

Exploring genderlect markers in a corpus of Nineteenth century Spanish novels

Bermúdez Sabel, Helena

helena.b.sabel@gmail.com

Université de Neuchâtel, Switzerland

Introduction

Gender has been studied profusely in Sociolinguistics due to its relevance in social structure, and the fact that women and men use language differently is largely proven (Eakins and Eakins, 1978; Eckert, 1989; Hobbs, 1990; e.g. Cheshire, 1997). However, gender-based variation regarding literary writing style has received little attention from a sociolinguistics perspective, except for studies whose focus was a specific author or work (e.g. Costello, 1980; Kamuf, 1980; Miller, 1988; Hamdan, 2011). In this paper I will tease out certain gender-specific linguistic features in fiction writing by implementing a stylometric analysis of 86 Spanish novels written between 1840 and 1919.¹ The goal is to test whether trends detected in Contemporary history (post-1945) concerning linguistic change and the role of women are observed in 19th century Spain. The authors belong to a stable sociolinguistic stratification and as such the following sociolinguistic principle is expected: men would use a higher frequency of nonstandard forms than women, unless in change from above, that is, a conscious change with social repercussion, in which women favour the incoming prestige form (Labov, 1990, p.213).

Methodology

I first conducted a quantitative analysis to identify chronolect markers, that is, linguistic elements specific to a given period in comparison to another. The corpus was divided in four periods spanning twenty years each. The presence of distinctive markers of a later period in an earlier one can be interpreted as innovative uses, and likewise, distinctive markers of the earlier periods can be read as conservative uses when present in later periods. Once the chronolect markers were teased out, the next step was to verify the specificity of those elements throughout a gender-based diachronic comparison to attest whether the sociolinguistic principle mentioned above applies.

Firstly, lemma and part-of-speech frequencies were calculated by conducting a set of analyses based on Zeta (Schöch et al., 2018), a measure of distinctiveness whose initial formula was developed by Burrows (2002). This type of measures produces a preferred and avoided list of words by one dataset in comparison to another dataset, and they have already been used in gender-based studies with literary corpora (e.g. Rybicki, 2016; Weidman and O'Sullivan, 2018).

Findings and Discussion

The comparison of the novels written between 1840 and 1859 with those written from 1900 to 1919 reveal that honorifics are a crucial period marker with *vos* being the most distinct word of the first period and *usted* of the later (both forms designate the receiver in a communicative act). Considering these results, I calculated the specificity index (Lafon, 1984) of the conservative form *vos* and the innovative form *usted* in the different subcorpora divided by genre and period. The innovative form *usted* is specific in the first two decades to female writing while *vos* is underrepresented. These findings can be used to explore the linguistic change related to these honorifics, since we can argue that the analysed corpus reflects the expected behaviour: women authors favoured the innovative form earlier than men which suggests that the generalisation of *usted* had already covert prestige before the second half of the 19th century (this means that even if *usted* was a nonstandard form in certain registers, its replacement of *vos* was highly valued). The diachronic study carried out by Blayer (2001) addressing honorifics, based on theatre plays, reports the progressive decline of *vos* in the 19th century which is consistent with the findings of the current study. However, her analysis makes no reference to the gender of either the playwright or the plays' characters.

In addition, the results derived from a gender-based stylometric analysis using Zeta measures can be put in relation with corpus linguistics studies carried out in other languages. There are many similarities between the distinctive semantic fields for women authors found in this Spanish novel corpus with those teased out in the study of English novels (1840s–1990s) by Weidman and O'Sullivan (2018).

Concerning the use of epistemic adverbs, a relation can also be established between the results obtained in this study and those in Pennebaker (2011): he reports the use of what he calls “certainty words” (like *always* or *absolutely*) as a female stylistic trait in English; similarly, *jamás* (‘never’) and *siempre* (‘always’) are overrepresented in the female author corpus of Spanish novels.

Future work

In order to carry out a comprehensive sociolinguistic analysis of the use of *vos/usted*, I would need to establish socio-pragmatic categories (who talks to whom) as in Blayer (2001). However, Blayer (2001) does not differentiate whether the speaker is a man or a woman, thus it is not possible to trace from her results whether authors were implementing a gender-based linguistic characterisation of the speakers. The specific analysis of character language by gender also needs to be addressed in the studied novel corpus.

Notes

1. The analysed corpus corresponds to the Spanish section of the *European Literary Text Collection* (ELTeC-spa, <https://distantreading.github.io/ELTeC/spa/index.html>) excluding American authors.

Bibliography

Blayer, I. M. F. (2001). (1) Las formas de tratamiento en el español peninsular: un estudio diacrónico. *Si-*

gnum: Estudos da Linguagem , 4 (1), pp. 27–58. DOI: 10.5433/2237-4876.2001v4n1p27.

Burrows, J. (2002). ‘Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing* , 17 (3), pp. 267–87. DOI: 10.1093/lc/17.3.267.

Cheshire, J. (1997). Linguistic Variation and Social Function. In Coupland, N. and Jaworski, A. (eds), *Sociolinguistics: A Reader* . Modern Linguistics Series. London: Macmillan Education UK, pp. 185–98. DOI: 10.1007/978-1-349-25582-5_16.

Costello, B. (1980). The “Feminine” Language of Marianne Moore. In McConnell-Ginet, S., Borker, R. and Furman, N. (eds), *Women and Language in Literature and Society* . New York: Praeger, pp. 239–57.

Eakins, B. W. and Eakins, R. G. (1978). *Sex Differences in Human Communication* . Houghton Mifflin.

Eckert, P. (1989). The Whole Woman: Sex and Gender Differences in Variation. *Language Variation and Change* , 1 (3), pp. 245–67. DOI: 10.1017/S095439450000017X.

Hamdan, S. (2011). Identifying the Linguistic Genderlects of the Style of Writing of Arab Male and Female Novelists. *Journal of Education Culture and Society* , 2011 (2), pp. 55–63. DOI: 10.15503/jecs20112-55-63.

Hobbs, D. R. (1990). *Gender-Based Strategies in Issuing Directives in Mexican Spanish* . [Doctoral Thesis Thesis]. Austin, Texas: University of Texas at Austin.

Kamuf, P. (1980). Writing like a Woman. In McConnell-Ginet, S., Borker, R. and Furman, N. (eds), *Women and Language in Literature and Society* . New York: Praeger, pp. 284–99.

Labov, W. (1990). The Intersection of Sex and Social Class in the Course of Linguistic Change. *Language Variation and Change* , 2 (2), pp. 205–54. DOI: 10.1017/S0954394500000338.

Lafon, P. (1984). *Dépouillements et statistiques en lexicométrie* . Genève, Paris: Slatkine.

Miller, N. K. (1988). Women’s Autobiography in France: For a Dialectics of Identification. In Brodzki, B., Schenck, C. and Furman, N. (eds), *Life/Lines: Theorizing Women’s Autobiography* . Ithaca: Cornell University Press, pp. 45–61.

Pennebaker, J. W. (2011). *The Secret Life of Pronouns: What Our Words Say About Us* . Bloomsbury USA.

Rybicki, J. (2016). Vive La Différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies. *Digital Scholarship in the Humanities* , 31 (4), pp. 746–61. DOI: 10.1093/lc/fqv023.

Schöch, C. et al. (2018). Burrows’ Zeta: Exploring and Evaluating Variants and Parameters. In *DH 2018* . <https://dh2018.adho.org/burrows-zeta-exploring-and-evaluating-variants-and-parameters/> [12/11/2022] .

Weidman, S. G. and O’Sullivan, J. (2018). The Limits of Distinctive Words: Re-Evaluating Literature’s Gender Marker Debate. *Digital Scholarship in the Humanities* , 33 (2), pp. 374–90. DOI: 10.1093/lc/fqx017.