A + A_{echo}$

.....

Examples:

Pocos suecos son muy muy bajos

(Few Swedes are very very short)

Pocos suecos son bajos bajos

(Few Swedes are short short)

A similar remark as for **G₅** is in order, in that repetition of *muy* merely emphasizes degree, whereas repetition of *bajos* emphasizes authenticity. Note also that repetition of modifiers such as *suficientemente* (sufficiently) or *preferentemente* (preferably) would lead to ungrammatical sentences.

G₇

Suppression of *N* and $[]$ in R_{NP}^0 affords **G₇**:

$R_S^7 : S \rightarrow NP + VP$

$R_{NP}^7 : NP \rightarrow Q$

.....

Example:

Pocos son muy bajos

(Few are very short)

Suppression of the noun from the noun phrase leads to a sentence whose meaning is indeterminate without a context.

G₈

Suppression of M in R_{VP}^0 affords **G₈**:

$$R_S^8 : S \rightarrow NP + VP$$

$$R_{VP}^8 : VP \rightarrow V + A$$

.....

Examples:

Pocos suecos son bajos

(Few Swedes are very short)

La mayoría de los suecos son altos

(Most of Swedes are tall)

In these sentences, the suppression of the modifier has brought about a reduction in imprecision.

G₉

Suppression of Q in R_0^{NP} affords **G₉**:

$$R_S^9 : S \rightarrow NP + VP$$

$$R_{NP}^9 : NP \rightarrow \square + N$$

.....

Example:

Los suecos son muy bajos

(Swedes are very short)

Note that the suppression of the quantifier in the noun phrase reduces imprecision more than the suppression of the modifier in the verb phrase (in **G₈**), doubtless because it leaves an unqualified noun, which in general will have a more precise meaning than the unmodified adjective left by suppression of the modifier.

G₁₀

Suppression of Q in R_{NP}^0 and M in R_{VP}^0 affords **G₁₀**:

$$R_S^{10} : S \rightarrow NP + VP$$

$$R_{NP}^{10} : NP \rightarrow \square + N$$

$$R_{VP}^{10} : VP \rightarrow V + A$$

.....

Example:

Los suecos son altos

(Swedes are tall)

Here, the only imprecision left attaches to *altos*. Replacement of *altos* with an adjective of precise meaning would create a sentence of precise meaning.

2.2 Variants affording semigrammatical sentences

G₁₁

Inversion of V and $M + A$ in R_{VP}^0 affords **G₁₁**:

$$R_S^{11} : S \rightarrow NP + VP$$

$$R_{NP}^{11} : NP \rightarrow Q + \square + N$$

$$R_{VP}^{11} : VP \rightarrow M + A + V$$

.....

Examples:

Pocos suecos muy bajos son

(Lit. few Swedes very short are)
La mayoría de los suecos muy altos son
 (Lit. the most of the Swedes very tall are)

The result of the \mathbf{G}_{11} translocation is that the subject is not followed by the expected verb, which is instead placed at the end of the sentence.

\mathbf{G}_{12}

Inversion of M and A in R_{11}^{VP} affords \mathbf{G}_{12} :

$R_S^{12} : S \rightarrow NP + VP$

$R_{NP}^{12} : NP \rightarrow Q + [] + N$

$R_{VP}^{12} : VP \rightarrow A + M + V$

.....

Examples:

La mayoría de los suecos, altos bastante son

(Lit. the most of the Swedes, tall quite are)

Pocos suecos, bajos más bien son

(Lit. few Swedes, short rather are)

The grammatical equivalents would be *La mayoría de los suecos son bastante altos* and *Pocos suecos son más bien bajos*. Note that without the comma, the second example would be quite ungrammatical because it would be impossible to interpret. Note also that this variant is restricted to certain modifiers (see $\mathbf{G}_{12'}$ below).

2.3 Variants affording ungrammatical sentences

\mathbf{G}_{13}

Inversion of M and A in R_{VP}^0 affords \mathbf{G}_{13} :

$R_S^{13} : S \rightarrow NP + VP$

$R_{NP}^{13} : NP \rightarrow Q + [] + N$

$R_{VP}^{13} : VP \rightarrow V + A + M$

.....

Examples:

Pocos suecos son bajos muy

(Lit. few Swedes are short very)

La mayoría de los suecos son altos muy

(Lit. the most of the Swedes are tall very)

Again, only certain modifiers are allowed to realize M .

$\mathbf{G}_{12'}$

$\mathbf{G}_{12'}$ is identical to \mathbf{G}_{12} except that a different set of modifiers realizes M . Example: *Pocos suecos bajos muy son* (Lit. few Swedes short very are). The radically ungrammatical and semantically unacceptable character of \mathbf{G}_{13} and $\mathbf{G}_{12'}$ derives from the semimorphological nature of certain adverbs.

\mathbf{G}_{14}

Suppression of A in R_{VP}^0 affords \mathbf{G}_{14} :

$R_S^{14} : S \rightarrow NP + VP$

$R_{NP}^{14} : NP \rightarrow C + [] + N$

$R_{VP}^{14} : VP \rightarrow V + M$

.....

Example:

Pocos suecos son muy
(Lit. Few Swedes are very)

3 A variant of the canonical grammar

The \mathbf{G}_0 variants generating grammatical sentences can be combined with each other and with \mathbf{G}_0 itself by introducing disjunctive sequents in the production rules of \mathbf{G}_0 . In order to do so, however, more attention must first be paid to the restrictions on the sets of modifiers and quantifiers that were imposed in certain variants.

Some of the \mathbf{G}_0 variants described above require recognition of the class of repeatable quantifiers, Q_r , exemplified by *poco* and *mucho*. In addition we may distinguish second order quantifiers (Q_2), which, possibly subject to partitive modification, can be followed by other quantifiers (example: *algunos de los* (some of the) in *algunos de los pocos suecos* (some of the few Swedes)). Distinguishing between the singular and plural uses of quantifiers is also useful, but will not be dealt with here.

Other \mathbf{G}_0 variants appearing above require recognition of modifiers M_i that can be permuted with their adjective, as in \mathbf{G}_{12} , or modifiers M_r that can be repeated, as in \mathbf{G}_6 , without loss of grammaticality or semigrammaticality. The M_i include *bastante* and *suficientemente* (and do not include *muy*), while the M_r include *muy* and *poco* (and do not include *más o menos* (more or less)). Furthermore, just as we may distinguish second order quantifiers, so may we also distinguish second order modifiers M_2 (such as *demasiado* (too) in *demasiado poco preciso* (too little precise)).

With the above definitions, some of the grammatical and semigrammatical variants set forth above can be combined in the following variant of the canonical grammar:

$\mathbf{G}_{0'}$

$$R_S^{0'} : S \rightarrow NP + VP$$

$$R_{NP}^{0'} : NP \rightarrow Q + \square + N|Q_r + \square + N_r|Q_i + \square + N_i|Q_2 + Q + \square + N$$

$$R_{VP}^{0'} : VP \rightarrow V + M + \square + A|V + M_r + \square + A|V + M + \square + A_r|V + M_i + \square + A_i|V + M_2 + M + \square + A|V + M_2 + M + \square + A_r$$

4 Discussion

Fuzzy logic has hitherto paid more attention to semantics than to syntax, in the sense that it has been more concerned to provide an adequate representation of sentences with imprecise meaning, and of their logical relationships, than to broaden its syntactic scope beyond the standard QNVMA format. However, the spirit of fuzzy logic - the desire to cope adequately with imprecise parlance - implies that it should also be able to handle imprecise syntax, i.e. the well-formed formulas of a

formal fuzzy language should include what in natural language are semigrammatical sentences. In this paper we have modified the canonical grammar generating the standard QNVMA format so as to allow generation of a variety of other formats typical of imprecise parlance. This has practical implications for the application of fuzzy logic to the construction of interfaces facilitating communication between human beings and computers, since such interfaces must be able to cope with syntactic imprecision as well as with imprecision associated with quantifiers and modifiers.

Acknowledgments

This work was supported by the PB92-0394 grant from the D.G.I.C.Y.T (Ministry of Education, Spain).

References

- [1] Katz, J. J., 'Semi-sentences', in Fodor, J. A. & Katz, J. J. (de.), *The Structure of Language. Readings in the Philosophy of Language*. Prentice-Hall, Inc. Englewood Cliffs, 1964, pp. 400-416.
- [2] Zadeh, L. A., 'Quantitative fuzzy semantics', *Information Sciences*, 3, 1971, pp. 159-176.