Occupational gender bias in ungendered languages and LLMs: Comparing Hungarian and Chinese

Anonymous ACL submission

Abstract

This paper examines occupational gender bias and stereotypes in a cross-linguistic setting. We analyze ratings of 50 job titles collected from speakers of two languages without grammatical gender: Hungarian and Mandarin Chinese. Participants were instructed to rate how typical it is for a certain job to be associated with men or women, according to their own perceptions. Our results show that in both languages the occupational nouns carry societal biases, despite the fact that the job titles themselves have no grammatical gender markings. We analyze the ratings by participant gender and perform intra-linguistic and cross-linguistic comparisons, highlighting differences in the two languages and offering insights that range from peculiarities in word formation to broader cross-cultural generalizations. Additionally, we compared the human raters' responses with that of a few popular generative artificial intelligence (AI) agents. Interestingly, the biases exhibited by the large language models (LLMs) in these chatbots were found to be even stronger than those shown by our human participants.

1 Introduction

"Hairdressers are mostly women." – one could say if during the course of their life they encountered more female than male hairdressers. But would this perception reflect reality?

Occupational gender stereotypes are generalizations about the roles and characteristics of individuals in a certain group of in the workforce, and due to the long history of sex-based division of labour, it is one of the key components of gender stereotypes as a whole (Deaux and Lewis, 1984). Stereotypes are present from a young age (Canessa-Pollard et al., 2022) based on various inputs, such as personal observations, in-group–out-group bias, (social) media influence, societal norms and cultural narratives, and much more. As such, they are an inherent part of the human condition, originally said to be mechanism to make sense of biological sex differences (Levanon and Grusky, 2016). These stereotypes however are often prejudiced, inaccurate, and are "misrepresenting the true ratios of gender in the workplace" (Garnham et al., 2015; Gygax et al., 2016) – as Kaukonen et al. (2025) have pointed out.

One of the factors in both shaping and reflecting social biases is language, and its role has been extensively studied in the past decades, including regarding occupational gender stereotypes (cf. Sabatini, 1985; Pauwels, 1997; Gygax et al., 2008; Misersky et al., 2014; Lewis and Lupyan, 2020; Kaukonen et al., 2025). Much of the previous research have been conducted on languages with a grammatical gender system, such as Italian, French, German, Dutch, or English where occupational nomenclature and human agent nouns are often gendered (e.g., *handyman*, *seamstress*).[1]

In recent years however, documentation of gender bias has been extented to other languages as well, such as Lithuanian, Icelandic, and Polish, also including languages without grammatical gender like Chinese, Japanese, and Thai (Hellinger and Bußmann, 2003; Pauwels, 1998, 2003). Most recently, Kaukonen et al.'s (2025) work compared Estonian and Russian, where the former lacks grammatical gender.

Our study compares Hungarian and Chinese, both of which lack grammatical gender, which in turn does not mean that these languages are free of gender biases. We are interested in how speakers of these two languages navigate the stereotypes of the workplace and how much they impose gender

[1] One important notion here is the asymetry between masculine and feminine forms of occupational nouns, that is, a lexical gap in occupational titles resulting in "male as norm" principle and the absence of words denoting a variety of female occupations (Baron, 1986; Hellinger, 1990; Sabatini, 1985; Yaguello et al., 1978; Pauwels, 2003; Lassonde and O'Brien, 2013).

biases present in society onto jobs and occupational titles without explicit gender markings.

1.1 Background

1.1.1 Hungarian

Hungarian is a Finno-Ugric language in the Uralic language family, without a grammatical gender system,[2] and most occupational nouns and job titles are realized in linguistically gender-neutral terms. On the surface level, the feminine form is created by appending *-nő* ‘woman’ to the unmarked base word, but in reality, these female-marked nouns are formed by compounding *nő* ‘woman’ (a noun) to the base word (noun or adjective), which is a regular way of producing female occupational titles. In practice however, this does not equal to a symmetric male-female pair, as the unmarked form is not necessarily “masculine”, it is also neutral. We can observe 3 types in the pragmatic usage of occupational nouns when it comes to the unmarked–marked pairing and its implications for the gendering of the unmarked nouns at the discourse level:

1) Both forms are common. Frequently occurring word-pairs in Hungarian would be for example *énekes* ‘singer’ – *énekesnő* ‘female singer’ (not in our dataset). In cases where both versions are well established – i.e., both occur with a relatively high frequency in a balanced corpus – the unmarked word seems to carry some male bias, as the frequent use of a feminine form indicates a need and/or custom for differentiation. We wanted to test if raters perceived this bias or not. In this example, the absolute and relative frequencies (occurrance per a million words) of the two lemmatized nouns in the Hungarian National Corpus (HNC) are 1441/9.4001 for *énekes* and 748/4.8795 for *énekesnő* (Váradi, 2002; Oravecz et al., 2014); the frequency difference here is roughly half (51.9%).

The stark deviations in frequencies for marked–unmarked word pairs such as the above are not an indicator for a strong gender bias – we can assume that both men and women singers would be equally represented in the Hungarian corpus – but reflect that in general, the unmarked, neutral forms are used for either males or females when talking about one’s occupation. The female-marked forms are used when there is an explicit intention to specify the gender of the individual, and when it is otherwise not known from context or from proper names.[3] A special situation would be the use of the vocative case, which requires the marked form when addressing female professionals, e.g., *Tanárnő!* ‘teacher (f.)’ or *Doktornő!* ‘doctor (f.)’.

2) Only unmarked form is common. There are many cases where the unmarked form is the only one generally used for both genders. Take for example *ügyész* ‘prosecutor’ (8451/55,1287) vs. *ügyésznő* ‘female prosecutor’ (56/0,3653), or *fodrász* ‘hairdresser’ (944/6.1580) vs. *fodrásznő* ‘female hairdresser’ (35/0.2283); the deviations in frequency here are over multiple orders of magnitude. In these instances, the unmarked form is the default word to describe anyone practicing the occupation regardless of gender, and appending *’-nő* ‘woman’ to it – although possible – would render it unusual and a bit awkward; but still not as uncanny as Modern English *singress* would be.[4]

3) Marked form is common. Furthermore, there are cases, where the female-marked version ending in *-nő* is so ubiquitous, that it is the unmarked version that will sound a bit unusual, such as *házvezető* ‘housekeeper’ (10/0.0652) vs. *házvezetőnő* ‘female housekeeper’ (92/0.6001), or, to a small extent *takarító* ‘cleaner’ (169/1.1024) vs. *takarítónő* ‘female cleaner’ (392/2.5571). In our opinion these instances reflect deeply engrained societal biases.

1.1.2 Chinese

Chinese, a Sino-Tibetan language, only marks gender when writing the 3rd person singular pronoun (他 *tā* ‘he’ / 她 *tā* ‘she’) but that too is a relatively recent invention, going back to the May Fourth Movement of 1919 (Bi, 2013), and similarly to Hungarian, most occupations are unmarked for gender.

Chinese is generally acknowledged as a language that lacks grammatical gender from a structural perspective (Li and Thompson, 1989). In contrast to languages with mandatory noun gender systems such as French, German, or Spanish, Chinese nouns, adjectives, and articles lack distinctions of masculine, feminine, or neuter gender. Thus, certain scholars contend that the Chinese linguistic system is fundamentally gender-neutral and may not inher-

[2]The problematics of linguistic gender in Hungarian has been discussed by Vasvári (2014).

[3]For example, the sentence *Anyukám tanár.* is the canonical way of saying ‘My mom is a teacher.’. It uses the unmarked form, primarily because we want to channel information about the individual’s vocation, and secondly because it is obvious from the subject that she is a woman, so the marking would be redundant.

[4]Although English had a form *singeress* from Middle English, it is now obsolete.

ently reflect gender classification (Li and Thompson, 1989; Packard, 2000).

Nonetheless, sociolinguistic research argues that language use, particularly at the pragmatic level, is significantly shaped by socio-cultural attitudes (Labov, 1972), and many occupations are associated with relatively strong gender stereotypes (Sun, 1997; Su et al., 2021). For instance, 护士 *hùshì* translates to 'nurse' and 保姆 *báomǔ* to 'nanny', both of which are predominantly female occupations, whereas 警察 *jǐngchá* means 'police officers' and 高管 *gāoguǎn* refers to 'executive; manager', which are predominantly male roles. This societal perception results in pragmatic asymmetry: although occupational terms are grammatically gender-neutral, speakers or language users frequently form default gender assumptions in communication contexts based on societal stereotypes. Consequently, when a practitioner's gender diverges from the conventional stereotype associated with their profession, or when particular circumstances require explicit gender identification, speakers often precede occupational terms with gender markers 男 *nán* for 'male' or 女 *nǚ* for 'female'. This results in Chinese occupational phrases such as 男护士 *nánhùshì* for 'male nurse', 女警察 *nǚjǐngchá* for 'female police officer', 女高管 *nǚgāoguǎn* for 'female executive', and 男保姆 *nánbáomǔ* for 'male nanny'. This practice of incorporating gender markers demonstrates a dual effect (Sun, 1997; Gustafsson Sendén et al., 2015; Su et al., 2021). On one hand, it emphasises the transcendence of gender constraints by individuals, thereby increasing the visibility of specific groups, particularly those in non-traditional gender roles. Conversely, scholars suggest that this gender-focused terminology may, particularly in contradictory contexts, reinforce prevailing gender stereotypes (Sun, 1997; Gustafsson Sendén et al., 2015). For instance, under societal cognition specific to China, the aforementioned occupation title 高管 *gāoguǎn* 'executive' subconsciously defaults to a male referent. Highlighting 女高管 *nǚgāoguǎn* 'female executive' designates the individual as an anomaly and may inadvertently maintain the gender stereotype in China that associates the term 'executive' predominantly with males, showing an intrinsic uniqueness or necessity for explicit distinction when women assume this position.

In summary, Chinese occupational titles are inherently gender-neutral, specifically due to the lack of grammatical gender and their primary function of solely describing the work itself. The addition of gender modifiers such as 男 *nán* 'male' or 女 *nǚ* 'female' operates as a pragmatic device rather than a grammatical requirement. This strategy primarily signals socio-cognitive markedness by highlighting deviations from occupation-gender stereotypes (i.e., denoting non-prototypical associations), as empirically demonstrated in Chinese corpora where terms like 女总统 *nǚzóngtǒng*, 'female president' occur more frequently than their male-marked counterparts in leadership roles (Su et al., 2021; Farris, 1988). Concurrently, it fulfills referential specificity in contexts demanding explicit gender identification, such as legal depositions or medical documentation where deixis resolution is pragmatically necessary (Hellinger and Bußmann, 2003; Stahlberg et al., 2011).

Regarding both languages, we were interested in the unmarked occupational nouns, as they do not inherently possess linguistic gender bias, but according to our expectations they will be rated reflecting prevailing societal stereotypes, and thus show varying degrees of gender bias. Our research questions then can be formulated as follows: To what extent do speakers of Hungarian and Chinese exhibit gender bias when rating occupations and job titles? Are there significant in gender bias between male and female raters within the same language? What are the key differences and similarities in occupational gender stereotypes between Hungarian and Chinese speakers?

In recent years, with the advent of Large Language Models (LLMs), the challenge of addressing occupational gender stereotypes has been a prominent issue in language technology as well (cf. Kirk et al., 2021; Ju et al., 2024; An et al., 2025). To see how LLMs handle gender biases compared to human participants, we also conducted a series of experiments using the LLMs of popular artificial intelligence (AI) companies and analyzed their outputs, both in Hungarian and Chinese.

In section 2 we describe the experimental setup and our methods of analysis, followed by the results and their interpretation in section 3. Finally, we discuss the findings in section 3.4. FIX

2 Experimental setup

For both languages we designed a survey-based experiment in which participants were asked to rate job titles on a 7-point Likert scale, ranging from *completely male* (-3) to *completely female* (+3), with an equal/neutral midpoint (0). Participants were instructed to make decisions on how likely is an occupation to be pursued by men or by women.[5]

2.1 Participants

A total of 24 native Hungarian speakers filled our questionnaire, and after validating the responses (reviewing attention checks and manually checking for anomalies) 2 were rejected. Most of the participants were recruited online using the participant recruitment platform Prolific (https://www.prolific.com/), with screeners set for location (Hungary) and first language (Hungarian); raters were compensated for their time with a small monetary reward. In the end, the Hungarian ratings dataset had 22 participants (11 female), with age-ranges of 25-35 (n=14), 35-45 (n=3), and 45-55 (n=5). See Figure 9 for the distribution.

The Chinese survey was completed by 30 native Mandarin Chinese speakers, 6 of which were rejected after failing attention checks. Chinese participants were recruited from career development interest groups on WeChat, and they were required to have been born and raised in mainland China with Mandarin as their primary language and the primary language spoken at home. Additionally, they must have resided in China continuously for the past five years. Participants were paid a small fee for completing the questionnaire. The 24 accepted participants (10 female) were mostly university students, aged <25 (n=15), 25-35 (n=8), or 35-45 (n=1). See Figure 10 for the distribution.

2.2 Materials

The Hungarian survey contained 44 items and 6 attention-check items, each a well-known occupational title in Hungary, such as: *modell* 'model' or *katona* 'soldier'. The attention-check items were removed from the final analysis, these 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 and all should be rated 'completely female' (3). Participants who rated any of these lower than 2, or rated them lower than 3 more then once were rejected.

The 6 words above have counterparts without the *nő* 'woman' element, e.g., *pincér* 'waiter', and all were included in our survey. As explained, these words are unmarked for gender, but they do not explicitly refer to men. Therefore, the expected gender bias in the ratings of these occupations – if any – is due to social factors, not linguistic ones. For the full list, see the appendix.[6]

The Chinese survey contained 44 items of commonly occurring job titles in Mandarin Chinese (Simplified), with 6 attention checks. The attention checks were 妈妈 *māmā* 'mother', 爸爸 *bàba* 'father', 女作家 *nǚzuòjiā* 'female writer', 男作家 *nánzuòjiā* 'male writer', 女画家 *nǚhuàjiā* 'female painter', 男画家 *nánhuàjiā* 'male painter'. These words are inherently feminine or masculine in meaning, or explicitly determine gender by prepending 女 *nǚ* 'woman' and 男 *nán* 'man', helping to filter out inattentive responses. Participants who failed to rate these with the highest scores of either *completely male* or *completely female* were rejected. See the full list of items in the appendix.

2.2.1 Procedure

Hungarian and Chinese participants were both instructed to rate each word between *completely male* and *completely female*, the ratings were then converted to numerical values from -3 to +3. 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). The exact wording of the main question of the Hungarian survey was: "Ön szerint a foglalkozás tipikusan férfi foglalkozás, vagy tipikusan női foglalkozás?" (Is this occupation typically a man's occupation or a woman's occupation?), while in Chinese it was: "对于XX这个职业，您认为通常担任该职业的男女性别比例是多少？" (What do you think is the ratio of men to women in __?). The surveys were created in Microsoft Forms with essentially identical instructions. Hungarian participants were presented with the words in a list format in no particular order without context, each word with a corresponding rating scale next to it (cf. Fig. 8), Chinese participants saw each word highlighted in a question above, and were presented with the

[5]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.

[6]We have included *diák* 'student' out of curiosity. Although being a student is not a job per se, but it is beyond doubt the only truly gender-neutral "occupation" there is, since it is mandatory for everyone 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-marked form for it – *diáklány* 'girl student', adding *-lány* 'girl' to the base noun – and could be considered to belong to the 1st type with a potential male bias when considered in contrast.

scale directly below, in a random order (cf. Fig. 7). Time limit was not set, but the survey was designed to take around 5 minutes, and participants have finished under 5 minutes on average.

In addition to the two surveys, we have prompted 5 different popular AI chatbots: Mistral AI's Le Chat (Free), Microsoft's Copilot (+Think Deeper), OpenAI's ChatGPT (GPT-5 Preview), Google's Gemini (2.5 Flash), and DeepSeek (Deepthink R1) to elicit ratings on the same job titles. In both languages we used the same scale, the same instructions with a pretext: "You are participating in an experiment and your answers will help our research.".[7] Between August 7 and 14, 2025 we have prompted each AI agent 10 different times (2×5×10=100 total prompts), and then aggregated the results.[8]

2.3 Methods

For analyzing the data, we used the `scipy` library in Python, and performed one-sample *t*-tests to determine the significance of gender bias in the mean ratings of occupations, and independent sample (two-sample) *t*-tests to analyze the differences between different groups, such as the ratings of male and female raters and the differences between the ratings of Hungarian and Chinese raters for the same occupations.[9]

When comparing the ratings of humans vs. AI systems, we performed Pearson's correlation to determine the strength of the relationship between the ratings and determine which LLM is more aligned with human judgments.

3 Results & Analysis

For both languages, the majority of occupations were rated with a significant gender bias. In Hungarian, 36 out of 44 occupations showed significant bias, and in Chinese, 39 out of 40 were biased.

3.1 Hungarian

The Hungarian data was first analyzed using a one-sample *t*-test to determine which of the occupations showed significant deviations, 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 (p<0.05), with 14 showing female, and 22 showing male bias. See Figure 1 for a visualization of the mean ratings, with the gender biases highlighted.

3.1.1 Overall distribution of ratings

In general, occupations were rated according to expectations, following societal stereotypes and realities. Words with the highest female bias were *kozmetikus* 'beautician' (2.27), *nővér* 'nurse' (2.23), *házvezető* 'housekeeper' (1.77), *légiutas-kísérő* 'flight attendant' (1.45), and *takarító* 'cleaner' (1.14), while words with the highest male bias included *munkás* 'worker' (-1.55), *pilóta* 'pilot' (-1.68), *katona* 'soldier' (-1.86), *biztonsági őr* 'security guard' (-1.95), and *tűzoltó* 'firefighter' (-2.27).

Nővér 'nurse' (2.23) – 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 female bias that is encoded in its literal meaning. Interestingly, it was not rated as an exclusively female job, probably because male nurses are now also common. The gender-neutral word *ápoló* 'nurse' for the same job was also tested, and it received a neutral rating of 0.41.

The 8 job titles that came back as not significantly biased were: *ápoló* 'nurse' (0.41), *jegyárus* 'ticket seller' (0.23), *PR munkatárs* 'PR specialist' (0.23), *felszolgáló* 'server' (0.18), *diák* 'student' (0), *tanár* 'teacher' (0), *bíró* 'judge' (-0.18), and *titkár* 'secretary' (-0.23). It is worth noting that while *diák* 'student' was rated 0 by everyone, *tanár* 'teacher' had more individual variation in the ratings, leading to a higher standard deviation.

The strongest agreement were on *diák* 'student' (0, std=0), *biztonsági őr* 'security guard' (-1.95; std=0.4857), *bíró* 'judge' (-0.18; std=0.5011), *felszolgáló* 'server' (0.18; std=0.5885), *tudós* 'scientist' (-0.41; std=0.5903).

3.1.2 Intra-language gender differences in the Hungarian data

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 jobs that showed a significant difference was *rendőr* 'police officer' (male=-1.73; female=-1.00) and *katona* 'soldier' (m=-2.18; f=-1.55), here the male biases were much higher by male raters. Two marginally significant different items were also found in *modell* 'model' (m=1.36; f=0.73) and *nővér* 'nurse' (m=2.55; f=1.91) showing a similar trend. The results are summarized in Figure 11.

[7] See the instructions here (HU) and here (ZH).

[8] You can download the raw data from here.

[9] All procedures and calculations can be inspected here.

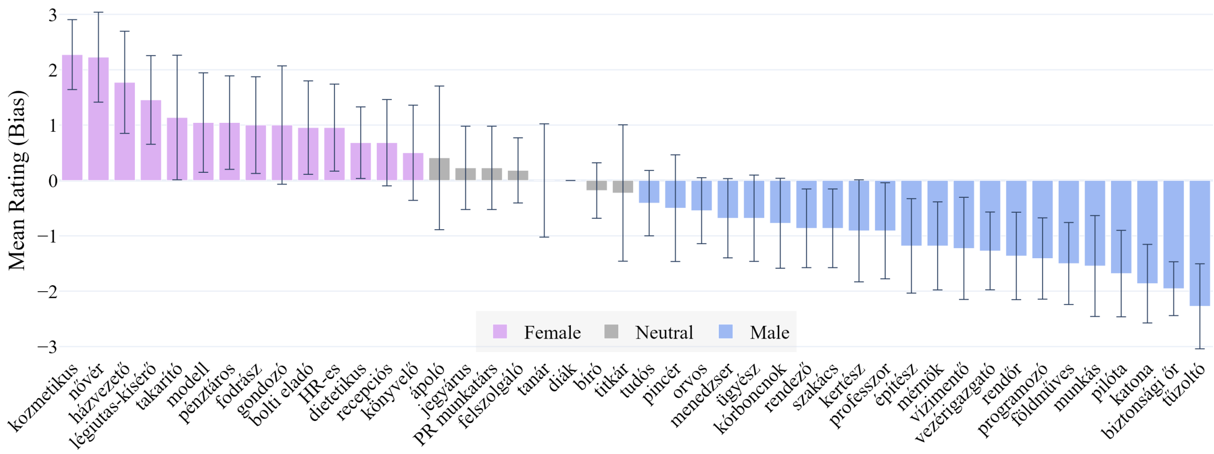

Figure 1: Mean ratings of occupational titles in Hungarian with standard deviations, significant gender bias highlighted – explore the interactive plot.

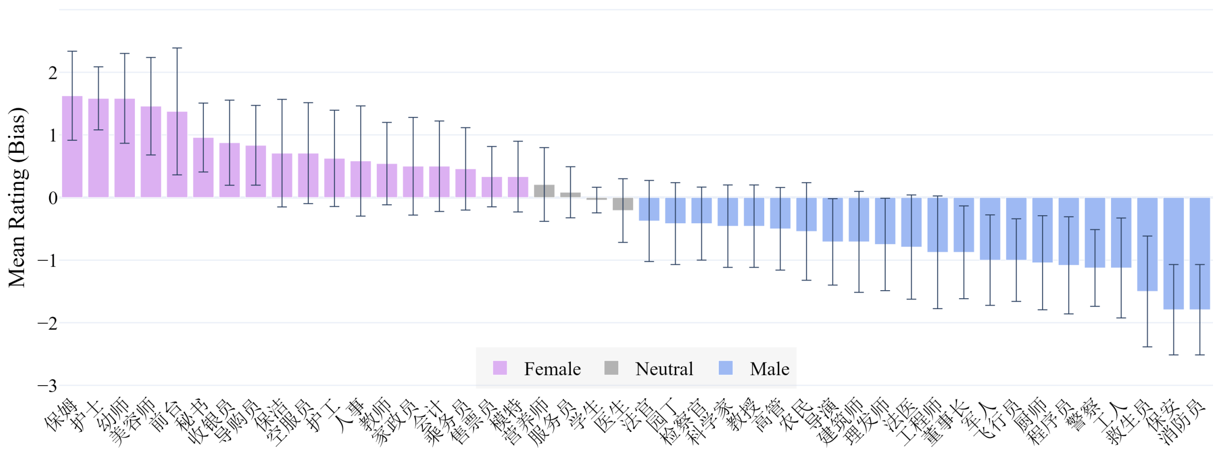

Figure 2: Mean ratings of occupational titles in Chinese with standard deviations, significant differences highlighted – explore the interactive plot.

Furthermore, these results seem to suggest that men's ratings tended to have a greater absolute bias for both male- and female-coded jobs (see Figure 4 below).

3.2 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 2.

3.2.1 Overall distribution of ratings

New from Wenhui from here...

The rating results largely aligned with our expectations concerning occupational gender stereotypes in Chinese society. Participants typically linked women to child-bearing, emotional labor, and service-oriented roles while associating men with high-intensity, high-risk physical labor, and order-maintaining roles.

...to here

In Chinese, the words with the highest female 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 'dietitian' (0.21), 服务员 'waiter/server' (0.08), 学生 'student' (-0.04), and

医生 'doctor' (-0.21).

Consistent with Hungarian results, Chinese data also showed a strong agreement on 学生 'student' (-0.04; std=0.2041) and 服务员 'water/server' (0.08; std=0.4082), the remaining words in the top 5 jobs with the lowest standard deviation were 售票员 'ticket seller' (0.33; std=0.4815), 护士 'nurse' (1.58; std=0.5036), and 医生 'doctor' (-0.21; std=0.5089).

3.2.2 Intra-language gender differences in the Chinese data

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). Marginally significant differences were found for 会计 'accountant' (m=0.71; f=0.20), 高管 'manager' (m=-0.29; f=-0.80), 工人 'worker' (m=-1.36; f=-0.80), and 保安 'security guard' (m=-2.00; f=-1.50). The results can be examined in Figure 12.

3.3 Cross-linguistic comparison

The wordlists of the two datasets were almost, but not exactly the same; by performing an inner join on the two lists, we could pair 42 items together according to their meanings. Then, using a two-sample *t*-test, we checked if there were significant differences between the ratings in the two languages. The results are summarized in Figure 3.

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 for example 'beautician' and 'nurse', as well as 'firefighter' and 'security guard'. Job titles that were neutral/unbiased in both datasets were 'server' (*felszolgáló*; 服务员), and 'student' (*diák*; 学生).

In terms of significant differences, we found 16 occupations that were rated differently in the two languages. These include 'beautician' (Hungarian=2.27 vs. Chinese=1.46), 'housekeeper' (H=1.77 vs. C=0.50), 'receptionist' (H=0.68 vs. C=1.38), 'flight attendant' (H=1.45 vs. C=0.46), 'model' (H=1.05 vs. C=0.33), 'dietitian' (H=0.68 vs. C=0.21), ''secretary' (H=-0.23 vs. C=0.96), 'teacher' (H=0.00 vs. C=0.54), 'hairdresser' (H=1.00 vs. C=-0.75), 'waiter' (H=-0.50 vs. C=0.08), 'doctor' (H=-0.55 vs. C=-0.21), 'gardener' (H=-0.91 vs. C=-0.42), 'farmer' (H=-1.50 vs. C=-0.54), 'pilot' (H=-1.68 vs. C=-1.00), and 'soldier' (H=-1.86 vs. C=-1.00), and 'firefighter' (H=-2.27 vs. C=-1.79).

New from Wenhui from here...

The occupational title 'secretary' elicits significantly different gender stereotypes among Hungarian and Chinese participants, which is a noteworthy observation. Hungarian participants perceive 'titkár' as a male-dominated profession (-0.23), whereas Chinese participants perceive it a predominantly female occupation (0.96). This discrepancy may arise from the interaction of lexical attributes and cultural context. In Hungarian, 'titkár' denotes both senior political officials and regular administrative employees. Conversely, the Chinese Occupational Classification Dictionary (Occupational Code: 3-01-02-02) delineates a 秘书 *mìshū* 'secretary' as an office service employee involved in clerical and meeting-related responsibilities. This definition may delay the manifestation of in societal perception and gender expectations, yet it largely corresponds with public expectations of the secretary profession. The semantic scope and societal expectations of identical occupational titles in Hungarian and Chinese illustrate how cultural background influences gender bias in vocabulary.

The gender stereotype linked to 'housekeeper' is predominantly female in both cultures; however, the extent of this bias varies considerably (Chinese: 0.5 vs. Hungarian: 1.77). In the results of Chinese participants, it is pertinent to compare the terms 家政员 *jiāzhèngyuán* meaning 'housekeeper' and 保姆 *baómŭ* meaning 'nanny', which are analogous in meaning yet vary in register. While the term 'housekeeper' (0.5) demonstrates less gender bias than the colloquial term 'nanny' (1.63), it remains categorised as a female-dominated profession due to the traditional female-associated traits of household tasks and emotional management suggested by the central term 家政 *jiāzhèng* 'housekeeping'. This domain-specific scoring disparity indicates that although the formal and informal characteristics of occupational terminology mitigate gender stereotypes, societal gender stereotypes regarding occupations may be intricately linked to the semantic implications of language. Particular terminology related to domestic or emotional labour is more prone to evoke conventional female gender bias perceptions, thus reinforcing the gendered classification of professions.

The disparities in gender perception between 'farmer' and 'worker' are particularly signifi-

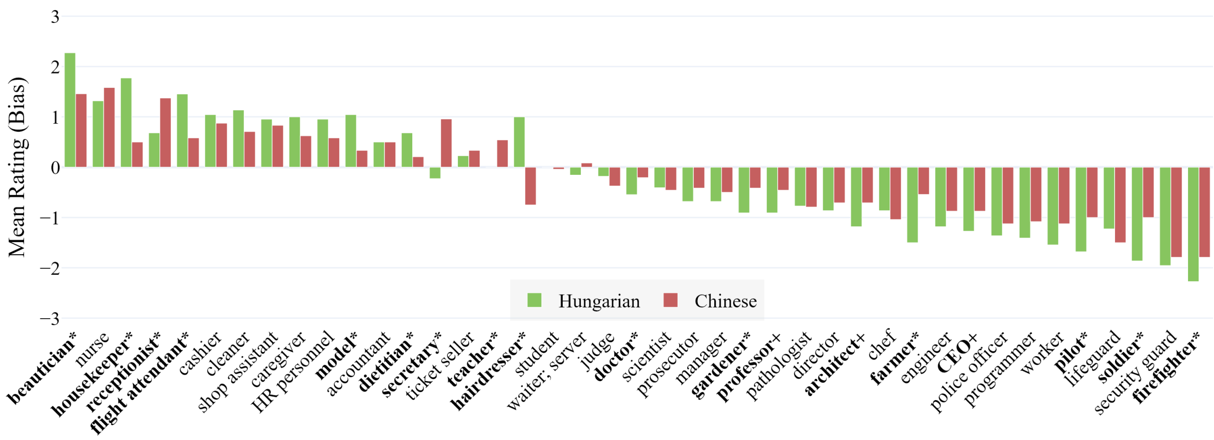

Figure 3: Mean ratings of common occupational titles in Hungarian and Chinese (significant*, and marginally significant+ in **bold**) – explore the interactive plot.

cant. Data from Chinese participants indicates that 'farmer' (-0.54) demonstrates a comparatively weaker male orientation than 'worker' (-1.13), whereas in Hungarian, the two are nearly identical (farmer: -1.50 vs. worker: -1.55). This discrepancy can be attributed to China's societal and economic development trajectory. Throughout the era of collective farming in China (1950s–-1970s), the work-point accounting system and production brigades promoted women's involvement in agricultural labour, while slogans such as "Women hold up half the sky" (妇女能顶半边天) enhanced the visibility of women in this sector (Jacka, 1997a,b). Despite the Household Contract Responsibility System established in the 1980s reinforcing the dominance of male household heads and consequently marginalising women's land rights (Zhu, 2009), the significance of women in agricultural production has progressively become apparent [data is needed here to show the proportion of women in the Chinese agriculture industry]. During this period, township enterprises and factories employed female labour; however, they confined women to light industrial positions through technical qualification prerequisites (Liu, 2007). This led to a gendered labour division in which women in industry participated in unskilled roles, while men retained technical monopolies (Bossen, 2002). This historical process has resulted in the ongoing divergence of gender perceptions between the agricultural and industrial sectors. Despite contemporary ideals of occupational gender equality, gender stereotypes in lower-tier professions continue to demonstrate significant historical persistence. The disparities in gender perception between 'farmer' and 'worker' are particularly significant. Data from Chinese participants indicates that 'farmer' (-0.54) demonstrates a comparatively weaker male orientation than 'worker' (-1.13), whereas in Hungarian, the two are nearly identical (farmer: -1.50 vs. worker: -1.55). This discrepancy can be attributed to China's societal and economic development trajectory. Throughout the era of collective farming in China (1950s–-1970s), the work-point accounting system and production brigades promoted women's involvement in agricultural labour, while slogans such as "Women hold up half the sky" (妇女能顶半边天) enhanced the visibility of women in this sector (Jacka, 1997a,b). Despite the Household Contract Responsibility System established in the 1980s reinforcing the dominance of male household heads and consequently marginalising women's land rights (Zhu, 2009), the significance of women in agricultural production has progressively become apparent [data is needed here to show the proportion of women in the Chinese agriculture industry]. During this period, township enterprises and factories employed female labour; however, they confined women to light industrial positions through technical qualification prerequisites (Liu, 2007). This led to a gendered labour division in which women in industry participated in unskilled roles, while men retained technical monopolies (Bossen, 2002). This historical process has resulted in the ongoing divergence of gender perceptions between the agricultural and industrial sectors. Despite contemporary ideals of occupational gender equality, gender stereotypes in lower-tier professions continue to demonstrate significant historical persistence.

...to here

The most striking dissimilarity was the word for 'hairdresser' (*fodrász* vs. 理发师), which shows a strong female bias in Hungarian (1.00) but a definite male bias in Chinese (-0.75). We believe that this is a neat example for a cultural difference, in Hungary hairdressers are perceived to be predominantly female, while in China the profession is perceived to be more male-dominated. It is not difficult to find evidence for the latter from the press, even in English-language media.[10]

3.4 Discussion

This in itself is not surprising...

The first obvious thing to notice when looking at Figure 3 is that the overall trends in gender bias are quite similar between the two languages, despite some notable differences in specific occupations. This suggests that stereotypes are comparable and – mostly – consistent across developing societies.

3.4.1 Gender bias by language

An interesting feature is that Hungarian raters tended to rate occupations with a stronger bias than Chinese speakers. Out of 42 shared occupations, 31 were rated with a significant bias in Hungarian, while only 10 were rated with a higher bias in Chinese (excluding the 1 item with equal value). This is a threefold difference, and warrants further exploration to fully explain.

Gabor: Tie in Kaukonen et al. here

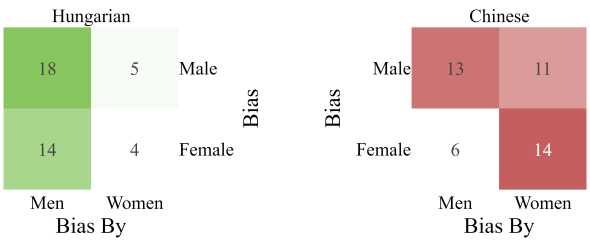

Figure 4: Confusion matrix of the Hungarian and Chinese ratings, showing the differences in the ratings of male and female participants.

3.4.2 Gender bias by language users' gender

An interesting divergence arose when comparing the two datasets and the differences between the ratings by gender. While in Hungarian, biases – both male and female – were stronger in the ratings

[10]https://www.chinadailyhk.com/hk/article/603100

Table 1: Correlation between human and AI ratings for occupational titles in Hungarian and Chinese, sorted by the correlation coefficient.

#	**Hungarian**	**R**	**p_value**
1	LeChat	0.9634	< 0.01
2	Gemini	0.9529	< 0.01
3	ChatGPT	0.9525	< 0.01
4	Copilot	0.9255	< 0.01
5	DeepSeek	0.8768	< 0.01
#	**Chinese**	**R**	**p_value**
1	DeepSeek	0.9504	< 0.01
2	Gemini	0.9452	< 0.01
3	LeChat	0.9313	< 0.01
4	ChatGPT	0.9235	< 0.01
5	Copilot	0.9156	< 0.01

of men, in Chinese women rated with a stronger bias on average, especially regarding female biases. You can compare the contrast in Figure 4.

4 Comparing human ratings with AI-generated ratings

Now, we were also curious how the human ratings compare to the ratings of LLMs used by popular AI agents and chatbots, so we repeated the same experiment using artificial intelligence instead of human raters. We prompted Mistral's Le Chat, Copilot, ChatGPT, Deepseek, and Gemini, as mentioned above. The mean AI ratings and the standard deviations were then superimposed on the human raters' plots to allow for a convenient comparison, you can see these in Figures 5 and 6.

The rationale behind this was to simulate how the general public and non-experts would turn to AI to do the job of human raters, so we purposefully avoided any pre-training or fine-tuning that requires a degree of technical know-how, and used the best available versions within a free quota on their respective web interfaces that were available in this timeframe. We performed a correlation test using Pearson's R to compare the human and AI ratings to find out which models did the best compared to humans. We found very strong positive correlations between the sets of ratings in both languages, see Table 1.

The results show that the LLMs in these AI agents are performing really well, showing converging trends. For Hungarian, the closest results showing the highest correlation of 0.9634 were achieved by using Mistral AI's Le Chat, the only European

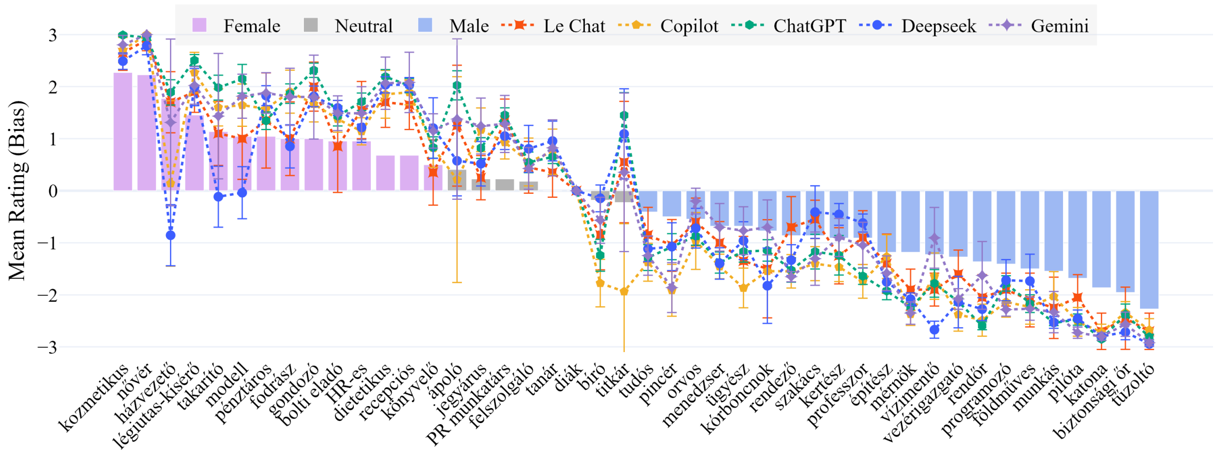

Figure 5: AI ratings for Hungarian occupations — explore the interactive plot.

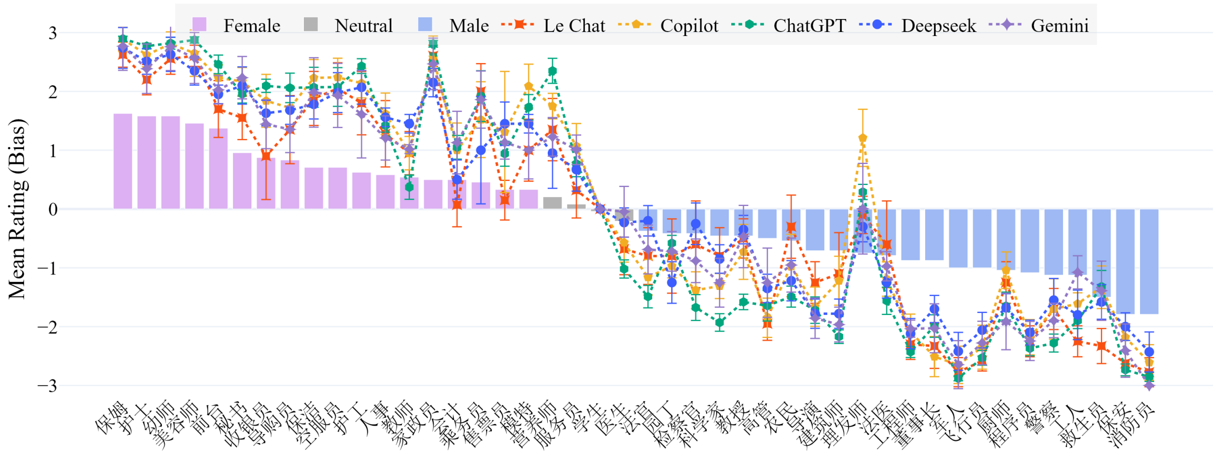

Figure 6: AI ratings for Chinese occupations — explore the interactive plot.

model in our experiment. In turn, the best results for Chinese were obtained with DeepSeek, which achieved a correlation of 0.9504.

However, it is to be emphasised that these practices can be highly problematic, as the biases encoded in the LLMs are even stronger than those of the human raters (see), and publishing studies that use LLMs instead of participants can perpetuate gender biases. We think that the results perfectly illustrate both the pros and cons/dangers of using AI agents instead of human raters, especially as they can reinforce social biases.

5 Conclusion

Finish conclusion...

…

For the sake of reproducibility and open science, we have made all the data and code available in this repository: https://anonymous.4open.science/r/occupational-bias-paclic39.

ANONYMIZE ALL

References

Haozhe An, Connor Baumler, Abhilasha Sancheti, and Rachel Rudinger. 2025. On the Mutual Influence of Gender and Occupation in LLM Representations. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1663–1680.

Dennis E. Baron. 1986. Grammar and Gender, volume 658. Yale University Press.

畢新偉 Xinwei Bi. 2013. 「她」字的來源與女性主體性——評《「她」字的文化史》 [The Origin of Chinese Characters and Female Subjectivity: A Review of “The Cultural History of the Chinese Character ‘She’ ”]. 二十一世紀 [21st century], 136:136–142.

Laurel Bossen. 2002. Chinese Women and Rural Development: Sixty Years of Change in Lu Village, Yunnan. CABI.

Valentina Canessa-Pollard, David Reby, Robin Banerjee, Jane Oakhill, and Alan Garnham. 2022. The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. Journal of Vocational Behavior, 134:103703.

Kay Deaux and Laurie L. Lewis. 1984. Structure of gender stereotypes: Interrelationships among components and gender label. Journal of personality and Social Psychology, 46(5):991.

Catherine S. Farris. 1988. Gender and Grammar in Chinese: With Implications for Language Universals. Modern China, 14(3):277–308.

Alan Garnham, Sam Doehren, and Pascal Gygax. 2015. True gender ratios and stereotype rating norms. Frontiers in Psychology, 6.

Marie Gustafsson Sendén, Emma A. Bäck, and Anna Lindqvist. 2015. Introducing a gender-neutral pronoun in a natural gender language: The influence of time on attitudes and behavior. Frontiers in psychology, 6:893.

Pascal Gygax, Ute Gabriel, Oriane Sarrasin, Jane Oakhill, and Alan Garnham. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. Language and Cognitive Processes, 23(3):464–485.

Pascal M. Gygax, Alan Garnham, and Sam Doehren. 2016. What Do True Gender Ratios and Stereotype Norms Really Tell Us? Frontiers in Psychology, 7.

Marlis Hellinger. 1990. Kontrastive Feministische Linguistik: Mechanismen Sprachlicher Diskriminierung Im Englischen Und Deutschen. Hueber, Ismaning.

Marlis Hellinger and Hadumod Bußmann. 2003. Gender across languages: The linguistic representation of women and men. In Marlis Hellinger and Hadumod Bußmann, editors, Gender Across Languages: The Linguistic Representation of Women and Men, number 11 in IMPACT: Studies in Language, Culture and Society, pages 1–25. John Benjamins Publishing Company, Amsterdam.

Tamara Jacka. 1997a. Agriculture. In Women's Work in Rural China: Change and Continuity in an Era of Reform, Cambridge Modern China Series, pages 120–142. Cambridge University Press, Cambridge.

Tamara Jacka. 1997b. Domestic work. In Women's Work in Rural China: Change and Continuity in an Era of Reform, Cambridge Modern China Series, pages 101–119. Cambridge University Press, Cambridge.

Da Ju, Karen Ullrich, and Adina Williams. 2024. Are female carpenters like blue bananas? a corpus investigation of occupation gender typicality. In Findings of the Association for Computational Linguistics: ACL 2024, pages 4254–4274, Bangkok, Thailand. Association for Computational Linguistics.

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.

Hannah Rose Kirk, Yennie Jun, Haider Iqbal, Elias Benussi, Filippo Volpin, Frederic A Dreyer, Aleksandar Shtedritski, and Yuki M Asano. 2021. Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. In 35th Conference on Neural Information Processing Systems (NeurIPS 2021).

William Labov. 1972. Sociolinguistic Patterns. University of Pennsylvania Press.

Karla A. Lassonde and Edward J. O'Brien. 2013. Occupational stereotypes: Activation of male bias in a gender-neutral world. Journal of Applied Social Psychology, 43(2):387–396.

Asaf Levanon and David B. Grusky. 2016. The persistence of extreme gender segregation in the twenty-first century. American Journal of Sociology, 122(2):573–619.

Molly Lewis and Gary Lupyan. 2020. Gender stereotypes are reflected in the distributional structure of 25 languages. Nature Human Behaviour, 4(10):1021–1028.

Charles N Li and Sandra A Thompson. 1989. Mandarin Chinese: A functional reference grammar. University of California Press.

Jieyu Liu. 2007. Gender and Work in Urban China: Women Workers of the Unlucky Generation. Routledge.

Julia Misersky, Pascal M. Gygax, Paolo Canal, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdrova, Lisa Von Stockhausen, and Sabine Sczesny. 2014. Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. Behavior Research Methods, 46(3):841–871.

Csaba Oravecz, Tamás Váradi, and Bálint Sass. 2014. The Hungarian Gigaword Corpus. In LREC 2014, volume 9, pages 1719–1723.

Jerome L. Packard. 2000. The Morphology of Chinese: A Linguistic and Cognitive Approach. Cambridge University Press.

Anne Pauwels. 1997. Of handymen and waitpersons: A linguistic evaluation of job classifieds. Australian journal of communication, 24(1):58–69.

Anne Pauwels. 1998. Women Changing Language. Longman London.

Anne Pauwels. 2003. Linguistic sexism and feminist linguistic activism. The handbook of language and gender, pages 550–570.

Alma Sabatini. 1985. Occupational Titles in Italian: Changing the Sexist Usage. In Marlis Hellinger, editor, Sprachwandel Und Feministische Sprachpolitik: Internationale Perspektiven, pages 64–75. VS Verlag für Sozialwissenschaften, Wiesbaden.

Dagmar Stahlberg, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2011. Representation of the sexes in language. In Social communication, pages 163–187. Psychology Press.

Qi Su, Pengyuan Liu, Wei Wei, Shucheng Zhu, and Chu-Ren Huang. 2021. Occupational gender segregation and gendered language in a language without gender: Trends, variations, implications for social development in China. Humanities and Social Sciences Communications, 8(133):1–10.

J. R. Sun. 1997. 语言与性别 [Language and Gender]. Jiangsu Education Press.

Louise O Vasvári. 2014. Problémás nyelvi nem a nem nélküli magyar nyelvben [Problematic linguistic gender in the genderless Hungarian language]. Társadalmi Nemek Tudománya Interdiszciplináris eFolyóirat, 4(1):130–166.

Tamás Váradi. 2002. The Hungarian National Corpus. In LREC 2002, pages 385–389.

Marina Yaguello, Gérard Lauzier, and Claire Bretécher. 1978. Les Mots et Les Femmes: Essai d'approche Socio-Linguistique de La Condition Féminine, volume 75. Payot Paris.

Ling Zhu. 2009. Gender Inequality in the Land Tenure System of Rural China. In China's Economy, volume Volume 1 of Series on Developing China ? Translated Research from China, pages 21–36. WORLD SCIENTIFIC.

Appendix

Hungarian items

modell, *katona* 'soldier', *kórboncnok* 'pathologist', *vezérigazgató* 'CEO', *menedzser*, 'manager' *nővér*, 'nurse' *szakács* 'chef', *felszolgáló* 'server', *könyvelő* 'accountant', *professzor* 'professor', *építész* 'architect', *tudós* 'scientist', *ápoló* 'nurse', *pénztáros* 'cashier', *bíró* 'judge', *munkás* 'worker', *vízimentő* 'lifeguard', *jegyárus* 'ticket seller', *tűzoltó* 'firefighter', *mérnök* 'engineer', *rendező* 'director', *takarító* 'cleaner', *HR-es* 'HR specialist', *házvezető* 'housekeeper', *légiutas-kísérő* 'flight attendant', *pincér* 'waiter', *orvos* 'doctor', *fodrász* 'hairdresser', *földműves* 'farmer', *gondozó* 'caregiver', *bolti eladó* 'shop assistant', *kertész* 'gardener', *titkár* 'secretary', *PR munkatárs* 'PR officer', *dietetikus* 'dietitian', *tanár* 'teacher', *rendőr* 'police officer', *pilóta* 'pilot', *recepciós* 'receptionist', *biztonsági őr* 'security guard', *ügyész* 'prosecutor', *kozmetikus* 'beautician', *programozó* 'programmer', *diák* 'student'.

Chinese items

警察 'police', 秘书 'secretary', 教授 'professor', 护士 'nurse', 高管 'manager', 教师 'teacher', 前台 'receptionist', 工人 'worker', 幼师 'kindergarten teacher', 模特 'model', 护工 'caregiver', 保姆 'nanny', 会计 'accountant', 工程师 'engineer', 保洁 'cleaner', 法官 'judge', 导购员 'shop assistant', 美容师 'beautician', 服务员 'waiter', 乘务员 'flight attendant'[11], 理发师 'hairdresser', 空服员 'flight attendant'[12], 售票员 'ticket seller', 厨师 'chef', 营养师 'nutritionist', 家政员 'housekeeper', 收银员 'cashier', 医生 'doctor', 法医 'pathologist', 程序员 'programmer', 保安 'security guard', 导演 'director', 军人 'soldier', 董事长 'CEO', 农民 'farmer', 学生 'student', 园丁 'gardener', 飞行员 'pilot', 人事 'HR personnel', 消防员 'firefighter', 科学家 'scientist', 检察官 'prosecutor', 救生员 'lifeguard', 建筑师 'architect'.

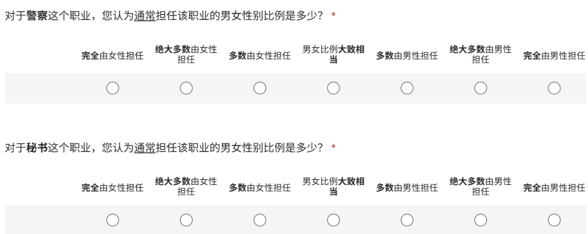

Figure 7: A sample of the the Chinese survey layout.

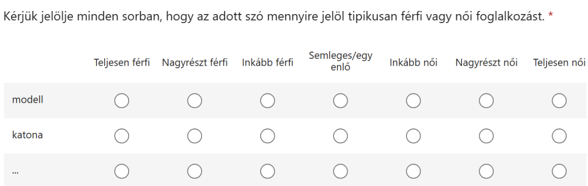

Figure 8: A sample of the the Hungarian survey layout.

[11] Also the attendant/crew on high-speed trains.
[12] Less frequent word; only on airplane.

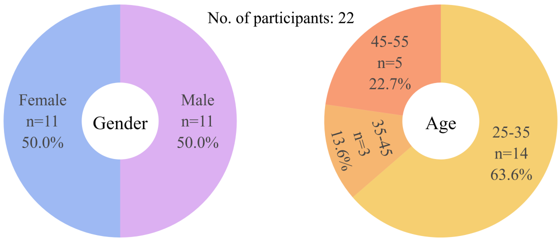

Figure 9: Demographics of the Hungarian participants.

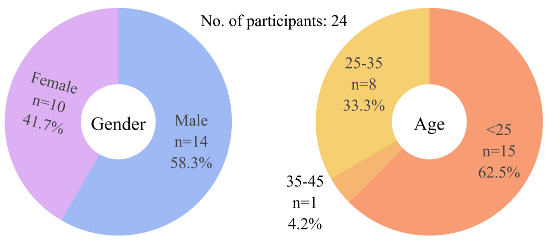

Figure 10: Demographics of the Chinese participants.

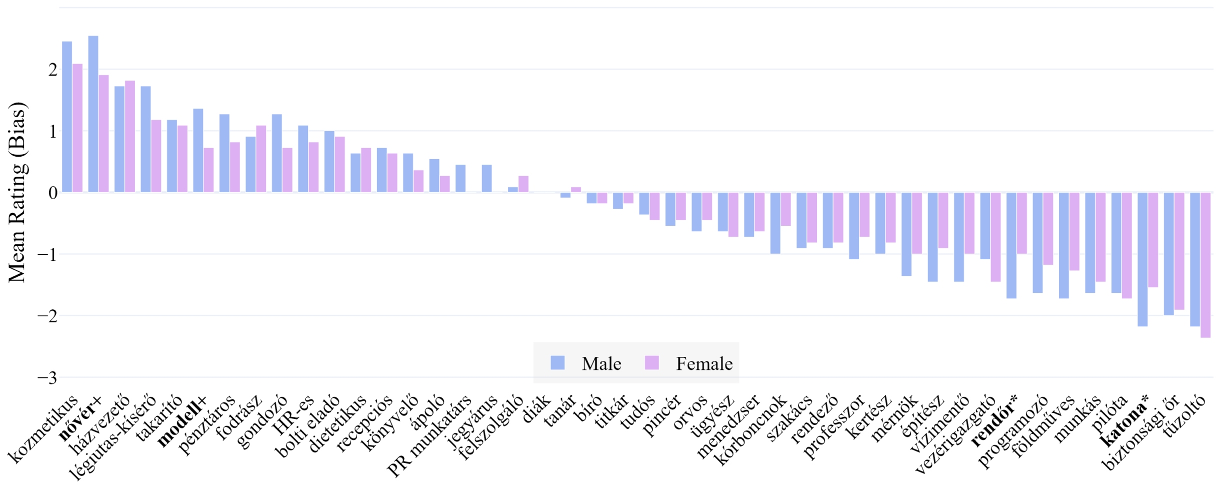

Figure 11: Mean ratings of occupational titles in Hungarian by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – explore the interactive plot.

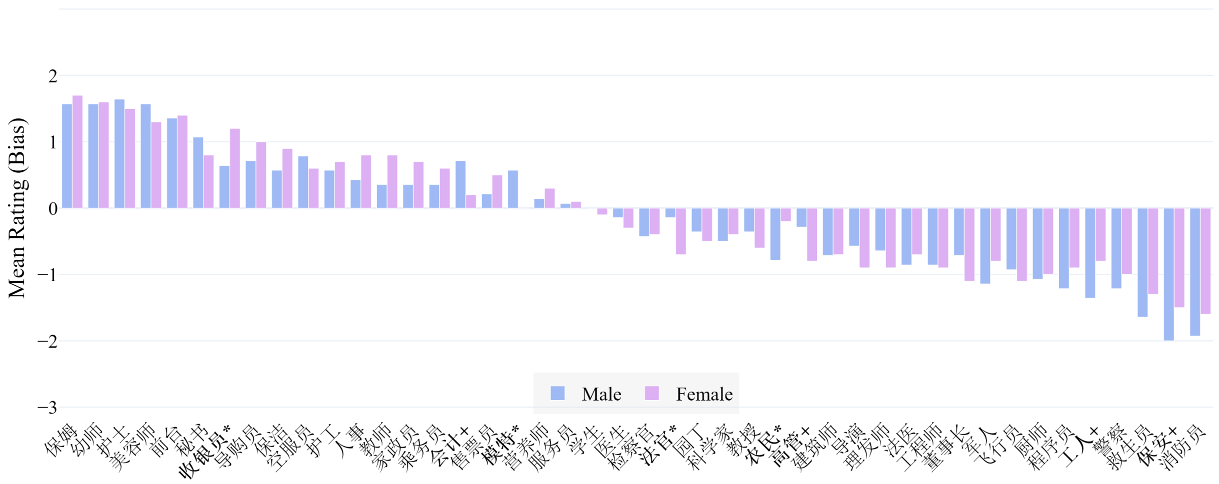

Figure 12: Mean ratings of occupational titles in Chinese by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – explore the interactive plot.