Universiteit Leiden

Master Media Technology

Reimagining Chinese Characters: Exploring Self-Perception Through Gendered Radical Interventions

Name: Huien Tan
Student ID: s3892026

Date: 07/07/2025

1st supervisor: Tessa Verhoef
2nd supervisor: Qinyu Chen

Master's Thesis in Media Technology

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

Abstract

According to the theory of linguistic relativity, the language we use can shape how we see the world (Whorf, 1956). This study explores whether changing radicals (the building blocks of Chinese characters) can affect how people see themselves. For Chinese words that have negative meanings and include the "女" (meaning: female) radical, two changes were applied: 1) replacing them with synonyms that do not contain the female radical (M1), and 2) replacing the female radical with a more neutral one (M2). For positive words that originally do not include the "女" (female) radical, the female radical was added in (M3). Participants were asked to rate sentences that described personal traits, some with original characters, the rest with modified versions, and to indicate how well each sentence matched the way they perceive themselves. The study result is that overall, the modified sentences didn't show a significant difference from the original ones. However, female participants were more influenced by M1, and the impact of M2 and M3 varied depending on the word's meaning. As a creative output, a user-friendly feminist Chinese input method is implemented based on users' feedback and suggestions on the basis of conducted interviews. It provides various typing options including original characters and modified characters, and also serves as a technical tool for exploring more female-friendly expressions in every day's digital communication.

Keywords: Linguistic relativity; Gendered Chinese characters; Feminist language reform; Radical modification; Self-perception; Chinese input method

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

This research is based on the theory of linguistic relativity, which suggests that the language we use can shape how we think and how we see the world (Whorf, 1956; Sapir, 1929). While many previous studies explored how language can shape cognition in the long term, this research focuses on something more immediate: how small changes in language can affect how people feel at the moment.

One earlier study by Vainapel and her colleagues (2015) showed just how powerful even tiny language tweaks can be. The study found that when women read a task description using only "he," they felt less motivated and confident than those who saw "he or she." Inspired by that, this paper explores whether micro-level changes to Chinese characters, especially changes to the gendered radical "女"(female), can affect how users feel about themselves immediately after exposure.

This focus on perception rather than cognition is deliberate as the study also aims to produce a technological intervention. As argued by Fogg (2002), the effectiveness of technological interventions often lies not in their ability to rewire belief systems, but in their capacity to produce immediate, tangible emotional effects—such as satisfaction, recognition, or resistance.

1.2 Motivation

In the Chinese writing system, female is not a neutral representation. Researchers (Chin and Burridge, 1993) have shown that the character "阴", which is a representation of female in Taoism, is around 90% likely to convey negative meaning or reinforce traditional gender stereotypes. Also, among all Chinese characters that carry the female radical, 18.6% of them are negative (Wang et al., 2023).

In an era where more and more people are influenced by feminism, various efforts have been made in China to tackle this situation. Influencers made videos to promote non-misogynist

curse words (e.g. *Why do so many swear words involve people's moms* on Xiaohongshu, 2024), editors re-interpretated characters that carry female radicals in a more positive way (*A "dictionary" composed solely of Chinese characters with the female radical* on Xiaohongshu, 2025), artists initiated exhibitions to engage women to resonate with positive characters with female radicals 2023), designers developed new vernacular of feminist-leaning words and phrases (Stinson, 2016). These efforts have raised awareness about the gendered nature of some Chinese characters, but they often rely on repeating existing narratives, limiting the influence to individual reflection rather than structural change.

What remains missing is a practical, scalable tool that allows people to take part in changing language from within, not just by seeing, but by using. Without such tools, gender bias in language stays untouched at the everyday level of communication.

Benjamin (2024) emphasized the importance of creating 'new stories' as a way to bring more just futures for everyone.In the current digital age, tools like Chinese input methods (e-keyboard) are widely used to produce online language content, and the UTF-8 standard has made the evolution of the shape and construction of Mandarin characters more static as it resists natural evolution. This makes conscious efforts toward language reform and gender inclusivity even more important. Therefore, an additional aim of this study is to offer a feminist e-keyboard as the final creative output to help widely spread opinions and bring changes to the society.

1.3 Research Question and Hyphothesis

This study explores how gendered radicals in Chinese characters influence users' self-perception. Specifically, the following questions guide the research:

RQ1: How do different types of linguistic intervention—synonym replacement (M1), radical modification (M2), and positively gendered character creation (M3)—affect participants' self-perception, compared to the original form (O)?

- **Hypothesis 1 for RQ1:** For answering RQ1, a hypothesis has been made that M1, M2, M3 will influence people significantly in a positive way, compared to their corresponding original from.

- **Hypgothesis2 for RQ1:** People who have stronger feminist beliefs will be more influenced by the modified characters (M1, M2, M3), because these changes are directly related to gender representation in language.

RQ2: When presented through a feminist Chinese input method (e-keyboard), how do users choose between different types of interventions (M1/M2/M3), and in what contexts are they most willing to use them?

1.4 Research Design

This study uses two-phase to explore how different types of changes to Chinese characters influence how people see themselves (RQ1), and how they feel about using these new forms in real writing situations (RQ2).

The study includes two parts (more details will be explained in Methodology):

- **Phase 1:** A three-week online survey is conducted to examine how people respond to the same sentence carried with different word forms. Participants are asked to rate how well each sentence described themselves.
- **Phase 2:** Interview within a small number of chosen participants will be conducted, questions about how they feel about the modified characters, and how they would like the e-keyboard to be like will be asked

The two phases are closely connected. The results from Phase 1 will help decide what to focus on in the interviews in Phase 2.

2 Literature Review

2.1 Introduction

Language is not only a tool for communication, it can also help shape thought processes and social structures (Boroditsky et al., 2003) and influence individual perception (Whorf,

1956), including our views on gender (Wolff and Holmes, 2011). While many languages contain embedded gender stereotypes (Lewis and Lupyan, 2020), the use of gender-fair language has been shown to reduce stereotypes and their negative effects such as the potential to lead to discrimination (Sczesny et al., 2016). Additionally, digital technology interventions have proven to be a potential reformative tool to change people's mind (Fogg, 2002).

This literature review explores the intersection of language, gender bias, and technological intervention. The key questions guiding this review are:

- How does language reflect and reinforce gender bias?
- How can technological interventions influence users' biased thought?
- What are the challenges and possibilities of gender language reform in the Chinese context?

2.2 Language and Gender Bias: Theoretical Foundations

2.2.1 Linguistic Relativity and Gendered Language

Gendered language shapes the world in a negative way and makes stereotypical thoughts on gender stronger. Lewis and Lupyan (2020) found that gender stereotypes are deeply embedded in the distribution of words across many languages, and the words associated with women tend to carry more negative connotations than those associated with men. Whorf (1956) and Sapir (1929) suggested that the structure of a language can shape how its speakers perceive and think about the world. Building on this theory, many scholars have dived into more specific fields.

The study done by Boroditsky and her colleagues (2003) provides further evidence for the influence of language on gender perception. They investigated grammatical gender in Spanish and German, the result showed that speakers unconsciously attribute characteristics to objects based on the languages' grammatical gender. For example, in German, the word for "bridge" (die Brücke) is feminine, making German speakers to describe bridges using words like "elegant" and "beautiful." In Spanish where "bridge" (el puente) is masculine, speakers were more likely to describe them as "strong" and "sturdy."

Beside grammatical gender, pronouns and generic terms also affect gender perception. Gastil (1990) found that even linguistic choices that seem neutral (such as using “he” as a default pronoun) can stress male-centered gender hierarchies. He demonstrated that when participants encountered ”he” as a generic pronoun, they were significantly more likely to visualize a male figure rather than a neutral or female figure.

Gendered language not only reflects but also sustains and actively constructs social gender inequalities. In Language and Woman’s Place, Lakoff (1973) analyzed English gendered expressions and found that women are encouraged to use more ”polite” and ”tentative” language, such as hedging expressions (e.g., sort of, kind of) and rising intonations. De Francisco (1992) looked at the differences in male and female conversational styles, and found that women tend to use more cooperative language while men use the more competitive ones. In her opinion, this kind of difference in communication styles would maintain the social stereotypes of male assertiveness and female passivity. Similar to this study, Cameron (2014) studied how young men use language to construct their sense of heterosexual masculinity. She found that they do it by making negative comments on women, as this help them feel more belonging to the group while keeping the feeling of being dominant.

2.2.2 The Social Effects of Gendered Language

The influence of using gendered language can also go beyond individuals and bring inequality to the more broadened fields such as education, employment and social interactions. Stahlberg and her colleagues (2007) investigated the impact of gendered language on how people form expectations about social roles. The research shows that masculine generics (e.g., Lehrer in German, referring to ”teacher” in a default masculine form) decrease the mental presence of women in professional settings. And the active use of masculine generics in hiring decisions decreases the likelihood of women being considered for male-dominated positions. Also, Prewitt-Freilino and her colleagues (2011) investigated the relationship between the gendered level of a language and how it relates to the gender equality level of the country who speaks it. Their study categorized languages into three types: gendered languages (e.g., Spanish, German); natural gender languages (e.g., English, Swedish), and genderless languages (e.g., Finnish, Turkish). The finding suggests that countries where gendered languages dominate tend to exhibit lower levels of gender equality in employment,

education, and political representation.

2.2.3 International Gender Language Reforms

Many societies have made efforts to encourage language reform in order to tackle the gender bias embedded in their language. For example, 'they/them' starts to become an inclusive pronoun in English (Saguy and Williams, 2021). Also in French, inclusive spelling was introduced into the language. Midpoint (e.g., "étudiant·e" for "student") was used to include both masculine and feminine forms (Pozniak et al., 2024). But this kind of reform also sparked wide discussion in the French society. As indicated by Viennot (2017), acceptancy of gender inclusive language varies in different French speaking countries, with Switzerland and Canada being more receptive, and France less. Similarly, in Spanish speaking countries, the term "Latinx" started to be used as a gender-neutral alternative to Latino/a, aiming at increasing inclusiveness for nonbinary people (Molina et al., 2024).

Beyond movements that happened within a specific language region, there is also an international movement called Gender Free Language movement. The movement was proposed in 1987 by Canadian and Nordic countries to UNESCO (Sczesny et al., 2016). In 1999, an official guideline was published. The guidelines advocate for gender-neutral language, and emphasize that "language not only reflects the ways of thinking but also shapes them. If words and expressions that includes inferiority against female are used frequently, it will become an assumption that's embedded in our cognition; therefore, as our perceptions evolve, our language must also adapt" (Desprez-Bouanchaud et al., 1999; Sczesny et al., 2016). Similarly, European Parliament issued language guidelines applicable to all official EU working languages to keep up with UNESCO's standpoint of GFL (Papadimoulis and Parliament, 2018; Sczesny et al., 2016). However, despite all those efforts, neither UNESCO nor the EU enforces these guidelines to be something legal for member states. The adoption of gender-fair language remains voluntary at the national level.

2.3 Historical Evolution of the ” 女” Radical and Current Challenges

In Mandarin, radicals are important building blocks of characters, and it also carries meanings. The radical “女 (female)” has seen the history of social changes, and is a reflection of family ethics and gender roles. In early China, women played significant roles and this can be told from ancient surnames such as “姜” (Jiang), “姬” (Ji), and “姚” (Yao). All of them were once associated with ruling clans and contains female radical. However, Chinese society gradually turned patriarchal, which leads to a decline in women’s status. Terms like “妻” (Qi, which means wife) and “妾” (Qie, which means concubine), which reflect women’s role in a hierarchical familial structure, started to appear (Zhao, 2003; Wang, 2016).

By Han Dynasty, influenced by Confucian ideology, characters started to reflect both moral expectations and societal stereotypes on women, which can be told from characters like “嫉” and “妒” (Ji and Du, both means jealous). This trend continued in the following dynasties, where words that describe women’s beauty and elegance, such as ” 姝” (Shu, fair/beautiful) and ” 娴” (Xian, elegant), coexisted with empirical social practices like foot-binding (Zhao, 2003; Wang, 2016).

Xie (2018) also investigated how the “女” (female) radical has been used in Chinese characters over time. She found that many characters with this radical have negative meanings, such as “奸” (Jian, treacherous), “妒” (Du, jealous), and “奴” (Nv, slave). She suggested that this situation showed how women were often seen negatively in traditional Chinese society.

In contemporary China, the Reformation of Simplified Chinese Character started in 1956. It aimed to improve literacy, but also leads to changes in gendered language. Some in positive ways: a few negative words that originally contained the “女” (female) radical got rid of its female radical. Some in negative ways: the character “她” (Ta, she/her) was introduced to mimic Western gendered pronouns, creating a binary gender distinction that wasn’t there in classical Chinese (Ling, 1989; Huang, 2023).

Nowadays, gender stereotypes remain prevalent in Chinese. According to an exibition focuses on Chinese characters with the female radical (Wang et al., 2023), artists and curators found

that in all 955 chracters that carries female radical, 178 of them have negative meanings.

2.4 Influential Technological Interventions

2.4.1 Technological Intervenetions to Combat Biases

Biased prediction happens in technology across many languages. Caliskan and her colleagues (2017) found that AI-driven autocomplete functions in English tend to recommend gender-biased occupational terms. For example, the word "doctor" is more frequently associated with "he", while "nurse" more with "she". Similarly, Zhao and his colleagues (2021) examined gender bias in Chinese pretrained language models and found that there are systematic sexist associations.

Technological interventions have been proved to be able to influence human behavior and attitude. Fogg (2002) introduced the concept of persuasive technology and explained how digital interfaces can be designed to subtly shape moves and minds. Research showed that biased predictive text and default options may unintentionally reinforce gender biases. As these models are commonly used in predictive text, their embedded biases can gradually influence the way people write and speak without even noticing (Bhat et al., 2021; Arnold et al., 2018). Also, nonverbal cues can also evoke changes. Bailenson and Yee (2005) found that even small nonverbal cues in virtual environments (e.g., avatar gestures) has an impact on how users behaved. Peck and his colleagues (2013) extended this finding by showing that virtual embodiment in a Black avatar can reduce racial bias.

2.4.2 Nudging Inclusive Language via Interface Design

Human language choices are not made in a vacuum—they are influenced by both cognitive efficiency and habitual exposure. Bybee (2010) wrote that users prefer frequently occurring, cognitively less demanding words and expressions, and over time, these patterns would become solidified in their mind. Also, Pickering and Garrod (2004) found that linguistic priming make people gradually adopt certain linguistic forms after repeated exposure.

Besides, the design of digital interfaces is also important in shaping user behavior. Keegan

and Evas (2012) demonstrated that software interfaces could employ nudging techniques (such as default settings) to encourage users to engage with minority language ICT interfaces, which indicates that user choices in digital environments can be shaped by subtle interface cues rather than active decision-making.

While there is limited research specifically on how input methods can promote gender-fair language, the above studies proves the feasibility of using e-keyboard to influence people to be more female friendly.

2.5 Research Gaps

From the above paragraphs, following research gaps were found:

- Lack of technological intervention in Chinese gender language reform, let alone analyzing its effectiveness in influencing people's thought.
- Limited exploration of user experience in adapting to gender-neutral input methods.

This study aims to fill the gap by examining whether modifications to the " 女" radical in Chinese characters influence users' perception of themselves, aiming to improve the linguistic inclusiveness of the Chinese language.

3 Methodology

3.1 Phase 1: Perception Measurement via Questionnaire

Participants and Timeline 60–80 native Chinese speakers will complete a weekly online questionnaire for three weeks. Each questionnaire takes 5–10 minutes.

Questionnaire Each week, participants evaluate 24 to 26 self-descriptive sentences, each embedding one of four variation types. The way this questionnaire is arranged is inspired by Vainapel(2015)'s study.

- Original (O): the original form of a word.
- Method 1 (M1): for negative-origin words with female radical, replace the whole word with a gender-neutral synonym that does not indlude female radical (e.g., 嫉妒 [jealous] → 眼红 [really want something]).
- Method 2 (M2): for negative-origin words with female radical, the 女-radical is replaced with a neutral radical (e.g., 妨碍 [hinder] → 仿碍). It is worth mentioning that this word does not have naturally embeded meaning as it is a newly designed word based on original form.
- Method 3 (M3): for positive-origin words without female radical, a new character is created by adding a female radical to the original one, or replacing the original radical with a female radical (e.g., 真诚 [being true] → 真娀). It is also worth mentioning that this word does not have naturally embeded meaning as it is a newly designed word based on original form.

Sentence samples are randomized, and each linguistic item appears once across the three weeks, in only one variation. In other words, no participant saw multiple versions of the same word. Please refer to appendix for full wordlist.

Each sentence is rated based on the question: To what extent does this sentence describe you? Likert scale range from 1 to 10 (1 = "Totally not me", 10 = "Totally me").

Examples For example, the word 妨碍 (hinder, negative-origin with female radical "女") will appear in three different versions across the three surveys in the same expression which means *Sometimes, I hinder others from doing things* :

- In Week 1 as the original form. (O: 妨碍)
  有时，我会妨碍 [hinder] 到别人做事情。
  (Sometimes, I hinder others from doing things.)
- In Week 2 as a gender-neutral synonym. (M1: 阻碍)
  有时，我会阻碍 [block] 到别人做事情。
  (Sometimes, I block others from doing things.)

- In Week 3 as a radical-modified form. (M2: 仿碍)
  有时，我会仿碍到别人做事情。
  (Sometimes, I hinder others from doing things.)

Likewise, the word 创意 (creativity, positive-origin word without 女 radical) will appear in two different versions across the two of three surveys:

- In Week 1 as the original form (O)
  我很有创意 [havaing creativity]，能提出很多新的点子。
  (I am creative and can come up with many new ideas.)
- In Week 2 as a positively gendered creation (M3: 创嬉)
  我很有创嬉，能提出很多新的点子。
  (I am creative and can come up with many new ideas.)

3.2 Phase 2: Input Method Usability and Interview

3.2.1 Content

The interview will focus on the two most extreme types of participants: those whose self-evaluation greatly increased, and those whose self-evaluation greatly decreased. They will be interviewed about the following content.

1) **Cognitive understanding of modified characters**
   - Could you understand the meaning of the modified characters when filling in the questionnaire?
   - Did the modified characters feel confusing or hard to understand?
2) **Personal preference to different versions**
   - Which version do you think sounds more negative or insulting?
   - Can you give an example of how you want to use it in a sentence?
   - Does whether or not having the ”female radical” make the word feel more or less related to you personally?

- Do you think changing the radical changes the meaning of the word?
- What do you think of how the characters look? Do they feel strange, uncomfortable, or fun to you?

3) **Contextual use and willingness to express**

- In what situations would you use these words?
- Would you feel okay using these characters when chatting with friends, commenting online, or during an argument?

3.2.2 Tools and Technical Implementation

Several tools were created by the researcher to support the experiment:

- **Custom M3 characters:** Since M3 words use newly created Chinese characters with 女 radicals, the researcher designed these characters manually using Font Creator. The characters were exported as web fonts (.woff files) so they could be shown properly online.
- **To show customized character on Qualtrics questionnaire:** In order to embed self-designed character (M2 and M3) into the self-evaluation questionnaire, the researcher first stores the new font on cloud servers based in Hong Kong, so that participants from in-and outside China can successfully see it, then uses CSS to embed the font into the questionnaire.
- **Custom input method:** The input method used in Phase 2 was built using the Rime Input Method Framework. It suggests both original words and modified words, allowing users to choose between different versions. The complete code can be found at https://github.com/enenmia/feministekeyboard

3.3 Ethical Considerations

The study has been approved by the Media Technology MSc Ethics Committee at Leiden University. All participants are adults. They are fully informed and gave consent, they

also know they have the right to withdraw at any time. Questionnaire responses are anonymous, and interview data is pseudonymised and securely stored. No personally identifying information is collected. The study involves no foreseeable risk or discomfort.

4 Data Analysis

4.1 Data Description and Preparation

4.1.1 Two Main Tables

A total of 61 participants completed the three-week questionnaire. There were 29 Chinese words involved in the study, of which 14 had originally negative meanings (each with three forms: original, M1, and M2), and 15 had originally positive meanings (each with two forms: original and M3).

That is, all 61 participants rated the same statements containing different forms of the 29 words, giving a score reflecting "to what extent does this statement fit me." After data cleaning, we obtained two tidy long-format tables for data analysis: one for scores and related data for words with originally negative meanings, the other for those with positive meanings.

id	**week**	**question_key**	**answer**	**meaning**	**method**	**sentiment**	**gender**
1A	1	Vulgar/M2/negative	10.0	Vulgar	M2	negative	female
22z	1	Vulgar/M2/negative	6.0	Vulgar	M2	negative	female
04f	1	Vulgar/M2/negative	8.0	Vulgar	M2	negative	female
...	...	...	...	...	...	...	...

Table 1: First rows of responses for Vulgar/M2/negative condition

id	week	question_key	answer	meaning	method	sentiment	gender
1A	1	True/o/positive	4.0	True	o	positive	female
22z	1	True/o/positive	8.0	True	o	positive	female
04f	1	True/o/positive	6.0	True	o	positive	female
...	...	...	...	...	...	...	...

Table 2: First rows of responses for True/o/positive condition

We will analyze these two tables separately to answer: For the same person and same meaning, do different methods produce different scores? This addresses the main research question(RQ1), and will prove if hypothesis 1 (that M1, M2, M3 will influence people significantly in a positive way compared to their corresponding original from) holds or not.

4.1.2 Feminist Score Table

To investigate whether hypothesis 2 for Research Question 1 (that people who have stronger feminist beliefs will be more influenced by the modified characters) holds, the study needs scientifically measure the feminist level of the participant, that is when Duncan and her colleagues' (2021) research comes into play. Their research introduced an original Feminist Consciousness Scale (FCS), which is an 8-item, two-factor instrument developed and validated by them. "Two-factor" means the scale measures both feminist identity (e.g., self-identifying as a feminist) and awareness of gender inequality (e.g., recognizing women's lack of power in society). The FCS has strong reliability and has been shown to work consistently across genders.

This study picked two representative questions covering these two key dimensions:

1) "I am a feminist."

2) "I believe that women in society do not yet have the power and influence they deserve."

participant_id	"I am a feminist."	"I believe that women in society do not yet have the power and influence they deserve."	feminist_score
1A	8.0	10.0	9.0
22z	7.0	10.0	8.5
04f	10.0	10.0	10.0
...	...	...	...

Table 3: First rows of feminist score table

Participants rated on how they agree with each feminist statement on a 10-point Likert scale (from 1 = strongly disagree to 10 = strongly agree). The final feminist score was calculated as the mean of both.

4.2 Model Selection

4.2.1 Why LMM

The model used is the Linear Mixed-Effects Model (will be called LMM afterwards).

It was chosen because the experimental data have a nested structure, which means each participant provided scores under multiple meanings (semantics/words) and multiple methods. Different individuals may have subjective differences in their overall scoring tendencies, and different meanings may inherently be more or less likely to receive high scores. If only paired t-tests or simple ANOVA are used, it is difficult to control for these "person" and "meaning" confounders, which can be misattributed to the effect of the intervention method.

LMM allows us to examine the main effect of the method while modeling participant and meaning as random effects, absorbing their natural variation. This ensures that the method main effect (e.g., the difference between M1 and O) only reflects the real rating change due to method, within the same person and same meaning. The model's significance and effect size thus have higher internal validity and interpretability, making it one of the most suitable

methods for multi-level designs.

4.2.2 The Selection of random Effects

Although each participant completed the questionnaire across three consecutive weeks, with different methods appearing in different weeks, the design ensured that the pairing of meaning and method was completely consistent for all participants (for example, for meaning1, the "O" form always appeared in week 1 for everyone, "M1" in week 2, and "M2" in week 3; for meaning2, "M1" might appear in week 1, "M2" in week 2, and "O" in week 3, and so on). As the effect of week is fully absorbed by meaning, it does not need to be modeled separately. This avoids overfitting or confusion in interpretation.

4.3 Results

4.3.1 For Originally Negative Words (With Three Forms: Original, M1, M2)

Do different methods significantly affect scores?

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	answer
No. Observations:	2562	Method:	REML
No. Groups:	61	Scale:	2.0394
Min. group size:	42	Log-Likelihood:	-5364.6601
Max. group size:	42	Converged:	Yes
Mean group size:	42.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	5.288	0.136	38.853	0.000	5.021	5.555
method[T.m1]	0.349	0.069	5.049	0.000	0.213	0.484
method[T.m2]	0.025	0.069	0.356	0.722	-0.111	0.160
Group Var	0.734	0.134				
meaning Var	3.498	0.178				

Figure 1: LMM regression results for originally negative words. The model includes *answer* as the dependent variable, with *method* as the fixed effect, and *participant_id* and *meaning* as random effects.

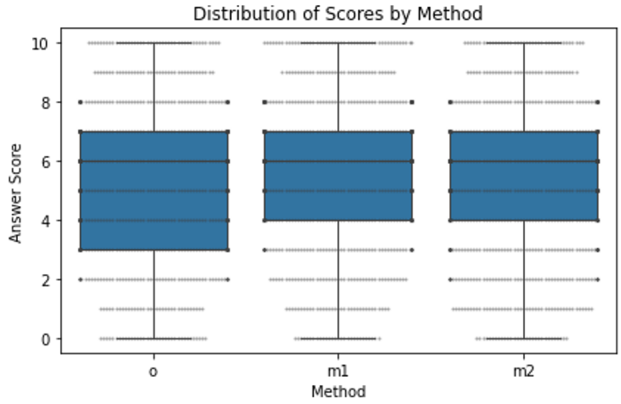

Figure 2: Distribution of scores by method for originally negative words

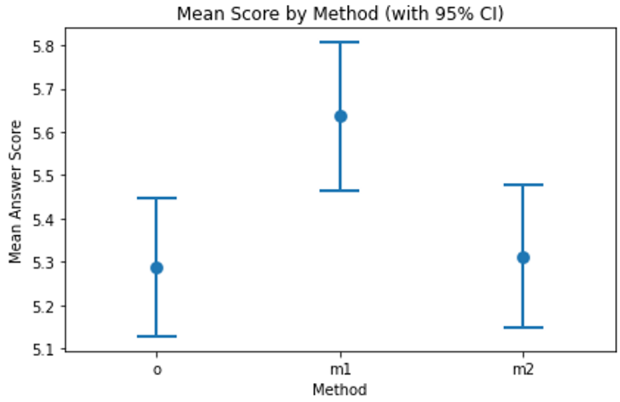

Figure 3: Mean score by method for originally negative words

The mixed-effects model analysis (see Figure 1) shows that, compared to group O, the M1 group's ratings significantly increased (estimate = 0.349, SE = 0.069, $p < 0.001$), while the M2 group's scores did not differ significantly from group O (estimate = 0.025, p = 0.722). Additionally, the model controlled for the random effects of participant and meaning, with variances of 0.734 and 3.498 respectively, indicating substantial variation between individuals and between meanings. This indicates that the M1 method systematically increases the rating, while the M2 method has no significant effect compared to the original expression (O).

It should be noted that in the negative meaning condition of this study, higher scores represent a more negative self-evaluation (i.e., the participant is more likely to agree "I have this negative trait"). Therefore, the increase in score caused by the M1 method actually reflects a greater tendency toward negative self-evaluation. This means that the M1 method

does not "weaken" the negative trait as assumed, but instead strengthens participants' echo with them. The M2 method does not show a significant difference from its original form, indicating that it does not significantly reduce negative self-evaluation. But at least, it also does not intensify it the way M1 method does.

Does participant gender influence the study?

Next, to assess whether the effect of intervention methods differs by gender, we included gender and its interaction with method in a mixed linear model.

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	answer
No. Observations:	2562	Method:	REML
No. Groups:	61	Scale:	2.0372
Min. group size:	42	Log-Likelihood:	-5364.2366
Max. group size:	42	Converged:	Yes
Mean group size:	42.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	5.283	0.147	35.906	0.000	4.995	5.571
C(method)[T.m1]	0.384	0.074	5.183	0.000	0.239	0.529
C(method)[T.m2]	0.008	0.074	0.109	0.913	-0.137	0.153
C(gender)[T.男]	0.038	0.406	0.095	0.925	-0.758	0.835
C(method)[T.m1]:C(gender)[T.男]	-0.268	0.205	-1.310	0.190	-0.669	0.133
C(method)[T.m2]:C(gender)[T.男]	0.126	0.205	0.615	0.539	-0.275	0.527
Group Var	0.752	0.137				
meaning Var	3.498	0.178				

Figure 4: LMM with Gender–Method Interaction Predicting Self-Evaluation Scores For M1&M2

In this mixed linear model, gender was included as a fixed effect (with female as the reference group), and method (M1/M2) was modeled as a fixed factor, along with interaction terms between method and gender. Meaning and participant ID were treated as random effects to account for variation across words and individuals. Compared to separate models for different gender groups, this approach allows for a more integrated interpretation of gender-method interactions.

According to the results, for female participants, M1 can significantly increase negative self-evaluation, while M2 has no significant effect; for other genders, neither M1 nor M2 differ significantly from O, which means that neither method has a systematic effect on self-evaluation for non-females.

In other words, female participants are more likely to be influenced by M1 (synonym-replaced) expressions, in the direction that increase their self-alignment with the negative

description.

Is the intervention effect related to meaning?

In order to examine whether the effectiveness of the interventions (M1: synonym-replacement; M2: radical-modification) varies depending on the semantic meaning of the word, paired t-tests were conducted separately for each word meaning. These tests compared participants' scores between the original form and the modified versions. Since multiple comparisons increase the risk of false positives, a Bonferroni correction was applied to adjust the significance threshold and control for the multiple comparisons problem.

This approach allowed us to identify whether the interventions produced consistent effects across all meanings or if they were only effective for certain word types. Specifically, we assessed whether either method significantly increased or decreased participants' self-association scores relative to the original version.

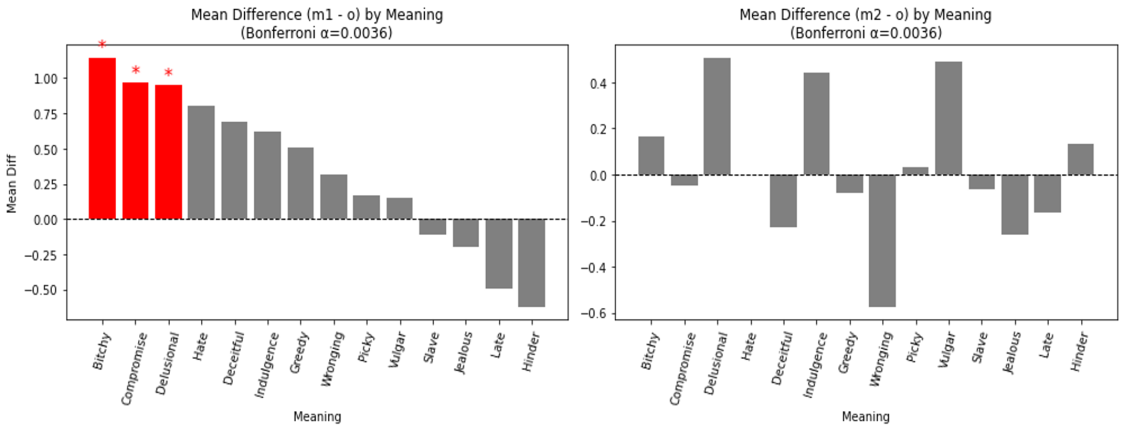

Figure 5: Meaning-Level Differences in Self-Evaluation Between Modified and Original Expressions (M1−O on the left and M2−O on the right)

In the results visualization (see Figure 5), colored bars represent meanings where the intervention effect was statistically significant, while grey bars indicate non-significant differences. The result shows that for M1, only three meanings (bitchy, compromise, delusional) have significantly different scores from its original forms, but in a way that decrease creases people's self-confidence, which counters the hypothesis. There was no significant difference found between M2 and O.

Is the effect related to the degree of feminism?

Here is how the study analyze the degree of feminism: for each participant, the mean of M1-o score differences across all meanings was calculated (higher M1, bigger difference = stronger M1 effect). Then the study uses feminism score to predict this effect size.

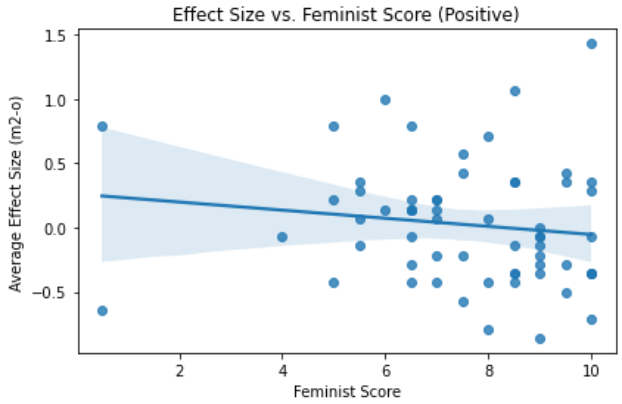

Figure 6: Relationship Between Feminist Score and the Effect of M1

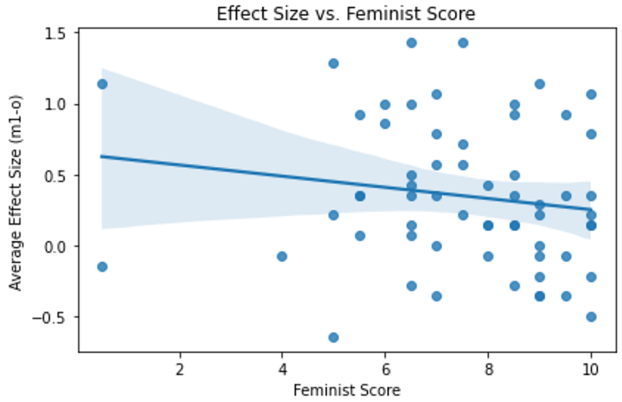

Figure 7: Relationship Between Feminist Scores and the Effect of M2

Correlation and regression analysis show no significant linear relationship between feminism score and the M1-o or M2-o differences (i.e., the degree of influence by synonym-replacement or radical-modification) in negative meanings. In other words, participants with higher feminism scores are not necessarily less susceptible to negative priming by modifications in this experiment; the trend is negative but not statistically significant.

4.3.2 For Originally Positive Words (Two Forms: original and M3)

Do different methods significantly affect scores?

```
         Mixed Linear Model Regression Results
============================================================
Model:             MixedLM Dependent Variable: answer
No. Observations:  1860    Method:             REML
No. Groups:        64      Scale:              1.5051
Min. group size:   10      Log-Likelihood:     -3618.2968
Max. group size:   30      Converged:          Yes
Mean group size:   29.1
------------------------------------------------------------
               Coef. Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------
Intercept      6.474    0.148 43.864 0.000  6.185  6.763
method[T.m3]   0.023    0.057  0.397 0.691 -0.089  0.134
Group Var      1.173    0.202
meaning Var    1.599    0.128
============================================================
```

Figure 8: LMM regression results for originally positive words

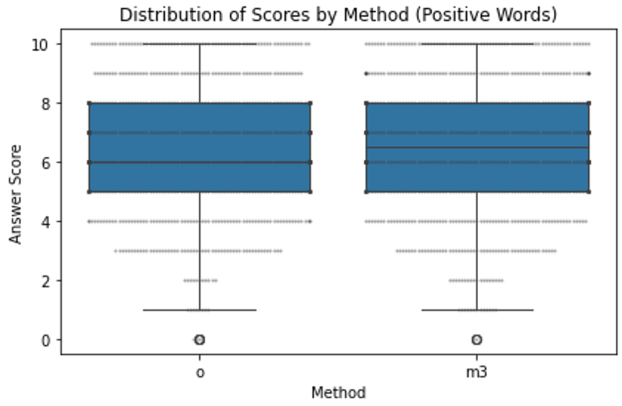

Figure 9: Distribution of scores by method for originally positive words

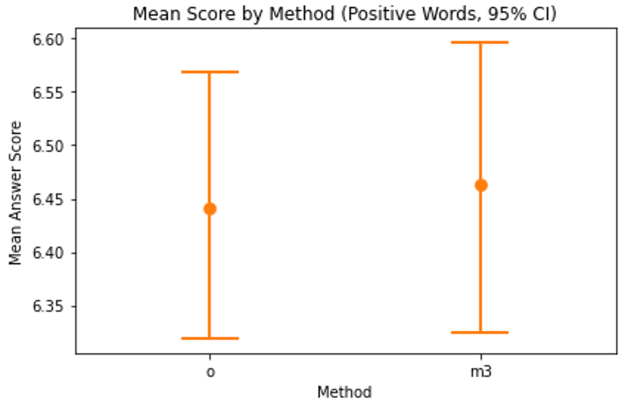

Figure 10: Mean score by method for originally positive words

The M3 method has no significant effect compared to the original expression (O). But at least, M3 does not increase negative self-evaluation.

Does participant gender influence the study?

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	answer
No. Observations:	1860	Method:	REML
No. Groups:	64	Scale:	1.5064
Min. group size:	10	Log-Likelihood:	-3618.9309
Max. group size:	30	Converged:	Yes
Mean group size:	29.1		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	6.484	0.159	40.756	0.000	6.172	6.796
C(method)[T.m3]	0.032	0.061	0.526	0.599	-0.087	0.152
C(gender)[T.男]	-0.075	0.448	-0.168	0.866	-0.953	0.802
C(method)[T.m3]:C(gender)[T.男]	-0.074	0.170	-0.434	0.664	-0.407	0.259
Group Var	1.194	0.206				
meaning Var	1.598	0.128				

Figure 11: Mixed Linear Model with Gender–Method Interaction Predicting Self-Evaluation Scores M3

Even when analyzed by gender, there is still no significant difference.

Is the intervention effect related to meaning?

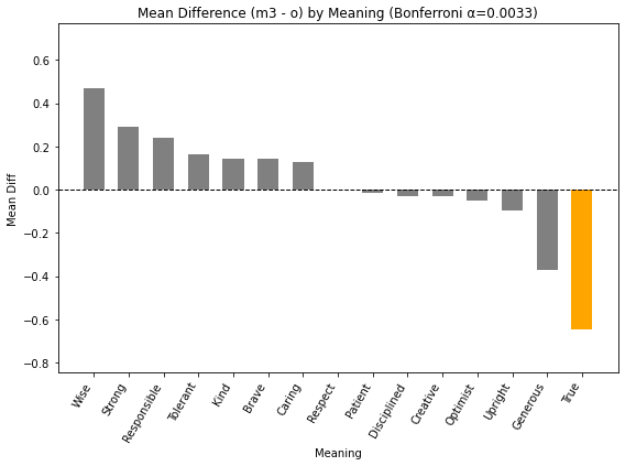

Figure 12: Word-Level Differences in Self-Evaluation Between Modified and Original Expressions (M3-O)

Paired t-test results show that for all positive meanings, the vast majority of words show no significant difference in self-evaluation between the M3 and o methods (all $p > 0.05$) except for word that means *True*.

True's M3 scores are significantly lower than its original form (mean difference = -0.65, $p = 0.0028$), suggesting M3 may trigger negative effects for this specific meaning.

Is the effect related to the degree of feminism?

Figure 13: Relationship Between Feminist Scores and the Effect of M3

The degree of feminism is also not linearly related to the degree of effect.

5 Interview Findings

5.1 Interviewee Selection

As mentioned in Methodology, in the interview phase, we focused on the two most extreme types of participants: those whose self-evaluation greatly increased, and those whose self-evaluation greatly decreased.

Self-evaluation increase type (high responders): Participants whose self-evaluation significantly improved under all interventions. Specifically, for negative words (M1, M2), scores significantly decreased (M1-O < 0, M2-O < 0, indicating weakened identification with negative traits), and for positive words (M3), scores significantly increased (M3-O > 0, indicating strengthened identification with positive traits). These participants best fit this study's hypothesis (most 'high' first): ['50y', '34n', '60q', '22x']

Self-evaluation decrease type (anti responders): Participants whose self-evaluation significantly declined under all interventions: for negative words, scores increased (M1-O $>$ 0, M2-O > 0, indicating strengthened identification with negative traits), and for positive words, scores decreased (M3-O < 0, indicating weakened identification with positive traits). These participants most contradict the study's hypothesis (most 'anti' first): ['24v', '41e', '59f', '36a', '23e', '29g', '61z', '05i', '17f', '52g', '61t']

The researcher contacted them starting from the beginning of the list. Eventually, the participants who took part in the interview were those with the IDs 50y, 34n, and 23e.

5.2 Insights from Interviews

Interviewees provided many insightful ideas that can help improve the creative output of this research project: a feminist Chinese input method (e-keyboard).

Based on their sharing of how they would use it in real life, the most important point would be that the E-keyboard should offer multiple modification choices for users to choose from, as the same user may prefer using different levels of modification in different circumstances. For example, when a female is chatting with her close friends who all hold feminist standpoints, she might prefer a softer or more neutral version (like M1 or M2). When she is sending

comments online under a post that is not relevant to feminist ideas, she might prefer using the original form, as she doesn't want to be perceived as having typos. If she is arguing with men online about gender issues and want to express more aggression, she would prefer using characters with male-related radicals (instead of gender-neutral modifications). These are not included in the M1/M2/M3 types discussed in this paper and could be considered a new type, M4, as they are even more progressive.

Different people also have different attitudes toward the same expression. For example, some think "矫揉造作" (with airs and graces) is better than "婊里婊气" (bitchy) as they think it decreased the part that discriminate women, while others prefer "婊里婊气" (bitchy), as they think it can be seen as a compliment in modern contexts, used to praise someone for being bold or having attitude.

Moreover, even participants who described themselves as more moderate feminists said that they wouldn't feel uncomfortable seeing others use the newly modified characters. They themselves might not use them every day, but they found the idea playful and creative.

For those who feel more connected to feminist ideas, the E-keyboard gives them a way to speak with more attitude. The shape of the character, together with the message, becomes part of how they express what they think or feel. It's not just about changing the words but about using writing itself to show who they are.

6 Discussion

6.1 Answers to RQs

According to the data and discussion above, we can already give out a clear answer to our research questions and hypotheses.

Hypothesis 1 for Research Question 1 is not supported. This means that in general, M1/M2/M3 does not influence people positively compared to original forms. More specifically, M1 influence people in the opposite direction (negatively); all M2 modifications do not influence people significantly; most M3 modifications do not influence people as well except for few

meaning.

Hypothesis 2 for Research Question 1 is not supported, which means people with stronger feminist belief will not be influenced by modifications more strongly. However, if viewed by gender as describe in Finding 4, the result can sometimes show meaningful differences. This suggest that gender may affect responses more than feminist beliefs.

The answer for Research Question 2 is that users change their preference on modifications. First, the baseline of acceptancy varies among individuals. Some prefer progressive modifications, some prefer the more moderate ones, or even the original ones. Also, the same person can change her choices on modified characters under different circumstances (elements include whether it is anonymous or not/whether she is in a female-friendly environment that makes her feel safe). In general, no matter the participant strongly identifies with feminism or not, most people are willing to and has the incentives to try out the newly designed input method. And even those with more moderate feminist views don't feel uncomfortable when they saw others using more progressive modifications.

We will dive into the more specific findings and the reason behind it in the following section.

6.2 Findings and Reasoning

Based on the data, the following findings can be concluded. The reason behind them can also be discussed based on conducted interview and literatures.

6.2.1 Finding 1: Synonym replacements (M1) increase negative self-evaluation.

For negative traits, using synonym-replaced words (M1) leads to higher negative self-evaluation compared to the original. These pairs have an M1 form which scores significantly higher than O:

Pair	O	Translation of O	Valence	Intensity	M1	Translation of M1	Valence	Intensity
1	婊里婊气	Bitchy	2	9	矫揉造作	Airs and graces/affected	2	5↓
2	妥协	Compromise/ trade-off	/	/	退让	Give in / yield	/	/
3	痴心妄想	Delusional / wishful thinking	2	9	胡思乱想	Let one's imagination go wild	2	1↓

Table 4: *Meaning of the O and M1 with significant difference, along with their valence and intensity scores.*

These significant differences may be due to the embedded differences in word intensity and valence, according to Database Chinese Emotional Lexicon Ontology (Xu et al., 2008), hereafter CELO database. The database is based on an extended version of Ekman(1999)'s six basic emotions. Each meaning has two dimensions: intensity and valence. Intensity level ranks from 1 to 9, with 9 being the most intensive. Valence has three categories of 0, 1, 2 and 3, which separately represent neutral, negative, positive and both.

For Bitchy The word "婊里婊气" is not explicitly included in the CELO database, but its strong variation "婊子" (bitch, the noun form of bitchy) has an intensity of 9 and a negative valence of 2, whereas "矫揉造作" (with airs and graces, M1) has an intensity of 5 and the same negative valence of 2. This means that the M1 version is inherently less negative. Therefore, it is expected that participants would rate the M1 word higher than the original, assuming their self-perception remains at the same level.

For Delusional The original word "痴心妄想" (wishful thinking) has an intensity of 9 and a negative valence of 2, while the M1 replacement "胡思乱想" (wild imagination) has an intensity of only 1 with the same valence of 2. Again, this indicates that in this pair, M1 is less intense and less pejorative, which may explain why M1's self-evaluation scores are higher.

For Compromise The original term "妥协" (compromise) and its M1 replacement "退让" (concession) did not appear in the database, and interviewees also have different opinions on which one is more 'negative'. Therefore, we cannot really determine whether the score difference is due to the difference in word meaning or not.

6.2.2 Finding 2: Following Finding 1, the study also found that the negative impact of synonym replacements (M1) is significant among female participants, but not for other genders.

This can be because female participants are more likely to relate to the original negative words that contain the "female radical," viewing them as personally relevant.

Previous neuroimaging research has shown that individuals belonging to marginalized groups will show stronger emotional and self-referential neural responses when exposed to derogatory labels targeting their identity (Naranowicz and Jankowiak, 2025). In contrast, neutral or reclaimed alternatives triggered lower amygdala activity, which suggested reduced emotional reactivity (Naranowicz and Jankowiak, 2025). Also, this gendered sensitivity may stem from the historically embedded connotations of the female radical in Chinese characters, which has been associated with negative traits and social subordination for a long time (Xie, 2018; Zhao, 2003).

In contrast, male participants typically do not associate themselves with those characters and therefore show less reaction. This gender-specific impact of M1 speaks with the existing literature that language both reflects and constructs gender identity (De Francisco, 1992; Cameron, 2014).

6.2.3 Finding 3: In general, radical modified forms (M2) have no significant influence compared to original form, neither in positive direction nor in negative direction

Based on the interviews, several possible reasons were identified:

Meanings were still recognizable

Many participants mentioned that although M2 characters looked unfamiliar at first, they could still quickly understand their meaning, and this did not affect their judgment or scores. For example:

- "It wasn't to the extent that I couldn't recognize the character."
- "Once I understood the meaning, it didn't affect how I judged it."

- "I was focused on answering the questions and didn't notice it was a newly created character."

This finding aligns with Bybee's (2010) research, where he suggested that cognitive efficiency is an important element for language processing. As participants were still able to recognize the M2 forms without too much effort, the modifications did not disrupt comprehension enough to change how participants evaluated them.

Participants had different responses to the removal of the female radical

Some participants clearly felt that switching to a gender-neutral version reduced the negative tone of the word, which led them to give higher self-relevance scores. Others focused more on the visual connection between the character and their own identity. For them, removing the female radical weakened the sense of personal relevance, and their scores went down as a result. As one participant put it:

- "Neutral radicals didn't give me that strong feeling of relevance like the '女' radical sometimes does."

Meanwhile, not all participants felt bothered by the presence of the female radical in the original negative words. One female participant explained:

- "I don't think the female radical is derogatory in some words···for example 嫌弃 (hateful), it doesn't feel like an insult to women for me."

Low emotional involvement in the task

Because the study used questionnaire as the carrier, most participants treated it as a task rather than a space for expressing opinions. They weren't particularly engaged with how the words looked:

- "I was focused on the answers, not the characters."
- "If I were on Xiaohongshu commenting under a feminist post, I might care more—but not in this kind of survey."

Some mentioned that in more emotionally-heavy situations such as arguing online, they might prefer stronger or more expressive modifications than M3, even using male radicals for emphasis:

- "If I'm arguing with a man and I really want to say something bad on his toxic masculinity, a gender-neutral word doesn't feel aggressive enough."

This supports Fogg's (2002) framework of persuasive technology, where he emphasizes that digital interfaces are heavily context-based. The same intervention (in this study: a modified character) may trigger very different reactions depending on whether it appears in a reflective survey or a public, expressive platform (Scharff, 2017).

Participants made sense of the new characters

Not everyone saw the M2 change as gender-related. Some thought it was just a style or cultural variation:

- "I didn't think too much about it. At first I thought it was a translated word or a different writing system."
- "I thought it was some expression you might see in a novel from Taiwan."

So even when participants noticed something different about the characters, they were able to accept it or explain it away, meaning it didn't affect how they felt or how they scored the word. This tendency of accepting a new language from is a well-documented phenomenon. Ferreira and her colleagues (2002) suggest that people will not process everything in language in a precise way. In language comprehension, a way of 'good enough' processing is adopted to maintain fluency in comprehension (Ferreira et al., 2002). Also, Clark and Marshall (1981) found that people turn to previous knowledge and cultural background to interpret unfamiliar expressions.

6.2.4 Finding 4: For positive traits, modification (M3) generally has no effect, except in rare cases like "True"

The lack of significant difference between M3 and O can be explained by the diversity in participants' perceptions. According to the interview, some participants said that if a word

already matched their self-identity, adding a gendered radical didn't make it feel more or less relatable —their rating stayed based on the meaning, not the visual form. One interviewee explained that after the female radical was added, the word 真诚 (Being true) felt "less true", and estimated that its positive tone had dropped to "about 80% or 90% of the original form". She added that their self-evaluation didn't change because of the word's meaning, but rather adjusted to match how the word now felt, hence, a lower score made sense to her in this case.

6.2.5 Finding 5: Participants' feminism score does not predict their response to language intervention.

Although Hypothesis 2 expected that participants with stronger feminist beliefs would be more influenced by the language changes, the results did not support this. A reason may be that sometimes there is a gap between what people believe and how they feel in the concurrent moment. Even if someone identifies as a feminist, their reactions to words often happen quickly and automatically, based on what they are used to seeing or hearing.

This argument is supported by Pickering and Garrod's (2004) study where they find that people don't always deeply process every word, they often respond based on habits and past experience. Bybee (2010) also introduced that the more often people see or use a word, the more natural it feels. Therefore, some people are more comfortable with the original forms than with modified forms, simply because they are more familiar with the original look. Even if they agree with the idea behind the change.

Also, as Scharff (2017) explains, many people express feminism differently depending on the situation. In a public or emotional setting, they may tend to speak more strongly. But in a quiet, neutral setting like a survey, they may choose not to react much. That could potentially explain the reason of this finding that even those participants with high feminism scores did not show stronger reactions to the changes: they just didn't feel like this was the place to do so.

6.3 Linking Back to Linguistic Relativity

My study can partially support and extend the theory of linguistic relativity.

Firstly, language structure can indeed influence individuals' self-perception, especially under the topic of gender. According to Vainapel and her colleagues' (2015) research which is an extension of linguistic relativity, the language's manipulative effect on self-description exerts mainly on women. Also, when exposed to gender-inclusive language, the study found that women are more influenced by men, who were not significantly affected. Finding 2 of this study which suggests that female is significantly influenced and male is not, clearly speaks with the finding of Vainapel's (2015) research.

Secondly, the findings agree with and also extend Boroditsky's (2003) finding that the way language categorizes the world can influence people's focus when thinking (i.e. attentional bias). If speakers of language without grammatical gender learn a new experimental language with grammatical gender, their attention change and they start to notice gender-related aspect of the language (Wolff and Holmes, 2011).

This idea can be extended to show that how exactly people's focus on language reflects back on themselves depends on their unique interpretation of the linguistic component. If we raise a more specific example: experimental characters like "真娀 (M3)" is semi familiar, semi novel to the participants. It belongs to a gray area where interpretation can be done subjectively, and this makes the newly designed characters especially sensitive to the reader's internalized associations with components like the "女" radical. If they view female positively, then they would interpret the new form positively; if they view female with inherent negative feeling, then the interpretation of the new form would be negative. This also explains why the difference between M2 and M3 with their original form is generally not significant.

Thirdly, Vainapel's (2015) experiment proves that when the statement includes women by using 'she and he' as reference rather than just using 'he', women participants feel more motivated and gain more self-efficacy. My study takes the question one step further by asking: what happens if this inclusion is not neutral as pronoun, but related to the positive or negative meaning of a word? The finding suggests that the effect of such inclusion is dependent on the tone of the word itself (as stated when explaining the result of M1) and

how each person interprets the female radical.

6.4 Limitations

There are three main limitations of this research. First, although there are 61 participants that have participated in phase 1, most of them are female. Though the study does plan to focus more on the reaction from female users, the gender ratio can still be improved to make the data more valid. Second, most participants happen to own a degree equals to or higher than bachelor, which makes the sample not comprehensive enough to reflect on various potential user groups. Third, the carrier of this study is questionnaire, which is relatively neutral and is not the best place to evoke emotional and social reactions as if in real life.

6.5 Suggestions for Future Work

For improving the limitations and for pointing out some interesting fields to be explored based on this study, there are following suggestions.

Firstly, a more comprehensive study among more participants with various backgrounds and better gender ratio can be conducted.

Secondly, more interaction scenario that goes beyond a passive setting (questionnaire) can be conducted, and it would be interesting to see which can best empower female. Settings such as using modified characters on social media, in schools, in social activities such as Women's march can be interesting.

Thirdly, the feminist e-keyboard can be improved before it is really put into market. As it is based on an open-source Chinese input method initiated by male, it still includes many terms that are not female-friendly. A cleaning up of the sexist terms will be necessary. Also, this input method can be used as a investigation tool for future study, to see if using and working with female-friendly characters can have long term cognition effects on users.

Lastly, it would be interesting to include a wearable Electroencephalography(EEG) headset

to objectively detect people's emotional response. It may provides more detailed reasoning for some findings of this study.

7 Conclusion

This study explores whether changing gendered radicals in Chinese character can influence how people see themselves. Synonym replacement (M1) based on word unit and Radical neutralization based on character unit (M2) are the modifications for originally negative expressions. Adding female radical (M3) is how originally positive expressions are modified. The study tested how participant evaluate themselves differently under modified and original forms. Main findings include: 1) M1 increase negative self-evaluation, and this negative impact of M1 is significant among female participants, but not for other genders; 2) the impact of M2 and M3 varied depending on the word's meaning; 3) participants' feminism level cannot predict their response to language intervention.

In general, the modifications do not have a consistent positive impact on self-perception. This result may be because 1) the questionnaire feels like a neutral and task-based setting and does not invite strong emotional reaction; 2) participant can make sense out of the modified characters because of human's embedded comprehension mechanism, and interpret them based on their original form; 3) there is a gap between what people believe and how they feel in the moment. Even if someone identifies as a feminist, their reactions to words can happen quickly and automatically based on what they are used to seeing or hearing.

The study also triggers a creative output, which is a user-friendly feminist Chinese input method that provides various typing options including original characters and modified characters.

This is the first study to bring feminist language reform and digital tool design together into the Chinese context. It is hoped that the result of this study can provide a unique perspective of how feminist ideas can affect people's self-perception, and to show how technology can help spread these ideas. The created input method can also serve as a tool for future researchers or developers who believe in feminism, tech for good, or inclusive language can build on. In doing so, it hopes to make feminist language more visible and more usable

in digital spaces.

References

Arnold, K., Chauncey, K., and Gajos, K. (2018). Sentiment bias in predictive text recommendations results in biased writing. In *Proceedings of Graphics Interface 2018*, GI 2018, pages 33 – 40. Canadian Human-Computer Communications Society / Societe canadienne du dialogue humain-machine.

Bailenson, J. N. and Yee, N. (2005). Digital chameleons. *Psychological Science*, 16(10):814–819.

Benjamin, R. (2024). *Imagination: a manifesto*. W. W. Norton Company.

Bhat, A., Agashe, S., and Joshi, A. (2021). How do people interact with biased text prediction models while writing? In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing (HCINLP 2021)*, pages 116–121, Online. Association for Computational Linguistics.

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). Sex, syntax, and semantics. In *The MIT Press eBooks*, pages 61–80.

Bybee, J. (2010). *Language, usage and cognition*.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Cameron, D. (2014). Gender and language ideologies. In *The Handbook of Language, Gender, and Sexuality*, pages 279–296.

Chin, N. B. and Burridge, K. (1993). The female radical: Portrayal of women in the chinese script. *Australian Review of Applied Linguistics. Series S*, 10:54–85.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I. A., editors, *Elements of discourse understanding*, pages 10–63. Cambridge University Press.

DayaZoom, X. U. (2024). 为什么脏话里的含妈量这么高啊? [why do so many swear words involve people's moms] - 小红书 [xiaohongshu].

De Francisco, V. L. (1992). Deborah tannen, you just don't understand: Women and men in conversation. *Language in Society*, 21(2):319–324.

Desprez-Bouanchaud, A., Doolaege, J., Ruprecht, L., and Unesco (1999). Guidelines on gender-neutral language.

Duncan, L. E., Garcia, R. L., and Teitelman, I. (2021). Assessing politicized gender identity: Validating the feminist consciousness scale for men and women. *The Journal of Social Psychology*, 161(5):570–592.

Ekman, P. (1999). Basic emotions. In Dalgleish, T. and Power, M. J., editors, *Handbook of Cognition and Emotion*, pages 45–60. John Wiley & Sons Ltd.

Ferreira, F., Bailey, K. G., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1):11–15.

Fogg, B. J. (2002). Persuasive technology. *Ubiquity*, (December):2.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, 23(11–12):629–643.

Huang, X. (2023). A cultural history of the chinese character “ta (她, she)”: Invention and adoption of a new feminine pronoun. In *A Cultural History of the Chinese Character “Ta (她, She)”*, chapter 1. Routledge, 1 edition.

Keegan, T. T. and Evas, J. (2012). Nudge! normalizing the use of minority language ict interfaces. *AlterNative: An International Journal of Indigenous Peoples*, 8(1):42–52.

Lakoff, R. (1973). Language and woman’s place. *Language in Society*, 2(1):45–80.

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

Ling, Y. (1989). “她”字的创造历史 [the history of the character ‘ta[she/her]’]. 语言教学与研究 *(Language Teaching and Research)*, pages 139–151.

Molina, M. F., Chary, A. N., Reyes, K., Ford, J. S., Chinnock, B., and Rodriguez, R. M. (2024). Demystifying hispanic versus latino/a versus latinx: Which do emergency department patients prefer? *Annals of Emergency Medicine*, 84(4):464–467.

Naranowicz, M. and Jankowiak, K. (2025). Positive mood enhances gender stereotype activation during semantic integration and re-analysis. *NeuroImage*, 310:121116.

Papadimoulis, D. and Parliament, E. (2018). Gender-neutral language in the european parliament.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3):779–787.

Pickering, M. J. and Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(2).

Pozniak, C., Corbeau, E., and Burnett, H. (2024). Contextual dilution in french gender inclusive writing: An experimental investigation. *Journal of French Language Studies*, 34(2):273–292.

Prewitt-Freilino, J. L., Caswell, T. A., and Laakso, E. K. (2011). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles*, 66(3–4):268–281.

Saguy, A. C. and Williams, J. A. (2021). A little word that means a lot: A reassessment of singular they in a new era of gender politics. *Gender & Society*, 36(1):5–31.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 5(4):207.

Scharff, C. (2017). *Gender, subjectivity, and cultural work*.

Sczesny, S., Formanowicz, M., and Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7.

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). Representation of the sexes in language. In Fiedler, K., editor, *Social Communication*, Frontiers of Social Psychology, pages 163–187. Psychology Press, New York.

Stinson, L. (2016). This little red book confronts sexism in the chinese language.

TaArt, X. U. (2025). 一本只有女字旁汉字的“字典”[a “dictionary” composed solely of chinese characters with the female radical] - 小红书 [xiaohongshu].

Vainapel, S., Shamir, O. Y., Tenenbaum, Y., and Gilam, G. (2015). The dark side of gendered language: The masculine-generic form as a cause for self-report bias. *Psychological Assessment*, 27(4):1513–1519.

Viennot, . (2017). Non, le masculin ne l’ emporte pas sur le féminin !

Wang, Q. (2016). 《说文解字》女部文化阐释研究综述 [a literature review on cultural interpretations of the “女” radical in shuowen jiezi].

Wang, Y., Mao, Y., and Deng, X. (2023). 三八新女性「她」艺术展 [her power art exihibition].

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press, Cambridge, MA. Originally written in the 1930s-40s, published posthumously.

Wolff, P. and Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):253–265.

Xie, Y. (2018). A brief analysis of derogatory chinese characters with the “女” radical: taking 奴, 嫉, 妒, 婪, and 嫌 as examples. *Modern Communication*, (3):73–74. CNKI:SUN:XKJJ.0.2018-03-043.

Xu, L., Lin, H., Pan, Y., Ren, H., and Chen, J. (2008). 情感词汇本体的构造 [database chinese emotional lexicon ontology]. 情报学报, 27(2):180–185.

Zhao, A. (2003). A glimpse of the change of female social status from words with radicals “女” in the origin of chinese character. *Journal of Xinyang Teachers College (Philosophy and Social Sciences Edition)*.

Zhao, J., Du, B., Zhu, S., and Liu, P. (2021). Construction of chinese sentence-level gender-unbiased data set and evaluation of gender bias in pre-training language model. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics (CCL 2021)*, pages 563–574, Huhhot, China. Chinese Information Processing Society of China.