

Gender differences in linguistic complexity through time

Freek Van de Velde^{*1}

^{*}Corresponding Author: freek.vandavelde@kuleuven.be

¹Dept. of Linguistics, Research Group QLV, KU Leuven, Leuven, Belgium

A large swath of linguistic studies have documented gender differences in language use, at the phonological, lexico-grammatical and stylistic level, both in spoken discourse and in written modes. (Labov 1990; Tannen 1994; Rayson et al. 1997; Coates 1998; Biber et al. 1998; Biber & Burges 2000; Härnqvist et al. 2003; Pennebaker et al. 2003; Newman et al. 2008; Yuasa 2010; Keune 2013; Warriner et al. 2013; Podesva & Kajino 2014; Verheijen & Spooren 2017; Hilte et al. 2020, 2022). Recurrent in the studies on the lexico-grammatical level is that men are more likely to display higher complexity, as indicated by average word frequency, type/token ratios, morphological complexity, and syntactic structure. This has been interpreted as men engaging more in ‘report (informative) style’ and women more in ‘rapport (involved) style’ (Tannen 1994; Biber and Burges 2000; Brownlow et al. 2003).

What motivates these gender differences? In addition to explanations in terms of gender as a social construct, where men and women conform to implicit or explicit norms and expectations, evolutionary-based accounts have recently been put forth as well in linguistics (Miller 2002; Foolen 2005; Rosenberg and Tunney 2008; Piersoul & Van de Velde 2021). In the latter line of research, language is seen as a costly trait, and verbal display can be used as a reliable fitness cue (in the sense of Zahavi 1997). While women outperform men in linguistic abilities, on average, men are more likely to use language display and women seem to use this cue in mate selection (Miller 2002: Ch. 10; Dunbar et al. 1997; Rosenberg and Tunney 2008; Lange 2011, Lange et al. 2014).

An historical angle on the gender differences in complexity can shed light on how fluid this evolutionary-evolved difference is. If the gender differences evolve in lock-step with the societal developments, for instance showing a convergence

between the genders when societal gender roles become less segregated and more malleable, then cultural factors weigh in more heavily.

This study looks at 120 years of written discourse by prolific writers in a 200 million word corpus of Dutch journalistic prose (CCLAMP, Piersoul et al. 2021), assessing aggregate measures of lexical, morphological and grammatical complexity, using the Tscan software (Pander Maat et al. 2014): lexical diversity (adjusted type/token ratio), average word frequency, morphological complexity, the number of abstract and general nouns and verbs, the hierarchical depth of embedding of composite clauses. This aggregate perspective, where we measure different complexity metrics, allows us to extrapolate beyond the findings of earlier diachronic studies into particular constructions (e.g. Palander-Collin 1999).

For each linguistic complexity metric, a linear mixed model was built, with an interaction effect of the year of publication and the gender as the explanatory fixed effect, and a random effect for the individual author and for the journal, to account for personal and editorial style. These models detect (i) a consistent effect size and sign on most of the syntactic and lexical measures (i.e. all measures except for word frequency and clause length), with men displaying more complex language, except for lexical diversity, and (ii) a diminishing gender gap, in the course of the 19th and 20th century, on all the metrics except for clause length. For all the lexical measures, women converge to men, in line with findings by Degaetano-Ortlieb et al. (2021). For the syntactic measures, either both genders approach each other, or men converge to women. The net result is that on all levels that display a statistically robust difference between the two genders in the 19th century, the difference has evaporated in the late 20th century.

These results can be interpreted as showing that the evolutionary account of Miller (2002) is not fully stable across time and culture, though the direction and the effect size of the difference in the 19th century is in conformity with such accounts. We have no ready explanation for the various convergence patterns in the lexicon and in the syntactic domain.

References

- Biber D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: CUP.
- Biber, D. & Burges, J. (2000). Historical change in the language use of women and men: gender differences in dramatic dialogue. *Journal of English Linguistics*, 28(1), 21-37.
- Brownlow, S., Rosamond, J. & Parker, J. (2003). Gender-linked linguistic behavior in television interviews. *Sex Roles*, 49, 121-132.

- Coates, J. 1998. *Language and gender: a reader*. Oxford: Blackwell.
- Degaetano-Ortlieb, S., Säily, T. & Bizzoni, Y. (2021). Registerial adaptation vs. innovation across situational contexts: 18th century women in transition. *Frontiers in Artificial Intelligence*, 4, 609970.
- Foolen, A. (2005). Language origins and sexual selection. In: H. Jacobs & L. Wetzels (Eds.), *Liber Amicorum Bernard Bichakjian* (pp. 37-58). Maastricht: Shaker.
- Härnqvist, K., Christianson, U., Ridings, D. & Tingsell, J.-G. (2003). Vocabulary in interviews as related to respondent characteristics. *Computers and the Humanities*, 37, 179-204.
- Hilte, L., Daelemans, W. & Vandekerckhove, R. (2020). Lexical patterns in adolescents' online writing: the impact of age, gender, and education. *Written Communication*, 37(3), 365-400.
- Hilte, L., Vandekerckhove, R. & Daelemans, W. (2022). Linguistic accommodation in teenagers' social media writing: convergence patterns in mixed-gender conversations. *Journal of Quantitative Linguistics*, 29(2), 241-268.
- Keune, K. (2013). *Explaining register and sociolinguistic variation in the lexicon: corpus studies on Dutch*. Utrecht: LOT Dissertation.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205-254.
- Lange, B.P. (2011). Male proneness to verbal display production. *Acta Linguistica*, 5, 97-104.
- Lange, B.P., Zaretsky, E., Schwarz, S. & Euler, H.A. (2014). Words won't fail: experimental evidence on the role of verbal proficiency in mate choice. *Journal of Language and Social Psychology*, 33(5), 482-499.
- Miller, G. (2002). *The mating mind: How sexual choice shaped the evolution of human nature*. New York: Doubleday.
- Newman, M., Groom, C., Handelman, L. & Pennebaker, J. (2008). Gender differences in language use: an analysis of 14,000 text samples. *Discourse Processes*, 45, 211-236.
- Palander-Collin, M. 1999. Male and female styles in 17th century correspondence: I THINK. *Language Variation and Change*, 11, 123-141.
- Pander Maat, H., Kraft, R., Van den Bosch, A., Van Gompel, M., Kleijn, S., Sanders, T. & Van der Sloot, K. (2014). T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal*, 4, 53-74.
- Pennebaker, J., Mehl, M. & Niederhoffer, K. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, 54, 547-577.
- Piersoul, J., De Troij, R. & Van de Velde, F. (2021). 150 Years of written Dutch: the construction of the Dutch corpus of contemporary and late modern periodicals. *Nederlandse Taalkunde*, 26(3), 339-362.
- Podesva, R. & Kajino, S. 2014. Sociophonetics, gender, and sexuality. In: S. Ehrlich, M. Meyerhoff and J. Holmes (Eds.), *The handbook of language, gender and sexuality*. 2nd edn. (pp. 103-122). Chichester: Wiley Blackwell.

- Rayson, P., Leech, G. & Hodges, M. (1997). Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics*, 2(1), 133-152.
- Rosenberg, J. & Tunney, R.J. (2008). Human vocabulary use as display. *Evolutionary Psychology* 6(3): 538-549.
- Tannen, D. 1994. *Gender and discourse*. Oxford: Oxford University Press.
- Verheijen, L. & Spooren, W. (2017). The impact of WhatsApp on Dutch youths' school writing. In: E. Stemle and C. Wigham (Eds.), *Proceedings of the 5th conference on CMC and social media corpora for the humanities (cmccorpora17)* (pp. 6-10).
- Warriner, A.B., Kuperman, V. & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207.
- Yuasa, I.P. (2010). Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile America in women? *American Speech*, 85, 315-337.
- Zahavi, A. (1997). *The handicap principle: a missing piece of Darwin's puzzle*. Oxford: Oxford University Press.