Variation and stability in conceptual representations: Insights from the socio-semantics of Czech

Mikuláš Preininger[1], James Brand[1], and Adam Kříž[1]

Faculty of Arts, Charles University, Czech Republic
mikulas.preininger@ff.cuni.cz

1 Introduction

Quantifying the meaning of concepts has been a long-standing goal of the cognitive sciences, with more and more studies emerging in the literature that provide normative ratings across a range of semantic dimensions. However, there have been relatively few investigations into the way socially meaningful dimensions of meaning are represented, for instance how we associate gender with different concepts [1, 2]. Such *socio-semantic* information has been demonstrated to be important for language processing, production and perception [3], and thus could be important for theories of learning, memory and cognition. We present here the first large-scale study that quantifies the socio-semantics of Czech, by providing normative ratings across 5 dimensions of meaning (AGE, GENDER, LOCATION, POLITICAL, VALENCE for a large set of words.

Furthermore, we use this data to explore how these socio-semantic representations might vary across groups of people, specifically by comparing the ratings from female and male participants. The aim of this analysis is to identify whether certain concepts exhibit variation or stability across the two groups, in order to gain a better understanding of which types of concepts are likely to have a shared socio-semantic representation, and which might not.

2 Methods

Our data are taken from the SocioLex dataset [4], an ongoing project where participants were asked to rate words in relation to how they associate the word's meaning to 5 different dimensions on a 7-point Likert scale – gender (feminine/masculine), location (rural/urban), political alignment (liberal/conservative), valence (negative/positive) and age (0-6, 7-17, 18-30, 31-50, 51-65, 66-80, and 81+ years). All dimensions (other than age) had a neutral option at the midpoint. Ratings for 2,700 words were collected (1,603 nouns, 766 adjectives, 331 verbs) for all dimensions.

1,156 participants were recruited from a university wide student database at Charles University, in addition to recruitment via Prolific, with 848 identifying as female ($M_{AGE} = 21.6$ years, $SD_{AGE} = 2.0$) and 308 as male ($M_{AGE} = 22.3$ years, $SD_{AGE} = 2.8$).
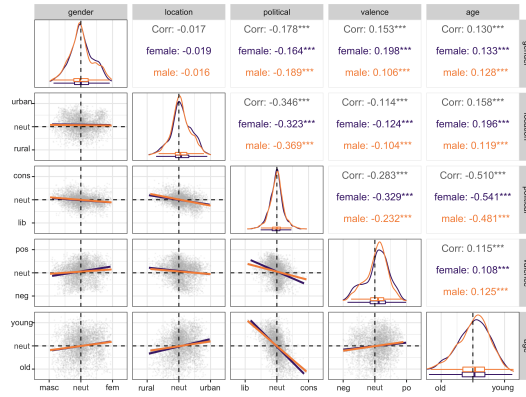
Fig. 1. Visualisation of the linear regression analyses.

We calculated separate mean ratings for females and males for each of the concepts, ranging from -3 (very masculine/rural/liberal/negative) to 3 (very feminine/urban/conservative/positive). For the age dimension, however, the ratings were treated as categorical values, so to compute an aggregated value comparable to the mean, we calculated the proportion of participants who selected each of the age categories, resulting in a 7-dimensional dataset. We then applied Principal Components Analysis to that data, which provided us with a principal component that can be interpreted as a numerical estimate for old-young. See `https://osf.io/e47u8` for all data and analyses.

3   Analysis

We ran linear regression models predicting the relationship between the mean ratings of each of the words, with a fixed effect of participant gender (female/male) interacting with the ratings from the other predictor dimensions. See Figure 1. Results suggest that the relationships between dimensions are fairly similar for both gender groups, however, not always the same. For instance, concepts associated with liberal/conservative stances tend to receive more positive/negative ratings from females than from males.

Next, we assessed the magnitude of differences between the females and males ratings for each of the concepts. This was achieved by calculating the difference in ratings for each of the dimensions, providing us with an estimate of which concepts had a stable representation (no difference) and which had varied representations (large difference), see Table 1. For example, in the dimension of valence, the concept *minisukně* [mini-skirt] was rated as more positive by males, whereas *feminismus* [feminism] was rated as more positive by females.

4   Discussion

The comparisons have shown conceptual representations exhibit variation and stability across gender groups for different dimensions of socio-semantic information. However, for a deeper understanding of these dynamics, we will aim to explore whether certain super-ordinate semantic categories may be predictive of this variation/stability. This will be achieved by obtaining semantic category tags for each of the words from new participants, e.g. golf → sport, and exploring if these categories cluster together words with large or small differences between the gender groups. Furthermore, we are also interested in exploring whether other socio-demographic groups show patterns of variation and stability in their conceptual representations, for instance by investigating socio-semantic ratings from participants who vary in age.

dimension	negative		stable		positive	
gender	*znásilnit* [to rape]	(-1.36)	*demokracie* [democracy]	(0)	*homosexuál* [homosexual.MASC]	(1.26)
location	*svatý* [holy.MASC]	(-1.11)	*opar* [herpes]	(0)	*tancovat* [to dance]	(1.52)
political	*miliardářka* [billionaire.FEM]	(-1.77)	*plakat* [to cry]	(0)	*golf* [golf]	(1.51)
valence	*minisukně* [mini-skirt]	(-1.61)	*covid* [covid]	(0)	*feminismus* [feminism]	(1.72)
age	*samotářská* [solitary.FEM]	(-4.1)	*zdraví* [health]	(0)	*stříbrná* [silver.FEM]	(2.54)

Table 1. Examples of concepts with the most varied/similar ratings. Czech words in italics are the concepts presented to participants. Square brackets contain their English translations and when relevant also gender tagging. Numbers in brackets are differences in mean ratings for each group.

References

1. Lewis, M., Cooper Borkenhagen, M., Converse, E., Lupyan, G., Seidenberg, M. S.: What might books be teaching young children about gender?. Psychological science, 09567976211024643 (2020)
2. Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., Sereno, S. C.: The Glasgow Norms: Ratings of 5,500 words on nine scales. Behavior research methods, **51**(3), 1258–1270 (2019)
3. Hay, J.: Sociophonetics: The role of words, the role of context, and the role of words in context. Topics in cognitive science, **10**(4), 696–706 (2018)
4. Preininger, M., Brand, J., Kříž, A.: Quantifying the Socio-semantic Representations of Words. In: Proceedings of CogSci (2022)