

Beyond the Alphabet, Beyond the Binary, Maybe Beyond ‘Chinese’?

Digital Input Constraints on Gender-Inclusive Chinese Pronouns

Janet Ganon-Davey

Abstract. The emergence of gender-inclusive pronouns in Chinese reveals how globalised technological infrastructure for writing can create institutional barriers to orthographic innovation. Written Chinese has only distinguished gender in third-person pronouns for about a century, with 他 ‘he’, 她 ‘she’ and 它 ‘it’ all pronounced as *tā* in Mandarin. Recently, several new Chinese gender-inclusive pronouns have emerged. The emerging pronouns include a romanised form based on pinyin (‘TA’), as well as Chinese character forms (Sinographs) encoding nonbinary gender (e.g., 「无也」). In this study, I analyse data from two large Chinese-language corpora to investigate the frequency and use of these emerging gender-inclusive pronouns alongside binary pronouns 他 ‘he’ and 她 ‘she’. Results of the corpus analysis highlight how institutional technical requirements—particularly Unicode encoding standards and input method compatibility—shape which linguistic innovations can succeed. Proposed Sinograph pronouns remain virtually absent from mainstream discourse, as they cannot be typed using standard input systems. Meanwhile, ‘TA’ is increasingly accepted in mainstream discourse, with its utility as a gender-neutral pronoun facilitating its emergence as a gender-inclusive form. These findings reveal how the parameters of digital writing can constrain language change through requirements for global standardisation that neither local solutions nor social acceptance alone can overcome. While the introduction of the feminine pronoun 她 ‘she’ in the 1920s required navigation of social and practical linguistic constraints in China, contemporary Chinese pronoun innovation must also work within the parameters of global technological infrastructure governing Sinograph creation. This study thus illuminates broader questions about how globalised contemporary writing technologies shape the possibilities for script reform and innovation.

1. Introduction

In recent years, growing recognition of gender diversity has prompted greater attention to using gender-inclusive language, and in particular,

Janet Ganon-Davey  0000-0001-6607-393X

The Australian Centre on China in the World, China in the World Building #188, Fellows Lane, The Australian National University, Canberra 2601, Australia
E-mail: janet.davey@anu.edu.au

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2024. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 12.
Fluxus Editions, Brest, 2026, pp. 533–550. <https://doi.org/10.36824/2024-graf-dave>
ISBN: 978-2-487055-12-4, e-ISBN: 978-2-487055-13-1

gender-neutral or gender-inclusive pronouns (Bradley, 2020; Robertson and Thompson, 2022; Troncoso, Schudson, and Gelman, 2023).¹ While much attention has focused on English singular ‘they’ and novel pronouns (‘neopronouns’) like ‘ze/zir’ (Arnold, Mayo, and Dong, 2021; Bjorkman, 2017; Sheydaei, 2021), the development of gender-inclusive pronouns in Chinese illuminates a distinct challenge: the relationship between linguistic innovation and digital writing infrastructure. In Chinese, the creation of new Sinographs now requires multiple layers of centralised institutional approval and technical implementation before widespread use becomes possible. This globally standardised constraint has major implications for how gender-inclusive language evolves in Chinese.

The Chinese third-person pronoun system has only distinguished gender in writing for about a century, with 他 ‘he’, 她 ‘she’, and 它 ‘it’ all pronounced *tā* in Mandarin (Ling, 1989).² In the past decade, several new gender-inclusive pronouns have emerged to cater to nonbinary and other gender-diverse Chinese language users. These innovative pronouns take various forms: some are standalone Sinographs (e.g., 「𠂇无也」), some combine Sinographic components with letters (e.g., 她), while others like ‘TA’ draw on the existing *tā* pronunciation (Lu, 2021). However, the practical constraints of digital Chinese writing significantly impact which of these new pronouns can achieve widespread use.

In this paper, I analyse how emerging gender-inclusive pronouns are used in contemporary Chinese texts compared to established binary pronouns 他 ‘he’ and 她 ‘she’. Through corpus analysis of pronoun frequency and usage patterns, I demonstrate that practical considerations—particularly those driven by technological constraints—are shaping which emerging gender-inclusive pronouns can succeed. While Sinograph neopronouns remain virtually absent, the pinyin-inspired pronoun ‘TA’ is becoming established in mainstream discourse. The technical requirements of typing and displaying new Sinographs thus appear to be creating significant barriers to Chinese pronoun innovation. These findings, reinforced by a recent survey of pronoun use among queer Chinese speakers (Davey, 2024), suggest that technological

1. Throughout the paper, I use ‘gender-neutral’ to refer to pronouns or other words that do not differentiate binary gender and/or are used for both men and women; ‘gender-inclusive’ to refer to words that either explicitly denote or implicitly encompass nonbinary genders; and ‘ungendered’ for words that do not contain or imply gender information. In some cases, the same pronoun can be gender-neutral or gender-inclusive. For example, English singular *they* can be used to refer to someone who is nonbinary, or someone whose gender is unknown (but is assumed to be a man or a woman).

2. Two additional third-person pronouns are used in Taiwan, *tā* 牠 for animals and *tā* 祂 for gods and deities (Cheng, 2016).

constraints may be steering Chinese gender-inclusive pronoun development toward solutions that work within existing digital parameters.

2. Background

Chinese characters, or Sinographs, are composed of distinct components that combine in rule-governed ways to convey both semantic and phonetic information (Myers, 2019; Wu, 2022). These semantic components play a crucial role in encoding meaning distinctions in writing that may not be reflected in speech. For instance, in modern Chinese pronouns, while all third-person singular forms are pronounced as *tā* in Mandarin, their written forms use different semantic components to distinguish gender. The masculine pronoun 他 ‘he’ features the component 亻 ‘human’ on the left side, reflecting its historical use as an ungendered pronoun. The feminine pronoun 她 ‘she’ instead uses 女 ‘woman’ on the left side to explicitly mark feminine gender (Liu, 1995). Both pronouns have 也 on the right side, which indicates the *tā* pronunciation in Mandarin.

This gender-distinguished pronoun system is relatively new—written Chinese has only marked gender in pronouns for a hundred years (Hong, 2020). However, the introduction of 她 ‘she’ in the 1920s has created an asymmetric system where the ‘human’ component of 他 permits its continued use as a default or generic pronoun, while 她 is exclusively used for women. Many scholars (e.g., Cheng, 2016; Farris, 1988; Hodge and Louie, 1998; Jung-Palandri, 1991; Moser, 1997) argue the specificity of 她 perpetuates default male subjectivity in written Chinese. This argument is supported by psycholinguistic research showing that the semantic components of Sinographs play a crucial role in gender processing during pronoun resolution, with 他 functioning as a default form that can apply to all genders while 她 is processed as specifically feminine (Su et al., 2012). This asymmetric gender marking in written but not spoken Chinese pronouns has sparked efforts to develop more inclusive alternatives for contemporary language users (Koon, 2022).

Several solutions have emerged to create more inclusive pronouns in contemporary Chinese. One approach advocates returning to the historical usage of 他 as a truly ungendered pronoun, arguing that its ‘human’ component better reflects this purpose (Cheng, 2016). Others suggest adopting pronouns from Chinese languages that lack written gender distinction, such as 佢 from Cantonese or 伊 from Wu, Southern Min and Taiwanese (Lu, 2021). More recently, two innovative proposed Sinographs have emerged specifically to represent nonbinary gender identity. The Sinograph neopronoun 𠄎, created in 2015 by intersex community group ‘雙性人The Missing Gender 0.972’, preserves the *tā* pronunciation through 也 and uses 乂 as its left semantic component to reflect growing use of ‘X’ as a gender or sex maker (Quinan and Hunt, 2023;

Tang, Yi, and Bedin, 2022).³ 她 was formally encoded in Unicode 17.0 in September 2025 at codepoint U+323BF in the CJK Unified Ideographs Extension J block. Likewise preserving the *tā* pronunciation, another Sinograph neopronoun 「𠂔无也」, which was proposed by Ziping Zhu in 2021, features *wú* 无 ‘none’ as its left-side semantic component, explicitly rejecting binary gender categorisation (Papadopoulos and Zhu, 2022). Alongside these Sinograph neopronouns, the pinyin-inspired form ‘TA’ has gained traction in digital spaces (Ferrari, 2022). ‘TA’ follows a precedent of romanised loanwords being adopted in modern Chinese (e.g., PK ‘player kill—to take on, go head to head, showdown’; APP ‘smartphone app’; COS ‘cosplay’; Q ‘cute; pleasantly chewy’) while avoiding gender specification entirely.

However, these proposals face major technological hurdles. While the introduction of 她 ‘she’ in the early twentieth century met significant social resistance and required technical effort to implement through printed works, these challenges could at least be addressed through local solutions in individual print shops and publishing houses (Brokaw and Reed, 2010; Huang, 2009). The shift to digital writing has fundamentally altered how new Sinographs can be introduced. Unlike many alphabetic writing systems where pressing a sequence of keys on a keyboard immediately produces those exact symbols on screen—what Mullaney (2017, p. 317) calls a ‘what-you-type-is-what-you-get’ system—Chinese input requires converting user actions to encoded graphs in computer memory. This process relies on Input Method Editors (IMEs), “middleware” programs that mediate between user input and graph output (Mullaney, 2024, p. 5). To type in Chinese, users either enter phonetic ‘spelling’ using systems like Sogou pinyin, or build Sinographs by selecting character components or drawing strokes on a screen. The IME intercepts these inputs and continuously searches its database, presenting users with candidate graphs that match their keystrokes or gestures in a manner akin to predictive text (for a detailed discussion, see Mullaney, 2024). However, this means users can only type graphs that are already encoded in the system’s database—a constraint that has transformed localised technical barriers into requirements for global institutional approval.

Chinese input databases operate within the encoding frameworks of Unicode and ISO/IEC 10646. The Ideographic Research Group (IRG) reviews and coordinates the inclusion of proposed new graphs into the CJK (Chinese, Japanese, Korean) Unified Ideographs set and the Uni-

3. In Chinese, the term *xìngbié* 性别 encompasses both ‘sex’ and ‘gender’ as understood in English. While the United Nations gender-inclusive language guidelines promote *shèhuì xìngbié* 社会性别 ‘social gender’ and *shēnglǐ xìngbié* 生理性别 ‘sex’ to distinguish these concepts (United Nations, 2012), these terms are not widely used in everyday discourse.

versal Multiple-Octet Coded Character Set (USC, ISO/IEC 10646; Tsu, 2022). These centralised technological requirements create fundamental constraints on Chinese orthographic innovation. Since 「𠂇无也」 has no Unicode encoding, it cannot be rendered as a unified graph, and must be represented here via its component parts, with 𠂇 indicating the left-right arrangement. By contrast, 𠂇 has now completed this stage. Following an IRG submission from the UK in 2021 (UK-20538; UK, 2021), 𠂇 was formally encoded in Unicode 17.0 in September 2025 as U+323BF. As Figure 1 shows, 𠂇 is now able to be rendered as a single graph with the same visual unity as established Sinographs like 她.



FIG. 1. The Sinograph 𠂇 neopronoun, encoded as U+323BF in Unicode 17.0

Yet Unicode encoding alone does not guarantee adoption. Before a newly encoded Sinograph can be typed or displayed, it must be incorporated into IMEs, predictive input algorithms, and handwriting recognition systems, each of which requires independent institutional action at a global scale (Mullaney, 2024). As of mid-2026, 𠂇 remains absent from mainstream IMEs and the font sets bundled with major operating systems, meaning it still cannot be typed or reliably displayed by most users. These technological hurdles fundamentally shape how gender-inclusive language can develop in contemporary Chinese.

3. Methodology

To examine how emerging gender-inclusive pronouns are used in contemporary Chinese, I analysed data from two large-scale Chinese language corpora: Sketch Engine’s Chinese Web Corpus 2017 (zhTenTen17; Sketch Engine, 2020) and WuDaoCorpora (Yuan et al., 2021). zhTenTen17 provided the primary dataset, containing 16.6 billion tokens (13.5 billion words) of internet texts collected in 2016. To verify these findings against more recent language use, I also examined a 15 GB subset (91+ million tokens) of WuDaoCorpora, which consists of internet texts collected in 2022 by the Beijing Academy of Artificial Intelligence for AI model training.

My analysis proceeded in two stages. First, I examined Key Word In Context (KWIC) results to determine how frequently each pronoun appeared in the corpora. KWIC analysis displays the target word with its surrounding context, allowing examination of how words are used in

natural language data. Using Sketch Engine's online interface, I conducted case-insensitive searches for each pronoun (including variants like 'ta' and 'Ta' for 'TA'). For comparative analysis of each pronoun, I extracted randomly-sampled datasets of 10,000 concordance lines each for 他 'he' and 她 'she', and all 9,366 concordance lines for 'TA'. Since the frequency results showed no instances of 「无也」 or 她 functioning as pronouns, I excluded these from further analysis.

Second, I conducted collocation analysis to understand how 'TA' is being used in online discourse relative to binary pronouns 他 'he' and 她 'she'. Collocation analysis reveals which words ('collocates') frequently appear near a target word (the 'node'). However, simple frequency counts can be misleading since some words are inherently more common in a language. I therefore used two complementary statistical measures to identify meaningful collocates within three words of each pronoun:

- Log-likelihood scores: identify typical collocates by measuring how significant the co-occurrence is, accounting for overall word frequencies in the corpus
- Mutual information (MI) scores: measure strength of association by calculating how much more likely words are to appear together than separately

Using both measures allowed me to consider not just which words typically appear near pronouns, but also which word combinations were distinctive (Desagulier, 2017). Since both statistical measures are affected by corpus size, using equivalent-sized concordance datasets was crucial for valid comparison (Pojanapunya and Todd, 2016). For this analysis, I used AntConc (Anthony, 2022a) and RStudio (Posit Team, 2023) with R v.4.2.1 (R Core Team, 2022). Because Chinese text lacks marked word boundaries, all data was first processed using TagAnt (Anthony, 2022b) for word segmentation, allowing me to identify meaningful word collocates rather than just adjacent Sinographs or punctuation marks. The processed data was then analysed in two complementary stages: I first examined overall frequency patterns to identify which pronouns appear in contemporary online discourse, then conducted a detailed collocation analysis to understand how technological constraints can influence Chinese language users' pronoun choices.

4. Corpus Results

The corpus analysis revealed three key patterns that illuminate how digital writing infrastructure shapes contemporary Chinese pronoun use. First, stark frequency differences between established and emerging pronouns highlight immediate barriers to adopting Chinese gender-inclusive pronouns. Second, collocation patterns reveal both similarities

and differences between ‘TA’ and binary pronouns—while shared patterns demonstrate ‘TA’ functioning as a standard pronoun, distinct patterns show its broader utility beyond binary alternatives. Third, specific discourse contexts favouring ‘TA’ over binary pronouns suggest practical considerations can outweigh concerns about maintaining traditional orthographic standards—a tension that echoes centuries of debate about Chinese language reforms (Tsu, 2022; Zhou, 2003).

4.1. Pronoun Frequency

The KWIC analysis revealed pronounced differences in the frequency of third-person pronouns in zhTenTen17. As shown in Table 1, 他 ‘he’ occurred over 34 million times (0.21% of corpus), more than twice as frequently as 她 ‘she’ (15+ million hits, 0.093%). The disparity between the corpus frequency of 他 compared to 她 reflects several factors: 他 appears in compound words (e.g., *qítā* 其他 ‘other’), transliterations (e.g., *jítā* 吉他 ‘guitar’), and crucially, in the default plural pronoun *tāmen* 他们 ‘they’. The higher frequency of 他 in news content, particularly in coverage of political figures, also contributes to this disparity.

TABLE 1. KWIC results for Chinese pronouns in Chinese Web Corpus 2017 (zh-TenTen17)

Pronoun	Total hits	Clean hits	Hits per million tokens	Percentage of corpus
他	34,456,382	–	2,076.54	0.21%
她	15,355,986	–	925.44	0.093%
TA	17,071	15,969	1.03	0.000093%
它	9,558,920	–	576.08	0.058%
祂	16,918	–	1.02	0.00010%
牠	4,832	–	0.29	–
牠	35	0	–	–
「无也」	54	0	–	–

The romanised form ‘TA’ appeared much less frequently than 他 ‘he’ and 她 ‘she’, with 17,071 total occurrences in zhTenTen17. After removing false hits (such as abbreviations for ‘teaching assistant’ or airline flight numbers), 15,969 instances remained where ‘TA’ functioned as a pronoun. Most notably for this investigation of technological constraints, the Sinograph neopronouns 「无也」 and 牠 were entirely absent as pronouns; while these graphs appeared 54 and 35 times respectively, manual examination revealed these were all false hits in mathematical or programming contexts. The WuDaoCorpora dataset showed similar

patterns, suggesting these frequency differences persist in more recent online Chinese discourse. More broadly, these frequency patterns provide initial evidence of how technological constraints—namely the ability to type and display characters—shape which pronouns can achieve widespread use in contemporary Chinese discourse.

4.2. Collocation Patterns

Having established pronounced differences in pronoun frequency, I next examined how the pronouns function in context through collocation analysis. Analysing the collocation patterns revealed that the use of ‘TA’ shows both similarities and differences with binary pronouns 他 ‘he’ and 她 ‘she’. The possessive marker *de* 的 appeared as the most frequent immediate right collocate for all pronouns (‘TA’: 28%, 她: 27%, 他: 25%; i.e., TA的 ‘TA’s’, 她的 ‘her’, 他的 ‘his’). This suggests that ‘TA’ serves a standard pronominal function in contemporary Chinese discourse. However, the three pronouns showed distinct patterns in their other common collocates. While 他 ‘he’ was associated strongly with verbs of speech and authority (e.g., *shuō* 说 ‘speak’, *biǎoshì* 表示 ‘express’, *zhīchū* 指出 ‘point out’), and 她 ‘she’ frequently appeared in passive constructions with *ràng* 让 and *gěi* 给, ‘TA’ demonstrated more versatile positioning. As shown in Figure 2, the top collocates of ‘TA’ by log-likelihood included both active and passive markers (*gěi* 给 #3, *ràng* 让 #4, *duì* 对 #8, *gàosù* 告诉 #9), suggesting ‘TA’ functions flexibly in various syntactic positions and discourse roles.

Two collocation patterns proved particularly revealing about how ‘TA’ functions in contemporary Chinese discourse, patterns that partially explain why this pinyin form has achieved greater success than Sinograph alternatives. First, the plural marker *mén* 们 ranked notably higher for ‘TA’ (#2 in both log-likelihood and MI rankings) compared to binary pronouns (*tāmen* 他们 #19, *tāmen* 她们 #124 for 他 and 她 respectively). This suggests ‘TA’ frequently serves to de-gender plural references, offering an alternative to *tāmen* 他们 ‘they’ for referring to mixed-gender groups. Second, a cluster of ‘love’ and ‘relationship’ related collocates appeared among the top MI-ranked collocates for ‘TA’. As shown in Figure 3, this cluster includes *xīnài* 心爱 ‘beloved’ (#4), *xīnyí* 心仪 ‘admire’ (#6), and *ài* 爱 ‘love’ (#17), along with *xiàngmào* 相貌 ‘appearance’ (#1), *yuǎnlí* 远离 ‘be far away from’ (#3) and *ānwèi* 安慰 ‘comfort’ (#5).

To investigate whether this apparent clustering of ‘love’ and ‘relationship’ collocates was distinctive to ‘TA’, I conducted a keyword analysis comparing the ‘TA’ concordance data against reference corpora of 他 ‘he’ and 她 ‘she’. A keyword is a word that appears with a statistically significant higher frequency in the target corpus compared to the reference corpus (Anthony, 2023; Pojanapunya and Todd, 2016). The key-

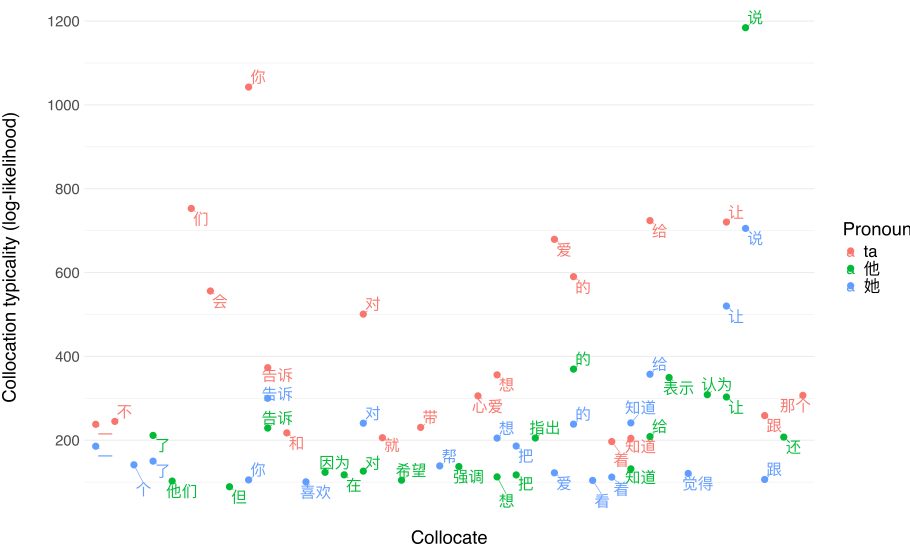


FIG. 2. Top 20 collocates of each pronoun by log-likelihood score

word analysis returned similar results regardless of which binary pronoun served as the reference for ‘TA’. As shown in Table 2, relationship-related words were much more likely to appear with ‘TA’ than either 他 ‘he’ or 她 ‘she’.

TABLE 2. Keyword analysis of Chinese pronouns in Chinese Web Corpus 2017 (zhTenTen17)

Rank	TA vs. 他	TA vs. 她
1	ta ‘they’	ta ‘they’
2	你 ‘you’	你 ‘you’
3	会 ‘can/will’	会 ‘can/will’
4	爱 ‘love’	爱 ‘love’
5	不 ‘not’	宝宝 ‘baby’
6	人 ‘person’	对方 ‘other party; significant other’
7	宝宝 ‘baby’	可以 ‘can/may’
8	爱情 ‘love/romance’	如果 ‘if’
9	喜欢 ‘like’	人 ‘person’
10	对方 ‘other party; significant other’	爱情 ‘love/romance’
11	如果 ‘if’	用户 ‘user’
12	可以 ‘can/may’	浪漫 ‘romantic’
13	吧 (particle)	app ‘app’
14	就 ‘then’	情人节 ‘Valentine’s Day’
15	时候 ‘time/moment’	或者 ‘or’

TABLE 2 (continued)

Rank	TA vs. 他	TA vs. 她
16	让 'let/make'	可能 'possible'
17	浪漫 'romantic'	最 'most'
18	朋友 'friend'	款 'sum; section'
19	想 'think/want'	有 'have'
20	是 'is/are'	们 (plural suffix)
21	那么 'so/then'	分手 'break up'
22	有 'have'	喜欢 'like'
23	孩子 'child'	软件 'software'
24	你们 'you (plural)'	的 'of'
25	分手 'break up'	伴侣 'partner'

Relationship-related words with high keyness scores for 'TA' included *bǎobǎo* 宝宝 'baby' (#5/7), *duìfāng* 对方 'other party; significant other' (#8/10), *àiqíng* 爱情 'romantic love' (#11/10) and *làngmàn* 浪漫 'romance' (#13/17). These words all had a normed frequency in the 'TA' corpus that was significantly greater than their normed frequency in the binary pronoun reference corpora, providing evidence to suggest that this cluster of 'love'/'relationship' collocates distinctly appear with 'TA'.

4.3. Discourse Contexts For 'TA'

To better understand the contexts in which 'TA' appears in contemporary Chinese discourse, I conducted detailed analysis of the concordance data. This examination revealed two main contexts where 'TA' is frequently chosen over binary pronouns 他 'he', 她 'she' or the construction 他/她 'he/she' in contemporary online discourse. First, 'TA' is used as a convenient gender-neutral way for social media and e-commerce platforms to communicate with or represent their users. For example, the social media platform Renren could generate this message to users: "TA们还在人人网等着你" ("**they** [TA] are waiting for you on Renren.com").

Second, 'TA' often appears in online discussions of love or relationships where the referent's gender is deliberately unspecified. In many cases, this appears to facilitate more open discussion of romantic sentiments towards another person by representing them as an abstract, ideal person (e.g., my 'Miss/Mr. Right')

(1) 第二步：选择你喜欢的QQ个性红包吧，有很多个性红包供你选择。第三步：把你满满的祝福和爱意发给心爱的TA吧。

Step two: Choose a QQ personalised red envelope that you like. There are many personalised red envelopes to choose from. Step three: Send your heartfelt wishes and love to your **beloved** [TA].



FIG. 3. Top 20 collocates of each pronoun by mutual information (MI) score

(2) 不管是男人还是女人，爱上一个人时，可以图TA的相貌，图TA的内在，甚至是TA的物质，但千万别掏心掏肺的去爱上TA，得想办法让TA爱上你。

No matter if it's a man or a woman, when you fall in love with someone, you may be attracted to **their** [TA's] appearance, **their** [TA's] inner qualities or even **their** [TA's] material possessions. But never ever pour your whole heart and soul into loving **that person** [TA]. You must find a way to make **them** [TA] fall in love with you.

These usage patterns suggest that, in mainstream contemporary discourse, 'TA' functions primarily as a gender-neutral pronoun—one that avoids specifying gender while still operating within the binary framework—and only infrequently as an explicitly gender-inclusive pronoun. In Example (2), the phrase “不管是男人还是女人” (“no matter if it's a man or woman”) frames 'TA' as operating within, rather than challenging, the gender binary. The broader utility of 'TA' as a gender-neutral pronoun—particularly in relationship discourse and platform user communication—appears to be key to its growing mainstream acceptance. Meanwhile, the technical barriers to typing and displaying Sinograph neopronouns appear to be severely limiting their adoption beyond specialised queer spaces. This contrast highlights how both practical utility and technical accessibility shape which new pronouns can gain widespread use in contemporary Chinese.

5. Discussion

The corpus analysis reveals how globalised technological infrastructure actively shapes, rather than simply enables or constrains, the emergence of gender-inclusive pronouns in contemporary Chinese. The data demonstrates three interconnected patterns: the complete absence of Sinograph neopronouns despite their creation predating many instances of ‘TA’; the emergence of ‘TA’ as a functionally flexible alternative serving both standard pronominal and distinctive discourse roles; and the adoption of ‘TA’ primarily for its practical utility in avoiding gender specification rather than explicitly for gender inclusivity. Together, these patterns illuminate how technical constraints actively shape the possibilities for Chinese pronoun innovation.

First, the absence of Sinograph neopronouns from a 13.5 billion word corpus is striking. This cannot be attributed solely to their recent creation; 他 was created in 2015, predating many instances of ‘TA’ in the data. Rather, the corpus evidence suggests that technical barriers to typing and displaying new Sinographs effectively prevent linguistic innovations from entering general usage, regardless of their potential social acceptance. This mirrors historical tensions in Chinese writing reform, where practical implementation requirements have long shaped which innovations succeed (Huang, 2009; Tsu, 2022).

Second, the collocation patterns provide further evidence of how technological constraints influence contemporary pronoun development. ‘TA’ demonstrates both standard pronominal functions—with high frequencies of possessive and plural constructions and both active and passive markers—and distinctive usage patterns centred around relationship-focused discourse. These patterns suggest that ‘TA’ has achieved wider adoption not just as a substitute for binary pronouns, but as a practical solution to situations where gender specification is unnecessary or undesirable. Like English singular ‘they’, ‘TA’ can thus serve either as a gender-neutral pronoun for unspecified binary gender or as a gender-inclusive pronoun for nonbinary referents (Bjorkman, 2017).

Third, the specific discourse contexts revealed by the corpus data further illuminate how practical utility drives pronoun innovation. ‘TA’ appears frequently in platform communication and relationship discussions, contexts where gender neutrality serves communicative purposes beyond deliberate gender inclusivity. Survey data reinforces this finding, with queer Chinese language users noting that ‘TA’ “suits people’s everyday pronunciation habits” and “doesn’t create issues with having to specify someone’s gender” (Davey, 2024). This parallels the historical introduction of 她 ‘she’ in the 1920s, where successful adoption depended not just on social acceptance but on matching existing language conventions (for a detailed discussion, see Davey, 2024). However, while the technical hurdles for writing 她 could be addressed through local inno-

vation (Huang, 2009), contemporary Chinese pronoun innovation faces globalised technological constraints.

These three interconnected patterns revealed by the corpus analysis together highlight how contemporary Chinese pronoun innovation operates within fundamentally different constraints than both alphabetic systems and historical Chinese writing technologies. New English pronouns like ‘ze/zir’ can be typed as soon as they are conceived, and other writing technologies at least allowed for local creation and reproduction of new Sinographs through handwriting, or else woodblock or metal type printing (Brokaw and Reed, 2010). Conversely, while contemporary Chinese input systems are highly sophisticated—with typing speeds that can match or exceed those achieved in alphabetic systems (Mullaney, 2024)—new Sinographs require centralised institutional approval and global technical implementation. A newly encoded Sinograph must be incorporated into IMEs before it can be typed, into predictive input algorithms before it surfaces as a candidate, and into font sets before it can be displayed, with each step requiring independent institutional action at a global scale. The trajectory of 他 illustrates this: formally encoded in Unicode 17.0 in September 2025—more than a decade after its creation—it remains absent from mainstream IMEs and inaccessible to most users. Unlike historical writing technologies where local solutions allowed for Sinograph creation and reproduction, none of these steps can be achieved by individual language users or communities acting alone. The result is that orthographic innovation in digital Chinese writing remains effectively beyond the reach of those it is meant to serve.

The corpus results therefore reveal how institutional technological constraints are actively steering the emergence of Chinese gender-inclusive pronouns. While the identical pronunciation of all third-person pronouns as *tā* in Modern Standard Chinese might seem to facilitate gender-inclusive language development, the technical requirements of digital Sinograph creation have pushed development toward solutions that work within existing digital parameters. In particular, the corpus evidence shows that while ‘TA’ has found viable use cases that facilitate its adoption, Sinograph neopronouns remain effectively excluded from mainstream discourse. This disparity cannot be explained by social factors alone. Rather, the complete absence of Sinograph neopronouns from the corpus data demonstrates how the technical requirements of digital writing—particularly Unicode encoding and input method compatibility—create fundamental limitations on orthographic innovation that differ markedly from both historical Chinese writing technologies and contemporary alphabetic systems. Survey data from queer Chinese speakers reinforces this interpretation (see Davey, 2024). While many survey respondents expressed preference for a ‘real’ Chinese pronoun, they consistently noted that Sinograph neopronouns were “hard to type” and “cannot be displayed by many computer systems”. The emergence

of 'TA' in the corpus data, coupled with the complete absence of Sino-graph neopronouns, reveals how practical technological considerations can outweigh traditional orthographic concerns.

More broadly, these findings underscore how globalised technological infrastructure fundamentally reshapes the possibilities for linguistic innovation in contemporary Chinese. While historical barriers to Sinograph creation could be overcome through technological and social change at a local level, creating new Sinographs in modern digital writing systems requires multiple layers of global institutional approval and technical implementation that remain inaccessible to ordinary language users. The case of gender-inclusive pronouns illustrates this transformation clearly: despite the identical *tā* pronunciation providing a seemingly natural foundation for innovation, the technical parameters of digital writing have actively steered development toward solutions like 'TA' that work within existing systems. This highlights how global technological infrastructure plays a formative role in shaping contemporary language change, moving beyond merely enabling or constraining linguistic innovation to actively steering its development through institutionalised technical requirements.

6. Conclusion

The development of gender-inclusive pronouns in Chinese offers a window into how technological infrastructure shapes linguistic possibilities. While the introduction of gender-distinguished pronouns a century ago also required navigation of social resistance and existing linguistic conventions, these could at least be addressed within China. Contemporary Chinese pronoun innovation faces an additional challenge: the centralised technological parameters and institutional approval processes of digital Chinese writing. The corpus evidence presented in this paper reveals how these constraints manifest in actual language use, with Sino-graph neopronouns remaining absent from mainstream discourse while 'TA' gains acceptance through its broad utility as a dual gender-neutral and gender-inclusive pronoun. These findings have implications beyond the specific case of Chinese gender-inclusive pronouns. As written Chinese continues to evolve in the digital age, the requirements of Unicode encoding and input method compatibility create specific barriers to linguistic innovation even as Chinese digital writing demonstrates remarkable efficiency and adaptability in other respects. Yet the emergence of 'TA' suggests an alternative path—one where practical utility may sometimes outweigh adherence to traditional orthographic conventions. This highlights how the technological infrastructure of modern writing systems does not simply enable or constrain linguistic innovation but actively shapes the direction of language change.

References

- Anthony, Laurence (2022a). *AntConc (version 4.2.0)*. URL: <https://www.laurenceanthony.net/software>.
- (2022b). *TagAnt (version 2.0.5)*. URL: <https://www.laurenceanthony.net/software>.
- (Mar. 17, 2023). "Common statistics used in corpus linguistics". In: URL: https://laurenceanthony.net/resources/statistics/common_statistics_used_in_corpus_linguistics.pdf.
- Arnold, Jennifer E., Heather C. Mayo, and Lisa Dong (2021). "My pronouns are *they/them*: Talking about pronouns changes how pronouns are understood". In: *Psychonomic Bulletin & Review* 28.5, pp. 1688–1697. ISSN: 1069-9384, 1531-5320. DOI: 10.3758/s13423-021-01905-0. (Visited on 12/19/2024).
- Bjorkman, Bronwyn M. (2017). "Singular *they* and the syntactic representation of gender in English". In: *Glossa: A journal of general linguistics* 2.1. ISSN: 2397-1835. DOI: 10.5334/gjgl.374. (Visited on 12/19/2024).
- Bradley, Evan D. (Mar. 2020). "The influence of linguistic and social attitudes on grammaticality judgments of singular 'they'". In: *Language Sciences* 78, p. 101272. ISSN: 03880001. DOI: 10.1016/j.langsci.2020.101272.
- Brokaw, Cynthia Joanne and Christopher A. Reed, eds. (2010). *From woodblocks to the Internet: Chinese publishing and print culture in transition, circa 1800 to 2008*. Sinica Leidensia. Leiden ; Boston: Brill. 440 pp.
- Cheng, Vincent T.-W. (2016). "Radical equity: A case study of the use of Chinese third-person pronouns in Taiwan's third grade social studies textbook". In: *Intercultural Communication Studies* 25.3, pp. 99–111.
- Davey, Janet (2024). "Taking 'TA' beyond the binary: In search of multi-modal gender-inclusive pronouns in Chinese". In: *Image [ε] Narrative* 25.3.
- Desagulier, Guillaume (2017). *Corpus linguistics and statistics with R*. New York, NY: Springer Berlin Heidelberg.
- Farris, Catherine S. (1988). "Gender and grammar in Chinese: With implications for language universals". In: *Modern China* 14.3, pp. 277–308.
- Ferrari, Pisana (Apr. 25, 2022). "The gradual rise of gender-neutral pronouns in Hong Kong and mainland China". In: *cApStAn*. URL: <https://web.archive.org/capstan.be/the-gradual-rise-of-gender-neutral-pronouns-in-hong-kong-and-mainland-china/> (visited on 05/10/2022).
- Hodge, Bob and Kam Louie (1998). *The politics of Chinese language and culture: The art of reading dragons*. Culture and communication in Asia. London, New York: Routledge. 182 pp.
- Hong, Yeonok (2020). "Change in Chinese personal pronouns from a typological perspective". In: *The Linguistics Journal* 14.1, pp. 233–258.
- Huang, Xingtao (2009). "她"字的文化史: 女性新代词的发明与认同研究 [*The cultural history of the Chinese character "她": A study of the creation and acceptance*

- of a new feminine pronoun]. 福建 [Fujian]: 福建教育出版社 [Fujian Education Press].
- Jung-Palandri, Angela (1991). "Gender and sexism in Chinese language and literature". In: *Schriftfestschrift: Essays on writing and language in honor of John DeFrancis on his eightieth birthday, Sino-Platonic Papers*. Ed. by Victor H. Mair. Vol. 27. Sino-Platonic Papers. Philadelphia, PA: Department of Oriental Studies, University of Pennsylvania, pp. 167–170.
- Koon, Wee Kek (Feb. 24, 2022). "How non-binary Chinese people seeking a gender-neutral third-person pronoun for themselves are coming up with solutions, including a new Chinese character". In: *South China Morning Post*. URL: <https://www.scmp.com/magazines/post-magazine/short-reads/article/3168100/how-non-binary-chinese-people-seeking-gender> (visited on 05/10/2022).
- Ling, Yuanzheng (1989). "“她”字的创造历史 [The history of the Chinese character “她”]". In: *语言教学与研究 [Language Teaching and Linguistic Studies]* 4, pp. 139–151.
- Liu, Lydia He (1995). *Translingual practice: Literature, national culture, and translated modernity—China, 1900-1937*. Stanford, CA: Stanford University Press. 474 pp.
- Lu, Zhao (June 26, 2021). "他/TA/X也: What pronouns do Chinese queer people use?" In: *RADII*. URL: <https://radii.co/article/pronouns-chinese-queers> (visited on 05/01/2022).
- Moser, David (1997). "Covert sexism in Mandarin Chinese". In: *Sino-Platonic Papers*. Ed. by Victor H. Mair. Vol. 74. Philadelphia, PA: Department of Asian and Middle Eastern Studies, University of Pennsylvania, pp. 1–23.
- Mullaney, Thomas S. (2017). *The Chinese typewriter: A history*. Cambridge, MA: MIT Press.
- (2024). *The Chinese computer: A global history of the information age*. Studies of the Weatherhead East Asian Institute, Columbia University. Cambridge, MA: MIT Press.
- Myers, James (2019). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. New York ; London: Routledge.
- Papadopoulos, Ben and Zipeng Zhu (Aug. 2, 2022). "A conversation with Zipeng Zhu". In: *Gender in Language Project*. URL: www.genderinlanguage.com/mandarin/zipengzhu (visited on 05/10/2022).
- Pojanapunya, Punjaporn and Richard Todd (2016). "Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis". In: *Corpus Linguistics and Linguistic Theory* 14. DOI: 10.1515/cllt-2015-0030.
- Posit Team (2023). *RStudio: Integrated development environment for R*. Boston, MA. URL: <http://www.posit.co/>.
- Qualtrics (2023). *Qualtrics XM*. Version April 2023. URL: <https://www.qualtrics.com>.

- Quinan, C. L. and Mina Hunt (2023). "Non-binary gender markers: Mobility, migration, and media reception in Europe and beyond". In: *European Journal of Women's Studies* 30.3, pp. 380–390. ISSN: 1350-5068, 1461-7420. DOI: 10.1177/13505068211024891. (Visited on 12/19/2024).
- R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria. URL: <https://www.R-project.org/>.
- Robertson, Meg and Riki Thompson (2022). "Shifting attitudes, shifting styles: Codifying gender-inclusive language use in style manuals". In: *Attitudes towards gender-inclusive language: A multinational perspective*, September 8–9, Queen Mary University of London, London, UK.
- Sheydaei, Iman (2021). "Gender identity and nonbinary pronoun use: Exploring reference strategies for referents of unknown gender". In: *Gender and Language* 15.3. ISSN: 1747-633X, 1747-6321. DOI: 10.1558/genl.18871. (Visited on 12/19/2024).
- Sketch Engine (Sept. 17, 2020). *ZhTenTen – Chinese corpus from the web*. URL: <https://www.sketchengine.eu/zhtenten-chinese-corpus/#toggle-id-1>.
- Su, Jui-Ju et al. (2012). "When 'he' can also be 'she': An ERP study of reflexive pronoun resolution in written Mandarin Chinese". In: *Frontiers in Psychology* 7. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2016.00151. (Visited on 12/19/2024).
- Tang, Chelsea, Irene Yi, and Cooper Bedin (2022). "Mandarin Chinese". In: *Gender in Language Project*. URL: www.genderinlanguage.com/mandarin (visited on 05/06/2022).
- Troncoso, Solangel C., Zach C. Schudson, and Susan A. Gelman (June 2023). "Women versus females: Gender essentialism in everyday language". In: *Journal of Psycholinguistic Research* 52.3, pp. 975–995. ISSN: 0090-6905, 1573-6555. DOI: 10.1007/s10936-022-09917-0. (Visited on 12/19/2024).
- Tsu, Jing (2022). *Kingdom of characters: The language revolution that made China modern*. New York: Riverhead Books.
- UK (May 1, 2021). "IRG N2487: UK Submission for IRG Working Set 2021". In: *IRG Collection 2021, Universal Multiple-Octet Coded Character Set, International Organization for Standardization*. URL: <https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2487UKData002.pdf> (visited on 05/22/2023).
- United Nations (2012). "中文性别包容性语言指南 [Guidelines for gender-inclusive language in Chinese]". In: URL: <https://www.un.org/zh/gender-inclusive-language/guidelines.shtml> (visited on 05/02/2022).
- Wu, Junru (2022). "Neurocognition in the reading of Chinese characters". In: *Journal of Chinese Writing Systems* 6.3, pp. 167–183. DOI: 10.1177/25138502211061605.
- Xu, Zhe and Patrick Sturt (2023). "The influence of stereotypical information on gender inference in Chinese". In: *SN Social Sciences* 3.2,

- p. 30. ISSN: 2662-9283. DOI: 10.1007/s43545-023-00618-6. (Visited on 12/19/2024).
- Yuan, Sha et al. (2021). “WuDaoCorpora: A super large-scale Chinese corpora for pre-training language models”. In: *AI Open* 2, pp. 65–68. ISSN: 2666-6510. DOI: 10.1016/j.aiopen.2021.06.001.
- Zhou, Youguang (2003). *The historical evolution of Chinese languages and scripts*. Pathways to Advanced Skills Series. Foreign Language Publications.