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

Anonymous ACL submission

Abstract

This paper is about occupational gender bias and stereotypes, presented in a cross-linguistic setting. In the study, we analyze experimental data collected from Hungarian and Mandarin Chinese speakers on their ratings of job titles, answering a question on how typically a job is done by either men or women according to their perceptions. Results show that in both of these languages the words carry societal biases, despite the fact that the job titles themselves have no gender markings. We analyze and compare the ratings across linguistic and gender lines, highlight the differences, and discuss the results with insights ranging from peculiarities in word formation to more cross-cultural, societal differences. Additionally, 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 the survey format, 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 would be: "Is the occupation typically a man's occupation or a woman's occupation?" (HU); "What do you think is the ratio of men to women in `occupation_name`? (ZH)".[1] The scale presented in both questionnaires followed the same logic, from -3 to +3, moving from men to 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 them lower than 3 more then once were rejected.

These words above also have their counterparts without the *nő* 'woman' element, i.e. *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 *al-*

---

[1] While 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.

kalmazottnő 'female employee' or *programozónő* '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 a bit 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 possess a male bias – and certainly not grammatically – but according to our expectations they will be rated according to the prevailing societal stereotypes nontheless.

— Mention some "default is masculine" theory if exists...

The full list of words is as follows: *modell*, *katona*, *kórboncnok*, *vezérigazgató*, *menedzser*, *nővér*, *szakács*, *pincérnő*, *felszolgáló*, *könyvelő*, *professzor*, *építész*, *tudós*, *ápoló*, *pénztáros*, *bíró*, *munkás*, *vízimentő*, *titkárnő*, *jegyárus*, *tűzoltó*, *mérnök*, *tanárnő*, *rendező*, *takarító*, *HR-es*, *házvezető*, *légiutas-kísérő*, *pincér*, *takarítónő*, *orvos*, *fodrász*, *földműves*, *ápolónő*, *gondozó*, *bolti eladó*, *kertész*, *titkár*, *PR munkatárs*, *dietetikus*, *tanár*, *rendőr*, *pilóta*, *házvezetőnő*, *recepciós*, *biztonsági őr*, *ügyész*, *kozmetikus*, *programozó*.

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 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]

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)

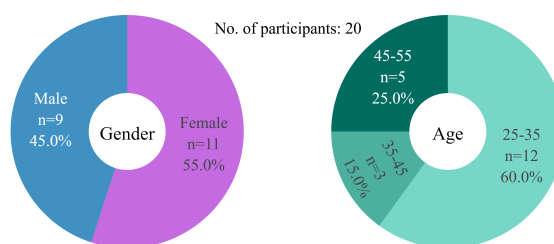[2] <mark>Please see a sample survey</mark> here.

Figure 1: Demographics of the Hungarian participants.

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.

3 Mandarin Chinese survey

3.1 Materials

The Chinese survey also contained 50 items with commonly occurring job titles in Mandarin Chinese (Simplified), also in randomized 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. Participants who failed to rate these with the highest scores of either 3 or -3 were rejected.

3.2 Procedure

The Chinese surveys were also distributed online with almost identical instructions. <mark>Ask Wenhui about the procedure for the Chinese survey.</mark>
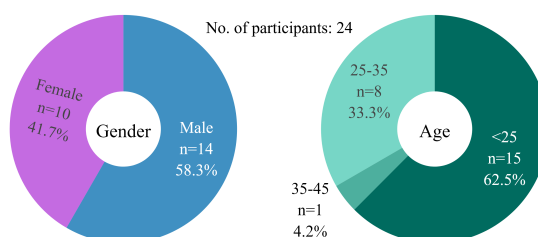
3.3 Participants

Figure 2: Demographics of the Chinese participants.

The Chinese survey was completed by 30 native Mandarin Chinese speakers, 6 of which were re-

jected after failing attention checks. Participants were paid a small fee for completing the questionnaire. The 24 accepted 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.

4   Results & Analysis

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 against 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, with the gender biases highlighted.

In general, occupations were rated according to expectations, following societal stereotypes and realities. Words with the highest female bias were *kozmetikus* 'beautician' (2.20), *házvezető* 'housekeeper' (1.80), *légiutas-kísérő* 'flight attendant' (1.40), and *takarító* 'cleaner' (1.25), *gondozó* 'caregiver' (1.05), 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 highest rated feminine word – *nővér* 'nurse' (2.20) – is a bit special, as it literally means 'sister' and goes back to the time when nuns were the ones taking care of the sick, hence the word carries a strong feminine bias that is encoded in its literal meaning. Interestingly, it was not rated as an exclusively female job, probably because male nurses are also common. The gender-neutral word *ápoló* 'nurse' for the same job was also tested, and it received a neutral rating of 0.45.

The 8 job titles that came back as not significantly biased were: *ápoló* 'nurse' (0.45), *PR munkatárs* 'PR worker' (0.30), *felszolgáló* '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 nearly everyone, *tanár* 'teacher' had more variety in the ratings, with a higher 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).

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 discrepancies 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.

<mark>...Expectations...</mark>

In Chinese, the words with the highest feminine bias were 保姆 'domestic helper' (1.63), 护士 'nurse' (1.58), 幼师 'kindergarten teacher' (1.58), 美容师 'beautician' (1.46), and 前台 'receptionist' (1.38). Words with the highest male bias were 警察 'police officer' (-1.13), 工人 'worker' (-1.13), 救生员 'lifeguard' (-1.50), 保安 'security guard' (-1.79), and 消防员 'firefighter' (-1.79), which shows a relatively strong similarity to the Hungarian trends.

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).

The two-sample *t*-test comparing the ratings of male vs. female participants showed that there were significant differences between what people think of a typical 收银员 'cashier' (m=0.64; f=1.20), 模特 'model' (m=0.57; f=0), 法官 'judge' (m=-0.14; f=0.70), and 农民 'farmer' (m=-0.79; f=-0.20). The results are summarized in Figure 6.

4.3   Cross-linguistic comparison

When comparing the two sets of ratings, the first noticeable trend is that in general, the two languages have similar biases for the same occupations. Shared items on the extreme ends of the scale include *nurse* and *beautician*, and *firefighter* and *security guard*. Whereas job titles that were unbiased in both datasets were *server* (*felszolgáló*, 服务员), and *student* (*diák*, 学生).
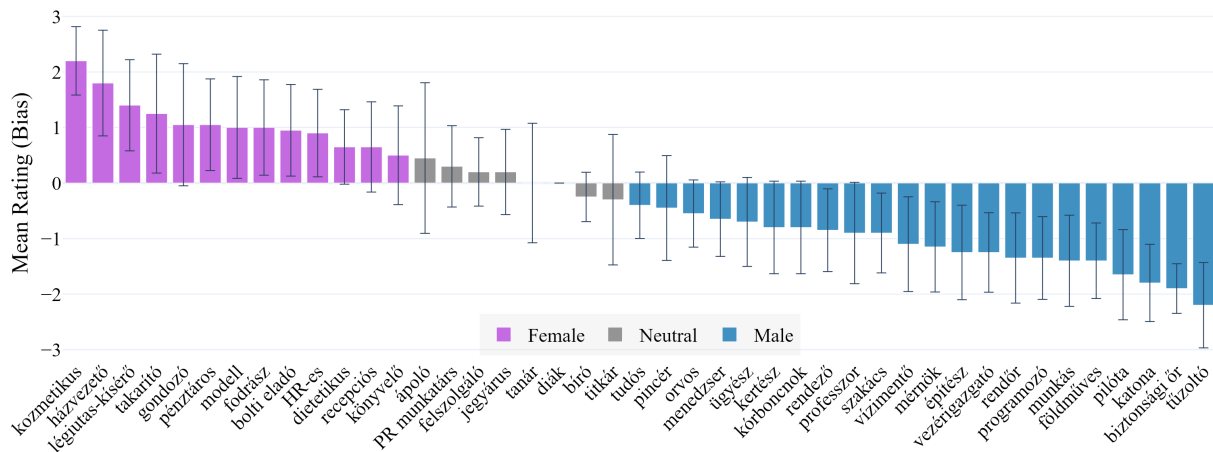
Figure 3: Mean ratings of occupational titles in Hungarian with standard deviations, significant gender bias highlighted – explore the interactive plot.
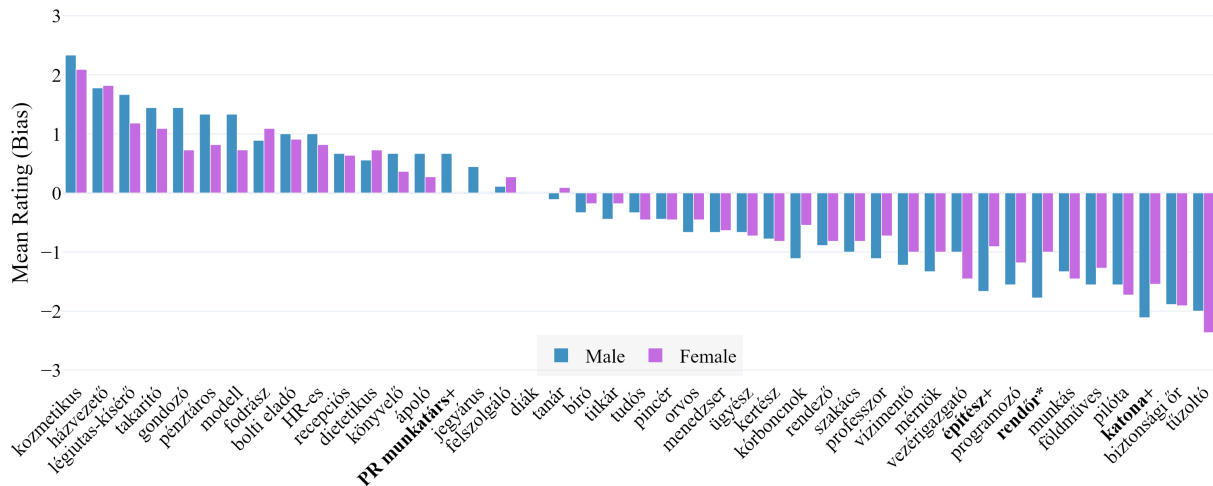

Figure 4: Mean ratings of occupational titles in Hungarian by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – explore the interactive plot.

The wordlists of the two datasets were not exactly the same, but by performing an inner join on the two lists, we could match 42 items together by their meanings. Using a two-sample t-test, we checked if there were significant differences between the two languages. The results are summarized in Figure 7. Explain...

The most striking difference was the word for 'hairdresser' (*fodrász* vs. □□□), which shows a strong female bias in Hungarian (1.00) and a strong male bias in Chinese (-0.75). Discuss this...

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 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 occupational titles in Chinese by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – explore the interactive plot.

Figure 7: Mean ratings of common occupational titles in Hungarian and Chinese, significant differences highlighted (significant*, and marginally significant+ in **bold**) – explore the interactive plot.