

Occupational gender bias in ungendered languages: Comparing experimental data from Hungarian and Chinese

Anonymous ACL submission

Abstract

This paper is about occupational gender bias and stereotypes, presented in a cross-cultural setting. In the study, we analyze experimental data collected from Hungarian and Chinese speakers on their ratings of occupational titles, answering a question on how typically a job is done by either men or women. Results show that in both of these languages the words carry societal biases, despite that the job titles themselves have no gender markings. We compare the ratings across linguistic and participant gender lines, highlight the differences, and discuss the results with insights ranging from peculiarities in word formation to more generic societal differences. We also compared the human raters' responses with that of a few popular generative AI engines, which will show that the biases we humans carry are even stronger in the Large Language Models (LLMs) underlying these chatbots.

1 Introduction

I like [Kaukonen et al. \(2025\)](#).

1.1 Background

2 Experiment setup

For both languages we devised a simple experiment in survey form, where we asked participants to rate job titles on a 7-point Likert scale. In both cases they were instructed to make decisions on how likely that occupation is to be done by men or by women, according to their own perception. The exact wording in translation were: HU: "Is the occupation typically a man's occupation or a woman's occupation?"; ZH: "What do you think is the ratio of men to women in `occupation_name`?"¹ The scale presented in both questionnaires followed the same logic, from -3 to +3, moving from men to

¹While this study focuses on people who identify or are identified as either male or female, we acknowledge the presence of non-binary people in the workforce.

women, with 0 in the middle, hence the choices were completely male (-3); mostly male (-2); somewhat male (-1); neutral/equal (0); somewhat female (+1); mostly female (+2); completely female (+3). First, we will introduce the Hungarian survey, then the Chinese one, and finally we will compare the results of the two.

2.1 Hungarian survey

2.1.1 Materials

The Hungarian survey contained 50 items, each a commonly occurring job title in Hungary, such as: *modell* 'model' or *katona* 'soldier', in no particular order. Six of the words were attention-check items, which were removed from the final analysis. The attention checks were *pincérnő* 'waitress', *titkárnő* 'secretary (female)', *tanárnő* 'teacher (female)', *takarítónő* 'cleaning lady', *ápolónő* 'nurse (female)', and *házvezetőnő* 'housekeeper (female)'. These words explicitly determine the gender of the worker by appending *-nő* 'woman' to the base word. If participants paid attention, all these items should be rated according to 'completely female' (3). Participants who rated any of these lower than 2, or rated it lower than 3 more than once were rejected.

These words above have also have their counterparts without the suffix, *pincér* 'waiter', *titkár* 'secretary', *tanár* 'teacher', etc., these are unmarked for gender. Common pairs include *énekes* 'singer' – *énekesnő* 'female singer', *színész* 'actor' – *színésznő* 'actress', and in such cases where both are well established, the unmarked word seems to have some male bias, but it does not explicitly refer to a man. More interestingly, there are occupations where the unmarked form is the only one generally used for both genders, and appending *'-nő* 'woman' to it – although possible – would render it a bit awkward, such as in *alkalmazott* 'female employee' or *programozó* 'female programmer', but not impossible such as English *singress* would

be. Furthermore, there are a few cases, where the female-marked version is so widespread, that it is the unmarked version that will sound odd, such as *házvezető* ‘housekeeper’, or to some extent *takarító* ‘cleaner’.

In short, we are interested in these unmarked words, as they do not inherently carry a male bias, but according to our expectations will nonetheless be rated according to the prevailing societal stereotypes.

We also included *diák* ‘student’ out of curiosity. Although being a student is not a job, but it is beyond doubt the only truly gender-neutral “occupation” there is, since it is mandatory for every child to go to school both in Hungary and in China. We wanted to see if there would be any bias regarding this word, especially that Hungarian has a female form for it, *diáklány* ‘girl student’.

2.1.2 Procedure

The questionnaire was distributed online, and after a brief welcome message and the instructions, the words were presented in a simple list format, each word with a corresponding rating scale next to it, with no context. Time limit was not set, but the survey was designed to take around 5 minutes; participants took 4 minutes 25 seconds on average to finish.

2.1.3 Participants

A total of 22 native Hungarian speakers filled the questionnaire, and after validating the responses (reviewing attention checks and manual checking for anomalies) 2 were rejected. The participants were mostly recruited through Prolific, an online platform, with screeners for current location (Hungary) and first language (Hungarian). Participants were compensated for their time with a small reward. In the end, the Hungarian rating dataset had 20 participants (11 female, 9 male), with ages ranges of 25-35 (n=11), 35-45 (n=4), and 45-55 (n=5). See Figure 1 for the distribution.

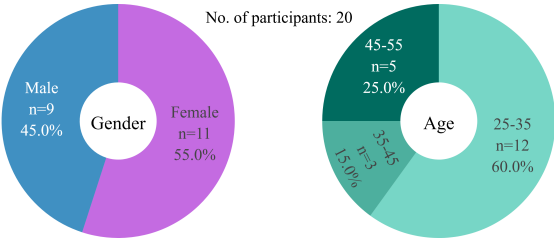


Figure 1: Demographics of the Hungarian participants.

3 Mandarin Chinese survey

3.0.1 Materials

The Chinese survey also contained 50 items with commonly occurring job titles in Mandarin Chinese (Simplified), also in random order. There were six attention checks included to ensure participant engagement and data quality, these were 妈妈 ‘mother’, 爸爸 ‘father’, 女作家 ‘female writer’, 男作家 ‘male writer’, 女画家 ‘female painter’, 男画家 ‘male painter’. Similarly to the Hungarian attention checks, these words are inherently feminine or masculine in meaning, or explicitly determine gender by prepending 女 ‘woman’ and 男 ‘man’, helping to filter out inattentive responses.

3.1 Procedure

Ask Wenhui about the procedure for the Chinese survey.

3.2 Participants

The Chinese survey was completed by 24 native Mandarin Chinese speakers, and none of them were rejected after validation. Participants were paid a small fee for completing the questionnaire. The 24 participants (10 female, 14 male) were mostly university students, aged <25 (n=15), 25-35 (n=8), or 35-45 (n=1). See Figure 2 for the distribution.

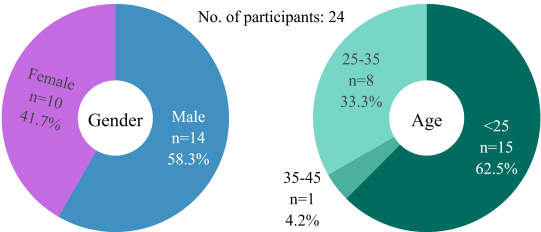


Figure 2: Demographics of the Chinese participants.

4 Results & Analysis

...overall...

4.1 Hungarian

The Hungarian data was first analyzed using a one-sample t-test to determine which of the occupations showed a significant bias measured from 0 (neutral/equal). The results showed that the majority of occupational titles (36 out of 44) were rated with a significant gender bias. See Figure 3 for a visualization of the mean ratings.

Occupations were rated according to expectations, following societal stereotypes. Words

with the highest female bias were *nővér* ‘nurse’ (2.20), *kozmetikus* ‘beautician’ (2.20), *házvezető* ‘housekeeper’ (1.80), *légiutas-kísérő* ‘flight attendant’ (1.40), and *takarító* ‘cleaner’ (1.25), while words with the highest male bias included *munkás* ‘worker’ (-1.40), *pilóta* ‘pilot’ (-1.65), *katona* ‘soldier’ (-1.80), *biztonsági őr* ‘security guard’ (-1.90), and *tűzoltó* ‘firefighter’ (-2.20).

The 8 job titles that came back as not significantly biased were: *ápoló* ‘nurse’ (0.45), *PR munkatárs* ‘PR worker’ (0.30), *felhasználó* ‘server’ 0.20, *jegyárus* ‘ticket seller’ (0.20), *diák* ‘student’ (0), *tanár* ‘teacher’ (0), *bíró* ‘judge’ (-0.25), and *titkár* ‘secretary’ (-0.30). It is worth noting that while *diák* ‘student’ was rated 0 by absolutely everyone, *tanár* ‘teacher’ had more variety in the ratings, with a high standard deviation.

The strongest agreement were on *diák* ‘student’ (0), *bíró* ‘judge’ (-0.25), *biztonsági őr* ‘security guard’ (-1.90), *tudós* ‘scientist’ (-0.40), and *orvos* ‘doctor’ (-0.55).

LLMs?

We also ran a two-sample *t*-test to compare the ratings of male and female participants for each occupation, and see if there was any significant difference between the two groups. The only job that showed a significant difference was *rendőr* ‘police’, where the male bias was much higher by male raters (-1.77 vs. -1.00). The results are summarized in Figure 4.

4.2 Chinese

Similarly to Hungarian, we found that a majority of occupations in Chinese were also rated with significant gender bias. The results of the one-sample *t*-test showed that 39 out of 44 occupations were biased. The mean ratings are shown in Figure 5.

...

The 4 job titles that were not significantly biased were: 营养师 ‘dietetician’ (0.21), 服务员 ‘waiter/server’ (0.08), 学生 ‘student’ (-0.04), and 医生 ‘doctor’ (-0.21).

In Chinese too, there was strong agreement on 学生 ‘student’ (-0.04) and 医生 ‘doctor’ (-0.21), the remaining words in the top 5 jobs with the lowest standard deviation were 服务员 ‘waiter’ (0.08), 售票员 ‘ticket seller’ (0.33), and 护士 ‘nurse’ (1.58).

4.3 Cross-linguistic comparison

The first noticeable trend is that in general, the two languages have similar biases for the same occupa-

tions.

Shared items on the extreme ends of the scale included *nurse* and *beautician*, and *firefighter* and *security guard*.

We compared the ratings of the 42 common items in the two datasets, again, using a two-sample *t*-test to see if there were significant differences between the two languages. The results are summarized in Figure 6.

Explain...

5 Experiment design

References

Elisabeth Kaukonen, Polina Oskolskaia, Liina Lindström, and Raili Marling. 2025. [Gender, language and labour: Gender perception of Estonian and Russian occupational titles](#). *Frontiers in Communication*, 9.

A Example Appendix

This is an appendix.

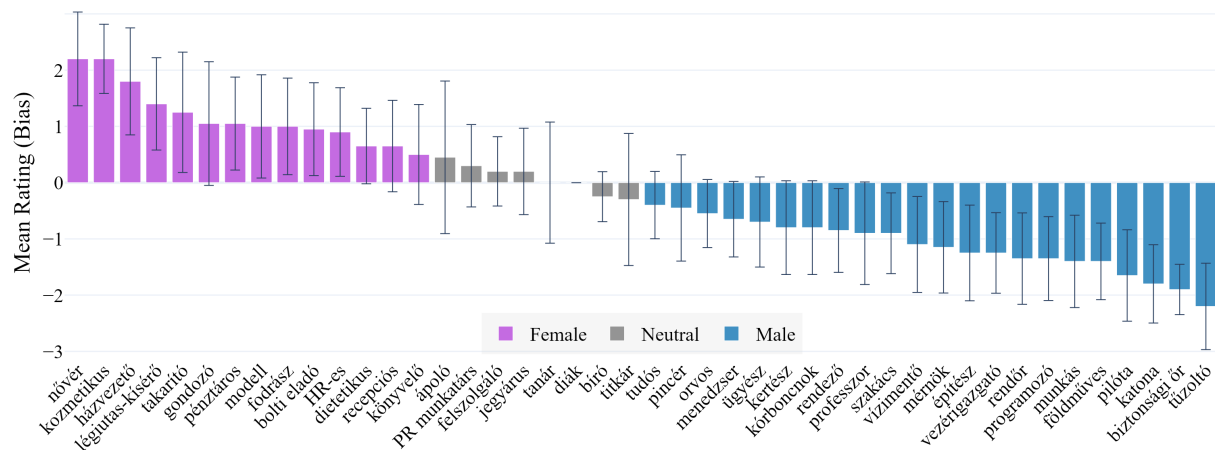


Figure 3: Mean ratings of occupational titles in Hungarian with standard deviations, significant gender bias highlighted – [explore the interactive plot](#).

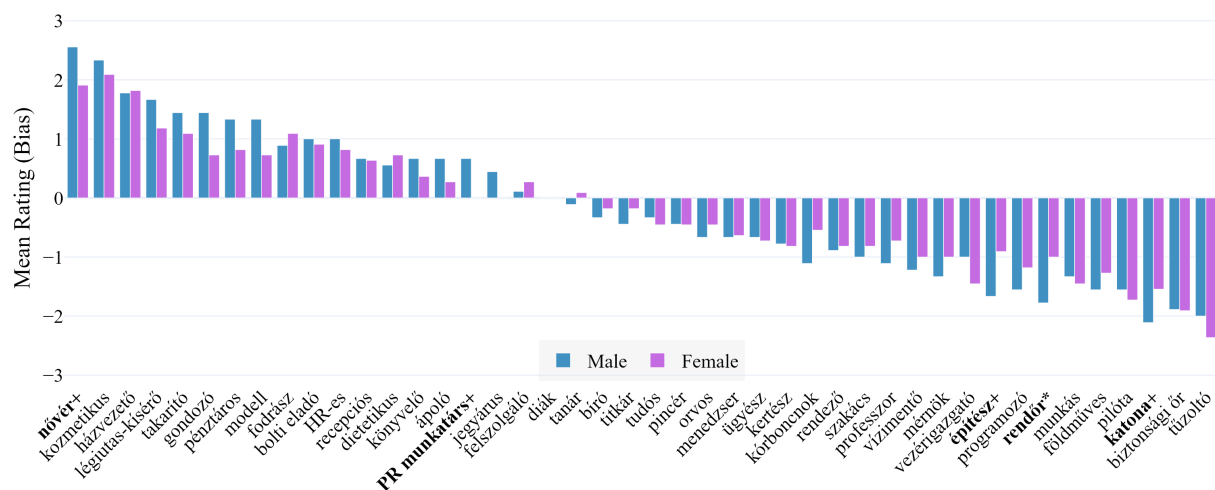


Figure 4: Mean ratings of occupational titles in Hungarian by gender, significant differences highlighted (significant*, and marginally significant+ in bold) – [explore the interactive plot](#).

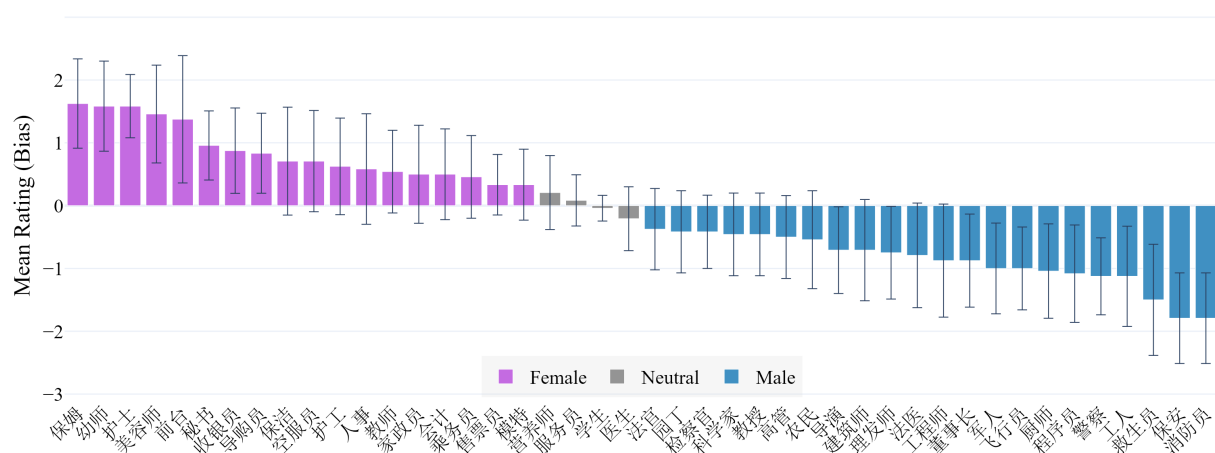


Figure 5: Mean ratings of occupational titles in Chinese with standard deviations, significant differences highlighted (significant*, and marginally significant+ in bold) – [explore the interactive plot](#).

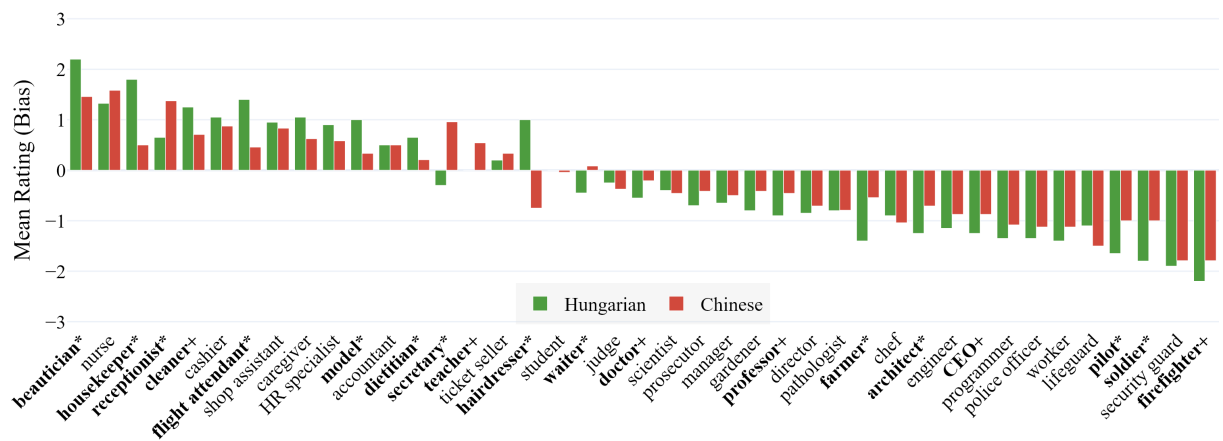


Figure 6: Mean ratings of common occupational titles in Hungarian and Chinese, significant differences highlighted (significant*, and marginally significant+ in **bold**) – [explore the interactive plot](#).