

Premodifying female and male adjectives in journalistic texts. A gender corpus analysis in Czech

Adrian Jan Zasina (The Czech National Corpus, Charles University, Czech Republic)

Currently, there are used a range of methods in the discourse analysis of gender (Baker 2014: 4) but most of these studies have been conducted on English. Our analysis engaging Czech data is one of the first in the field to analyse gender in Czech. In this research, we would like to describe the corpus approach to the analysis of gender in Czech and present similarities and differences between men and women as they appear in journalistic texts.

In the Czech environment, there have been previous qualitative gender studies (Hofmannová 2004) but a corpus approach to analysing gender remains unpopular. Therefore, our research was inspired by some previous discursive gender analysis of collocations conducted by Pearce (2008), Caldas-Coulthard & Moon (2010), Baker (2010), Macalister (2011), Taylor (2013), Moon (2014).

Research material and tools

For purpose of this study, we used a representative selection from the contemporary written Czech SYN2015 corpus (Křen et al. 2015; Křen et al. 2016; Cvrček, Čermáková & Křen 2016) which is divided into three text types: fiction (FIC), non-fiction, and newspapers and magazines (NMG). In our investigation, we took into account NMG, using a research sample which consists of 39,744,419 tokens.

For data analysis, we used two different interfaces: KonText (Machálek & Křen 2013) and a paradigmatic query interface (Cvrček 2017). KonText is an open-source application developed within the framework of the Institute of Czech National Corpus (ICNC, <http://ucnk.ff.cuni.cz>). The interface for paradigmatic query is not formally available as it is still in development at ICNC. This kind of search tool made it possible to classify adjectives into groups that collocate exclusively, almost exclusively or predominantly with the lexemes *muž* and *žena*.

Analysis

In our analysis, we dealt with the lexemes *žena/muž* in the journalistic part (NMG) of the SYN2015 corpus. We were looking for premodifying adjective collocates. In the next step, the adjectives we found were divided into groups according to exclusive, almost exclusive, or predominant appearance with one or another lexeme. Based upon this, we established four groups each sorted out into masculine (collocating with lexeme *muž*) and feminine (collocating with lexeme *žena*) as well and one group of adjectives common for both sexes.

Based on these four groups, we created ten semantic categories into which we sorted out the adjectives. These ten categories were labelled as follows: age, strength and supernatural power, appearance and attractiveness, character, psychological state and adjectives evoking positive/negative emotion, maternity, nationality/ethnicity, action, marital status, sexual orientation and others.

Our analysis has shown that a stereotypical view of men and women is still present in journalistic texts. We proved that certain adjectives are used only with one of the examined lexemes (e.g. *klíčový* 'key' only with *muž* and *practical* 'practical' only with *žena*). Our analysis shows that in the case of masculine adjectives, men are often described by age, strength and supernatural power. In the case of feminine adjectives, they are often related to motherhood and nationality/ethnicity. Adjectives evoking negative and positive emotions, appearance and attractiveness are common for both.

References

- Baker, P. (2014). *Using corpora to analyze gender*. London, England: Bloomsbury.
- Baker, P. (2010). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender & Language*, 4 (1) , 125-149.
- Caldas-Coulthard, C. R., & Moon, R. (2010). 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society*, 21 (2), 99-133.
- Cvrček, V. (2017). Paradigmatické korpusové dotazy a moderní diachronie. In M. Stluka & M. Škrabal (Eds.), *Lířka a czban – Sborník příspěvků k 70. narozeninám prof. Karla Kučery* (pp. 117-130). Praha, Czech Republic: Nakladatelství Lidové noviny.
- Cvrček, V., Čermáková, A. & Křen, M. 2016. Nová koncepce synchronních korpusů psané češtiny. *Slovo a slovesnost*, 77 (2), 83-101.
- Hoffmannová, J. (2004). Ženy a muži v časopisech pro ženy: role, perspektivy, výrazové stereotypy. *Stylistyka*, XIII, 27-34.
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondřička, P. & Zasina, A. (2015). *SYN2015: reprezentativní korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z WWW: <http://www.korpus.cz>
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondřička, P. & Zasina, A. (2016). SYN2015: Representative Corpus of Contemporary Written Czech. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 2522-2528). Portorož, Slovenia: ELRA.
- Macalister, J. (2011). Flower-girl and bugler-boy no more: changing gender representation in writing for children. *Corpora*, 6 (1), 25-44.
- Machálek, T. & Křen, M. 2013. Query interface for diverse corpus types. In K. Gajdošová & A. Žáková (Eds.), *Natural Language Processing, Corpus Linguistics, E-learning* (pp. 166-173). Lüdenscheid, Germany: RAM Verlag.
- Moon, R. (2014). From gorgeous to grumpy: adjectives, age and gender. *Gender & Language*, 8 (1), 99-133.
- Pearce, M. (2008). Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine 1. *Corpora*, 3 (1), 1-29.
- Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora*, 8 (1), 81-113.