

A CASE OF FREQUENCY-BASED MARKEDNESS

Carmen Muñoz
Universidad de Barcelona

This article studies the nature of an error which is common among Spanish intermediate learners of English: the confusion between the masculine singular personal pronoun and the feminine singular form. Data are drawn from written tasks produced by three groups of students, with different levels of proficiency. The errors that appear in the data are investigated in relation to case, level of proficiency and sex of the subjects. Results are also compared with those of a previous study with oral data. The analysis of the results shows that errors significantly consist of an overgeneralization of the masculine form at the expense of the feminine one, and that errors appear mostly in the possessive case. An explanation in terms of frequency-based markedness is discussed both from a linguistic and from a psycholinguistic point of view.*

1. Introduction

The discussion about the nature of second language learners' errors has a long tradition in the field of second language acquisition. On the one hand, researchers have emphasized the influence of the learners' mother tongue—either in the form of transfer errors or of errors produced by the difference in the parametric values between L1 and L2 (James 1980, Gass and Selinker 1983, White 1987)—, and on the other, the importance of developmental errors—errors common to all learners of the same language and independent of the mother tongue (Corder 1967, Richards 1974). It is our contention

* An earlier version of this paper was presented at ESSE/2, September 3, 1993. Our thanks to the students and faculty at the Escola d'Idiomes Moderns of the University of Barcelona.

in this paper that the explanation for certain persistent errors may lie in the interaction between L1, L2, and markedness.

When Jakobson (1968) first applied the notion of markedness to L1 acquisition, he claimed that those values which occur more widely in the world's languages are acquired earlier by children. Cross-linguistic frequency was thus seen as an explication for markedness; the more frequent values were seen as unmarked in opposition to the less frequent or marked values. Since then other criteria have been used to explicate markedness, among them distribution within a single language, structural simplicity (Eckman, Moravcsik and Wirth 1986); core vs peripheral phenomena (White 1986); ease of processing (Givón 1983). In the field of L2 acquisition the role of markedness has mainly been discussed in relation to issues such as learnability (Hyams 1986, White 1986, 1987), developmental sequences (Rutherford 1982, Mazurkevich 1985, Chaudron and Parker 1990), and the influence of the mother tongue (Eckmann 1977, Kellerman 1977, Gass 1979, Zobl 1984).

The general and more basic claim is that markedness may affect the order of acquisition. Furthermore markedness could explain the overproduction of certain L2 structures (Chaudron and Parker 1990), in which case performance could very likely be erroneous.

In this study we have looked at a very common and persistent error among Spanish learners of English: the confusion of the masculine and feminine singular forms of English personal pronouns. We believe that the generic use of the masculine pronoun increases the frequency of occurrence of masculine forms in the input data of learners, and that therefore a markedness relationship can be claimed to hold between masculine and feminine forms. Furthermore previous studies on the teaching of English possessive adjectives to French speaking learners of English by Zobl (1985) have confirmed that *his* is the unmarked member of the *his/her* pair.

In the next section we present the results of previous research on this area (Muñoz 1991), which motivated the present study.

2. A First Attempt towards Identification of the Error

In a previous study (Muñoz 1991) we attempted to observe the nature and frequency of the error, and in particular the influence of case and of natural or grammatical gender. By drawing the data from different

oral texts, we also tried to elucidate whether the kind of reference of the pronouns, that is exophoric or endophoric reference, had some bearing on this issue. The study was conducted with ten subjects, two of whom made no errors of this type; their results were therefore eliminated from the counting. All subjects were students of English Philology at the University of Barcelona —all but one bilingual Spanish/Catalan— with an intermediate to high level of spoken English.

The results confirmed that the only area of difficulty in the pronominal system was the third person singular, and that the errors consisted in the use of a masculine form for a feminine one and viceversa. From a total of 586 third person singular pronouns, masculine and feminine, 61 (9.6%) were incorrect. A chi-squared test showed that the difference between the number of incorrect possessive pronouns and incorrect subject and object pronouns was significant. That is, the number of incorrect possessive pronouns was significantly higher than the number of incorrect subject or object pronouns. This result was not surprising given the difference between the possessive pronoun systems of English and Spanish.

In Spanish (both Castilian and Catalan) the pronoun agrees with the possessed entity in number (all persons, with determinative and independent function), and also in (grammatical) gender, in most cases. Number agreement is illustrated in the following Spanish (Castilian) examples:

Sp.:	nuestro libro	nuestros libros
E.:	our book	our books

Likewise gender agreement is illustrated in the examples below:

Sp.:	nuestro libro	nuestra revista
E.:	our book	our magazine

By contrast, in English the number and gender agreement of the pronoun depends on the possessor, not on the possessed entity, as the previous and the following examples also illustrate:

E.:	his/her/their son	his/her/their daughters
Sp.:	su hijo	sus hijas

But most importantly one particular tendency was found to be statistically significant. There was an overgeneralization of the masculine forms at the expense of the feminine ones. That is, there were many more instances of the use of a masculine form instead of the feminine one,

than the other way round, the use of a feminine form where a masculine form was required. An overgeneralization of the masculine form has been observed in French learners of English (Zobl 1985) and in anglophone children learning French (Tarone, Frauenfelder and Selinker 1976), although to our knowledge it had not been previously reported in studies of Spanish learners of English.

This tendency was, however, independent of the gender that the possessed object had in Spanish in each case.¹ Neither the type of reference (deictic or anaphoric), nor the dominant language (Spanish or Catalan) seemed to be relevant.

In an attempt to shed more light on this issue, a new study was undertaken. First of all it seemed appropriate to try to confirm the tendency of the error with a different and larger group of subjects, and to take into account different levels of proficiency in English. The sex of the subjects was also taken into account in case a difference could be observed in the use of gender by male and female learners of English.

3. Hypotheses

The hypotheses of this research are partly based on the findings from the previous work (hypotheses 1 and 2):

(1) errors will be found in the third person singular forms, where there is variation of gender. The tendency of the error will be an overuse of the masculine form in place of the corresponding feminine form.

(2) errors will appear with significantly higher frequency in the possessive forms, that is, learners will most often overgeneralize *his* at the expense of *her*.

¹ It has been proposed (Grosu 1972, Zobl 1985) that with possessed human entities there can arise a conflict in natural sex between the possessor and the entity possessed, which would not happen with possessed inanimate entities. This conflict would increase the complexity of the possessive rule when it involves possessed human entities. In terms of markedness, based on perceptual complexity in Grosu's framework, possessed human entities would be marked, whereas possessed inanimate nouns would provide the unmarked contexts. The prediction that emerges from this dichotomy, i.e., that there would be more errors when this conflict in natural sex between the possessor and the entity possessed arises, has not been observed in our data, however.

In addition, it is predicted that:

(3) the sex of the subjects may be related to a preference in the use of masculine or feminine forms.

(4) the level of proficiency of the subjects may have an effect on the number of errors produced.

(5) when comparing the present study that uses written data with the previous one with oral data, the medium will not be found to be relevant, insofar as the language used in both studies is spontaneous.

4. Method

4.1. *Subjects*

The subjects in this study were 55 adult Spanish learners of English, 24 males and 31 females, studying on an intensive summer course at the Language School of the University of Barcelona. Of these 18 belonged to a low intermediate level (second year), 22 to an intermediate level (third year), and 15 to a high intermediate level (fourth year). All of them had been thoroughly tested before the course began, and assigned to the appropriate level course.

4.2. *Design*

Subjects were randomly given one of a set of picture booklets. Each booklet contained 6 stories, of four pictures each. They were instructed to describe the pictures and to write their sentences as quickly and as unreflectingly as possible.

The dependent variable of this study is the production of third person singular masculine and feminine personal pronouns. The principal independent variables are level of nonnative proficiency, and sex.

4.3. *Materials*

There was a total of 10 different story sequences, from which sets of six sequences were selected. Each sequence contained four pictures which showed two characters, a main one and a secondary one, doing different things.

Furthermore, in order to produce a similar number of masculine and feminine third person pronouns, there was a similar number of masculine and feminine characters.

5. Results

Table 1 displays the distribution as far as case is concerned of the third person masculine and feminine singular pronouns found in the data (six pronominal forms were eliminated because they involved errors of a different kind, for example *his* for *their*). A chi-square test of association showed that the number of errors was significantly higher in the possessive forms than in the other forms ($\chi^2 = 20.8$, $df = 1$, $p < .001$).

Table 1: Number of Pronouns and Errors per Case

Case	Pro.	Errors	%
Subjective (<i>he, she</i>)	408	11	2.6
Possessive (<i>his, her</i>)	215	26	12.0
Objective (<i>him, her</i>)	40	0	—
Total	663	37	5.5

Since there were no errors in the objective pronouns they were eliminated from all the successive countings.

The production of third person pronouns, subjective and possessive forms, increased with proficiency. The low level group, with 18 subjects, produced a total of 166 pronominal forms, with a mean number of 9.2 pronouns per subject. The middle level group, with 22 subjects, produced 239 pronouns, with a mean number of 10.8 of pronouns per subject, slightly higher than that of the low level group. The high level group, with 15 subjects, produced a total of 218 pronouns, and therefore the average number of pronouns produced by the subjects of this group was the highest: 14.5 per subject.

Table 2 shows the distribution of errors in the data. As can be seen there, males at all levels made more errors in pronoun choice than did females. In addition, males made increasing errors as their instructional level increased.

Table 2: Percentage of Errors per Level Group and per Sex

Group	Sex					
	Male			Female		
	Pro.	Errors	%	Pro.	Errors	%
Low	77	4	5.1	89	3	3.3
Middle	118	12	10.1	121	0	—
High	59	12	20.3	159	6	3.7
Total	254	28	11	369	9	2.4

Certainly, although male subjects presented fewer pronominal forms, they made more errors than female subjects. The difference was found to be significant ($\chi^2 = 12.7$, $df = 1$, $p < .001$).

The tendency of the error is shown in Table 3. As expected, there were more erroneous masculine forms, that is, a masculine pronominal form was used where there should have been the feminine one.² A chi-square test of association showed that the difference was significant ($\chi^2 = 12.7$, $df = 1$, $p < .001$).

Table 3: Percentage of Errors per Gender of Pronouns

Gender	Pro.	Errors	%
Masc.	259	27	10.4
Fem.	364	10	2.7

The tendency of the error was the same for male and female subjects. As shown in Table 4, out of the 28 errors made by males, as many as 20 (71.4 %) were masculine forms that should have been feminine. Likewise out of the 9 errors made by females, 7 (77.7 %) were in masculine forms.

²It is interesting to note at this point that although materials were balanced so that there were similar numbers of masculine and feminine pronouns, those productions used in the final analysis presented a higher frequency of feminine forms. Among the reasons why some of the texts were discarded were failure to describe picture sequences, and use of direct speech.

Table 4: Tendency of the Errors per Sex of Subjects

Sex	Erroneous pronominal forms			
	Masc. forms		Fem. forms	
	Pro.	%	Pro.	%
Male	20	71.4	8	28.5
Female	7	77.7	2	22.2

The direction of the bias was also independent of the case of pronouns, as shown in Table 5. The tendency to overuse masculine forms was similar in both subjective and possessive pronouns.

Table 5: Tendency of the Errors per Case of Pronouns

Case	Erroneous pronominal forms			
	Masc. forms		Fem. forms	
	Pro.	%	Pro.	%
Subj.	8	72.7	3	27.2
Poss.	19	73.0	7	26.9

We can compare these results with those of the previous study. Table 6 shows the percentages of errors by case. We can see that there were more errors in the first study, but their relative distribution remains the same: significantly more errors in the possessive forms than in the subjective forms.

Table 6: Errors per Case Compared

Case	Study 1			Study 2		
	Pro.	Errors	%	Pro.	Errors	%
Subj.	482	35	5.1	408	11	2.6
Poss.	57	23	40.3	215	26	12

Likewise when we compare the tendency of the error, we find that the percentages of overgeneralization of the masculine forms in the two studies are similar. In the case of subjective pronouns 72.7 % in this study (see Table 5 above), versus 74.2 % in the first study. In the case of possessive pronouns 73 % in this study (see Table 5 above), versus an even higher 91.3 % in the first study.

6. Discussion

First, as for the errors found in the third person singular pronouns, the results confirm predictions (1) and (2). Errors were found in the third person masculine and feminine pronominal forms, and particularly in the possessive case. The tendency of the bias was as expected, indicating an overuse of the masculine forms. Besides, the direction of the error was the same, regardless of case.

As for the third prediction, the sex of the subjects was not shown to be significant, as far as the tendency of the error was concerned. Males made more errors than females, however. And although this is not directly relevant to this study, it is not the first time research has shown females performing better in SLA (Farhady 1982, Eisenstein 1982).

Our fourth prediction was concerned with the possible effects of the different proficiency levels. As we have seen there is an increase in the number of pronouns per subject, which is related to an increase in fluency, length of sentences and consequently number of words produced. Surprisingly, there is a parallel increase in the number of errors produced, particularly by the male subjects. We have no explanation for this at the moment, and we would need to continue the research into higher levels of proficiency to see whether and when the number of errors diminishes again. This result does not, however, contradict our hypothesis that frequency-based markedness affects patterns of acquisition, since frequency can be considered to be an incremental factor.

Our fifth hypothesis was concerned with the medium, and thus with a comparison between the results obtained from the written data and those obtained from the oral data of the previous study. We have seen the same tendency in the errors in both studies, and consequently we can say that the tendency observed is characteristic of spontaneous language, independently of the medium used. The higher number of errors in the first study could, however, be an effect of the medium, of the experiment itself.

The results of this new study confirm our predictions and feasibility of an explanation in terms of markedness. The bias seems to be based on an asymmetrical opposition between the two series of forms. The masculine pronoun would be the unmarked member of the opposition, being distributionally more frequent and general (Zobl 1985), whereas the feminine pronoun, more restricted and specific, would be the marked member.

It has been frequently hypothesized that second language learners should learn less marked features more quickly than more marked features (Rutherford 1983, Chaudron and Parker 1990). In our case the less marked forms are used, and overused, much more frequently than the more marked forms.

However it is interesting to note that whereas markedness can satisfactorily explain the earlier acquisition and overgeneralization of the masculine forms, it fails to account for the difference between subjective and possessive forms. We believe the latter could be explained by the interaction between markedness and the mother tongue.

In Spanish topic continuity is not expressed by means of pronominal subjects like in English, and zero anaphora is used instead. Pronominal subjects are only used in case of contrast or emphasis (Muñoz 1988) and are then much more scarce. There is then structural correspondence between the forms, but not functional correspondence. On the other hand, as we have seen, Spanish has a single third person singular possessive form, whereas English has two forms (and number agreement is also marked differently). The lack of congruence between the two systems may have negative effects on acquisition, enhancing the effects of markedness.³

The differences between L1 and L2 could also explain that the error seems particularly persistent among Spanish learners. A significant difference could lie in the obligatoriness of the pronominal subject in English, which increases the frequency of generic uses of the English masculine pronoun. We could hypothesize that this would result in the reinforcement of the unmarked condition of masculine pronouns for a Spanish learner. However, appropriate cross-linguistic data, which is not available at the moment, would be needed to pursue this idea.

To conclude let us briefly reflect on the dimension of the error and its psycholinguistic relevance. Frequency affects the functioning of the lexicon in that frequently used words are easier to retrieve than less frequent words. Then since English masculine pronouns are highly frequent they would reach the activation level required for lexical access before the corresponding feminine pronominal forms. The overgeneralization of masculine pronouns could be explained if they

³ This situation where L1 has one form while L2 has two forms was known as "split" in Contrastive Analysis, where it occupied the highest level of difficulty.

had such a high default level of activation in the lexicon of the learner as a result of continuous use, both in comprehension and in production, that they would not be deactivated enough (de Bot and Schreuder 1993). A higher level of proficiency may free the learner from some cognitive strain, and give her more time to spend on the selection of the right pronominal form.

7. Conclusion

In this study we have provided data relevant to the understanding of a persistent error in the use of the third person singular pronouns by Spanish learners of English, and seemingly learners of English in general. The frequency of a particular linguistic item, that is, frequency-based markedness, has been seen to affect the overextension of the masculine form at the expense of the feminine pronoun.

In this case the frequency of the item is very high within a particular language, English, but also very high cross-linguistically. Therefore the overextension of the masculine form could also be considered to be culturally-induced. And in fact this case of frequency-based markedness could very well fit Comrie's idea according to which "the construction type which is least marked formally is also least marked in terms of properties of the real world" (1986, 86). If there is a correlation between linguistic markedness and real world markedness, the error can only be totally mitigated if the more marked item becomes less marked, both linguistically and extralinguistically. In this sense attempts such as the generic use of the feminine pronoun, the *he/she* forms, or the plural forms as alternatives to the generic use of the masculine pronoun, may be helpful.

At a different, more accessible, level teachers may help in class by focusing instruction on marked elements, which has been shown to have positive effects in related areas (Zobl 1985; see Larsen-Freeman and Long 1991, for a discussion), as well as by increasing the frequency of the marked forms in the input. And this is certainly a situation in which the role of instruction must be seen as relevant, since the mere exposure to natural input may do little to prevent fossilization.

WORKS CITED

- Chaudron, Craig, and Kate Parker. 1990. "Discourse Markedness and Structural Markedness". *Studies in Second Language Acquisition* 12, 43-64.
- Comrie, Bernard. 1986. "Markedness, Grammar, People, and the World". *Markedness*. Eds. Fred R. Eckman, Edith Moravcsik and Jessica Wirth. New York: Plenum. 85-106.
- Corder, S. Pit. 1967. "The Significance of Learners' Errors". *International Review of Applied Linguistics*, V, 161-169.
- De Bot, Kees, and Schreuder, Robert. 1993. "Word Production and the Bilingual Lexicon". *The Bilingual Lexicon*. Eds. Robert Schreuder and Bert Weltens. Amsterdam: John Benjamins. 191-214.
- Eckman, Fred R. 1977. "Markedness and the Contrastive Analysis Hypothesis". *Language Learning*, 27, 315-330.
- Eckman, Fred R., Edith Moravcsik and Jessica Wirth, eds. 1986. *Markedness*. New York: Plenum Press.
- Eisenstein, M. 1982. "A Study of Social Variation in Adult Second Language Acquisition". *Language Learning*, 32, 367-391.
- Farhady, Hossein. 1982. "Measures of language proficiency from the learner's perspective". *TESOL Quarterly*, 16, 43-59.
- Gass, Susan M. 1979. "Language Transfer in Universal Grammatical Relations". *Language Learning*, 29, 327-344.
- Gass, Susan M. and Larry Selinker (Eds.). 1983. *Language Transfer in Language Learning*. Rowley, MA: Newbury House.
- Givón, Talmy. 1983. "Topic Continuity in Discourse: an Introduction". *Topic Continuity in Discourse: a Quantitative Cross-Language Study*. Ed. Talmy Givón. Amsterdam: John Benjamins. (pp. 1-42).
- Grosu, A. 1972. "The Strategic Content of Island Constraints". *Working Papers in Linguistics No 13*, Department of Linguistics, Ohio State University.
- Hyams, Nina M. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: D. Reidel.
- Jakobson, Roman. 1968. *Child Language, Aphasia, and Phonological Universals*. The Hague: Mouton.
- James, Carl. 1980. *Contrastive Analysis*. London: Longman.
- Kellerman, Eric. 1977. "Towards a Characterization of the Strategies of Transfer in Second Language Learning". *Interlanguage Studies Bulletin*, 2, 58-145.
- Kellerman, Eric. 1985. "If at First You Do Succeed...". *Input in Second Language Acquisition*. Eds. Susan M. Gass and Carolyn G. Madden. Rowley, MA.: Newbury House. 345-353.

- Mazurkewich, Irene. 1985. "Syntactic Markedness and Language Acquisition". *Studies in Second Language Acquisition*, 7, 15-35.
- Muñoz, Carmen. 1988. "La referencia al sujeto tópico en el texto castellano y en el texto inglés." *Actas del X Congreso Nacional AEDEAN*, 399-406. Zaragoza, 1986.
- Muñoz, Carmen. 1991. "Why are HE and SHE a problem for Spanish learners of English?" *Revista Española de Lingüística Aplicada*, 7, 129-136.
- Richards, Jack C. (Ed.). 1974. *Error Analysis*. London: Longman.
- Rutherford, William A. 1982. "Markedness in Second Language Acquisition". *Language Learning*, 32, 85-108.
- Rutherford, William A. 1983. "Language typology and language transfer." *Language Transfer in Language Learning*. Eds. Susan M. Gass and Larry Selinker. Rowley, MA.: Newbury House. 358-370.
- Tarone, Elaine, Uli Frauenfelder and Larry Selinker. 1976. "Systematicity/Variability and Stability/Instability in Interlanguage Systems". *Papers in Second Language Acquisition* (Language Learning Special Issue No. 4). Ed. H. Douglas Brown, 93-134.
- White, Lydia. 1986. "Markedness and Parameter Setting: Some Implications for a Theory of Adult Second Language Acquisition". *Markedness*. Eds. Fred R. Eckman, Edith Moravcsik and Jessica Wirth. New York: Plenum Press. 309-327.
- White, Lydia. 1987. "Markedness and Second Language Acquisition: The Question of Transfer". *Studies in Second Language Acquisition* 9, 261-285.
- Zobl, Helmut. 1984. "Cross-Language Generalizations and the Contrastive Dimension of the Interlanguage Hypothesis". *Interlanguage*. Eds. Alan Davies, Clive Criper and Anthony Howatt. Edinburgh: Edinburgh University Press. (pp. 79-97).
- Zobl, Helmut. 1985. "Grammars in Search of Input and Intake". *Input in Second Language Acquisition*. Eds. Susan M. Gass and Carolyn G. Madden. Rowley, Mass.: Newbury House. 329-344.

