

## Female Representation in Han Characters Containing the Radical 女 (nǚ): A Corpus-Based Critical Discourse Analysis of Ideological Construction and Gender Bias

| Noviana Laurencia Surjadja<sup>1\*</sup> | Syihabuddin<sup>2</sup> | Rinaldi Supriadi<sup>3</sup> | Joni Soleman Nalenan<sup>4</sup> | Mohammad Ali<sup>5</sup> |

<sup>1,2,3,4</sup>School of Postgraduates  
Studies, Universitas  
Pendidikan Indonesia,  
Bandung, Indonesia

<sup>5</sup>Faculty of Cultural Studies,  
Universitas Komputer  
Indonesia, Bandung,  
Indonesia

novianalaurencia@upi.edu  
syihabuddin@upi.edu  
rinaldi.supriadi@upi.edu  
joninalenan@upi.edu  
mo.ali@email.unikom.ac.id

### ABSTRACT

Han characters (汉字) constitute one of the oldest continuously used logographic writing systems in the world, and as such they encode millennia of accumulated cultural values, social hierarchies, and ideological assumptions. This study empirically investigates the representation of women through Han characters containing the radical 女 (nǚ, 'woman') by employing a cultural-linguistic framework integrating corpus-based quantitative analysis, semiotic deconstruction, and critical discourse analysis. The research corpus comprises 214 characters drawn from the Kangxi Dictionary (康熙字典) and the Contemporary Chinese Dictionary (现代汉语词典, 7th edition, 2016); a supplementary diachronic corpus of 50 characters with documented oracle bone forms was drawn from the Xiaoxuetang Database of Ancient Chinese Scripts (小學堂). Frequency analysis was conducted against the MARKUS corpus (45 classical texts spanning the Han through Qing dynasties, totaling approximately 2.3 million characters). Results indicate that of the 214 characters analyzed, 61.2% carry connotations that subordinate, restrict, or pathologize the feminine; 22.4% are semantically neutral-descriptive; and only 16.4% bear positive connotations. These findings confirm that the Han writing system functions as an ideological apparatus that reproduces and consolidates patriarchal norms embedded in pre-modern and contemporary Chinese society. The study has implications for language pedagogy in the Indonesian context of Chinese-language education, for script reform policy, and for the broader scholarly debate on writing systems as ideological apparatus.

### KEYWORDS

Corpus-Based; Critical Discourse Analysis; Female Representation; Han Characters; Radical 女

## INTRODUCTION

Writing systems are never ideologically neutral. As materially inscribed technologies of meaning-making, they encode the cultural values, social hierarchies, and power relations of the civilizations that produced them, a proposition that acquires particular force when applied to the Han writing system (汉字, hànzi), used by over 920 million first-language Mandarin speakers. Unlike alphabetic systems, which encode primarily phonological information, Han characters encode meaning through visual composition: each character is a compressed ideological artifact carrying within its morphological structure the assumptions and hierarchies of three millennia of Chinese civilization.

This study investigates whether, and to what degree, this ideological content is systematically gendered, specifically, whether the several hundred Han characters containing the radical 女 (nǚ, 'woman') encode a disproportionately negative representation of femininity. The radical 女 is ranked thirty-eighth among the 214 Kangxi radicals and functions, in most characters containing it, as a genuine semantic determinative. Its appearance across domains as diverse as kinship, moral evaluation, social status, and supernatural agency makes it a particularly productive site for investigating how gender ideology is inscribed at the foundational level of the script.

The study is further motivated by its relevance to Chinese-language education in Indonesia, where Mandarin instruction has expanded significantly in recent decades and pedagogical materials introduce 女-radical characters without contextualizing the ideological logic embedded in their composition. Understanding this logic is prerequisite to developing critically informed instructional approaches in this context.

Prior scholarship has documented the negative skew of 女-radical characters, but has been limited by insufficient corpus size, the absence of inter-rater reliability testing, and limited diachronic evidence. The present study addresses these limitations and is organized around three research questions: (1) What is the distribution of semantic valence across the complete contemporary 女-radical inventory? (2) Through what structural and compositional mechanisms is patriarchal ideology encoded in these characters? (3) What is the diachronic trajectory of this encoding, and have PRC simplification reforms altered its ideological content? The purpose of this article is to provide empirically robust answers to all three questions, with implications for language pedagogy, script reform policy, and the scholarship on writing systems as ideological apparatus.

DeFrancis (1984), though primarily concerned with phonological dimensions, made observations relevant to ideological analysis of character structure. Blum (2002) offered an anthropological reading situating the encoding of gender hierarchy within the broader project of social classification in Chinese statecraft. Idema and Grant (2004) document 女-radical characters whose semantic history reflects genuine celebration of feminine qualities, providing a counterweight to uniformly negative accounts — a corrective the present study incorporates without losing sight of the dominant distributional pattern.

Adopting a lexicographic and corpus-based approach, Li (2020) examined gender representation in Chinese through an analysis of characters containing the radical 女, drawing on a mega-corpus of 287 million Chinese characters. A striking finding was that

more than 70% of 女-radical characters listed in classical Chinese dictionaries have disappeared from contemporary usage, a pattern Li attributes to language evolution and socio-cultural change rather than any deliberate ideological revision of the script.

Wang Binghua (2024), analyzing 女-radical characters through the lens of Chinese social and political history, argues that the ideological content of characters such as 安, 妇, 妒, and 奸 reflects the progressive consolidation of patriarchal norms from the patrilineal clan period onward, and demonstrates that the cultural psychology of male superiority (男尊女卑) encoded in the writing system is directly traceable to specific transformations in feudal political economy. This historically grounded account complements the present study's diachronic component. Tran (2019), drawing on Xu Shen's *Shuowen Jiezi* (说文解字) as a primary corpus, similarly documents the shift from matriarchal to patriarchal social organization as reflected in the semantic trajectory of 女-radical characters, and observes that the beauty standards encoded in characters such as 娴, 婉, and 孀 function as normative constraints on female behavior that are naturalized through the medium of the script itself. The convergence of these Chinese- and Vietnamese-based scholarly perspectives with the findings of the English-language tradition reviewed above strengthens the generalizability of the claim that the 女-radical inventory functions as an ideological apparatus across diverse scholarly and cultural contexts.

Across this body of scholarship, two persistent limitations recur: insufficient corpus size relative to the total 女-radical inventory, and the absence of systematic inter-rater reliability testing for valence coding. The present study addresses both.

### **Han Character Structure and the Kangxi System**

The structural taxonomy of Han characters identifies six morphological categories (六书, liù shū). The Kangxi Dictionary (康熙字典), published in 1716, standardized the radical system by designating exactly 214 radicals and assigning all 47,035 entries to exactly one radical. For the present study, the Kangxi assignment of 女 as the semantic determinative provides the operational definition of the corpus.

The script simplification reforms of 1956–1986 altered the graphic form of approximately 2,000 characters without, as Norman (1988) persuasively argues, substantively modifying the semantic content encoded in their morphological structure. Characters containing 女 were subject to simplification in their phonetic components, but the radical itself which carries the semantic payload relevant to this analysis was preserved intact.

### **Theoretical Framework**

This study draws on three complementary theoretical frameworks. The first is Critical Discourse Analysis (CDA) as developed by Fairclough (1992) and elaborated by van Dijk (1998). CDA proceeds from the premise that language use is simultaneously a social practice and a site of ideological struggle. Applied to the present context, CDA provides the conceptual apparatus for treating each 女-radical character as a 'frozen discourse', a

crystallized instance of social practice that encodes, and in the act of use reproduces, particular configurations of gender power.

The second framework is the theory of linguistic relativity, reformulated by Slobin (1996) as 'thinking for speaking.' The present study operates within the weaker version of this hypothesis: it does not claim that exposure to 女-radical characters with negative valence causes misogynistic beliefs, but argues consistently with psycholinguistic research on lexical priming and implicit association (Meyer & Schvaneveldt, 1971; Greenwald et al., 1998), that such exposure systematically biases the conceptual associations available to speakers when thinking and speaking about women.

The third framework is visual semiotics as articulated by Kress and van Leeuwen (1996). Because Han characters are visual artifacts as much as linguistic units, semiotic analysis of their graphic structure, the spatial arrangement of components, the iconicity of the 女 radical, the relationships of subordination or coordination encoded in compound formations, provides a level of insight that purely linguistic analysis cannot achieve. The original pictographic form of 女, depicting a figure with hands clasped in deference, is itself a semiotic act: it encodes a posture of submission as the defining visual representation of womanhood.

## RESEARCH METHODS

### Research Design

The present study employs a mixed-methods design integrating quantitative corpus analysis with qualitative semiotic and ideological interpretation. The rationale for this combination is straightforward: quantitative methods provide the distributional breadth necessary to establish the statistical significance and representativeness of observed patterns, while qualitative methods provide the interpretive depth necessary to move from description to explanation. Analyses were conducted using descriptive statistics and nonparametric correlation procedures; all calculations were performed and cross-verified manually and with SPSS v.26.

### Corpus Construction

The primary corpus consists of 214 characters identified as containing the radical 女 and currently in active use in standard Mandarin Chinese, as defined by inclusion in the Contemporary Chinese Dictionary (现代汉语词典, 7th ed., 2016), published by the Institute of Linguistics of the Chinese Academy of Social Sciences.

A supplementary historical corpus was assembled from two sources: the Kangxi Dictionary (n = 289 女-radical characters, for comparative purposes) and the Xiaoxuetang Database of Ancient Chinese Scripts (小學堂; <https://xiaoxuetang.org>; accessed January 2026), a comprehensive online repository of oracle bone, bronze, and seal script forms indexed by character entry. Of the 214 primary corpus characters, 50 had sufficiently documented oracle bone forms in this database to permit diachronic valence comparison; the remaining 164 characters were excluded from the diachronic subsample due to absence of attested oracle bone forms or paleographic ambiguity. Character forms were retrieved by

individual entry search, and semantic coding of oracle bone-period meanings drew on the etymological glosses provided within the database, cross-referenced with Xu Zhongshu's (1989) 甲骨文字典 as the primary paleographic authority. Of the 75 characters present in the Kangxi inventory but absent from the contemporary corpus, preliminary valence coding indicates approximately 58% carry negative connotations, 28% are neutral-descriptive, and 14% are positive, a distribution broadly similar to, though somewhat less negatively skewed than, the contemporary corpus (61.2% negative). Crucially, the attrition rate for positive-valence characters is not disproportionately higher than for negative-valence characters, indicating that the contemporary corpus does not dramatically overestimate negative bias relative to the fuller Kangxi inventory.

### **Coding Procedure**

Each character was coded along four dimensions: (1) the function of the radical 女 in the character: semantic (S), phonetic (F), or semantic-phonetic (SF); (2) primary semantic domain; (3) connotative valence: positive (P), negative (N), or neutral (Nt); and (4) frequency in classical texts, from the MARKUS corpus.

Valence coding was subjected to rigorous reliability procedures. Three independent coders each holding postgraduate qualifications in Chinese linguistics and each trained in feminist discourse analysis, coded all 214 characters independently. Positive valence was defined as inclusion of qualities regarded as desirable or admirable in contemporary usage; negative valence as inclusion of qualities regarded as undesirable or morally problematic; neutral valence as purely denotative or relational content without evident evaluative loading. Inter-rater reliability was assessed using Cohen's kappa ( $\kappa$ ). The mean kappa across all coding dimensions was 0.83 (substantial to near-perfect agreement). The three coders operated independently throughout the coding phase, with no consultation until the resolution stage, ensuring procedural independence. Discrepancies concentrated at the neutral-versus-negative boundary were resolved through structured discussion, with consensus decisions recorded for all cases.

### **Analytical Procedures**

Quantitative analysis consisted of frequency and percentage distributions for all coded dimensions; chi-square goodness-of-fit tests to assess whether the observed valence distribution differed significantly from empirically grounded null distributions; and Spearman rank-order correlation between connotative valence and frequency of occurrence in the classical corpus. Qualitative analysis proceeded at two levels: semiotic analysis of a purposively selected character subset, and ideological interpretation through the lens of CDA and Chinese social historiography.

## **RESULTS AND DISCUSSION**

### **Overall Distribution of Connotative Valence**

Table 1 presents the distribution of connotative valence across the 214 characters in the primary corpus. Characters appearing in multiple domain rows in subsequent tables reflect

polysemous entries coded to both primary and secondary semantic fields; this accounts for column totals slightly exceeding  $n = 214$ .

**Table 1.** Distribution of Connotative Valence in 女-Radical Han Characters ( $n = 214$ )

Connotative Valence	Count (n)	Percentage (%)	Description
<b>Negative</b>	131	61.2%	Restricting, subordinating, or pathologizing femininity
<b>Neutral-Descriptive</b>	48	22.4%	Denotative, relational, without clear evaluative loading
<b>Positive</b>	35	16.4%	Celebrating, honoring, or affirming femininity
<b>TOTAL</b>	<b>214</b>	<b>100%</b>	—

The dominance of negative valence characters (61.2%) is striking and statistically robust. To assess whether this distribution departs significantly from a non-biased baseline, two chi-square goodness-of-fit tests were conducted against theoretically motivated null distributions. The primary null hypothesis assumed no directional bias across the three valence categories, distributing expected frequencies as 35% negative, 40% neutral-descriptive, and 25% positive, a distribution consistent with the general expectation that evaluatively marked vocabulary constitutes a minority of any lexical inventory. A chi-square goodness-of-fit test comparing the full observed distribution (61.2% negative, 22.4% neutral, 16.4% positive) against this null yields  $\chi^2(2, N = 214) = 64.93, p < .001$ . As a secondary benchmark, a more conservative null (40% negative, 40% neutral-descriptive, 20% positive) yields  $\chi^2(2, N = 214) = 42.02, p < .001$ . The observed imbalance is thus not attributable to chance and reflects a systematic ideological pattern specific to the 女-radical inventory.

### Distribution by Semantic Domain

The distribution of valence across semantic domains reveals a highly differentiated picture that sheds light on the specific areas in which ideological bias is most concentrated.

**Table 2.** Distribution of 女-Radical Characters by Semantic Domain and Connotative Valence

Semantic Domain	Pos.	Neg.	Neu.	Representative Characters
Character Traits	2	34	4	嫉, 妾, 妒, 婪
Kinship & Family	8	3	22	姐, 妹, 妻, 姑
Social Status	1	28	2	奴, 婢, 嫔, 嬖
Reproductive Processes	5	8	6	娩, 妃, 娠, 媿
Physical Appearance	7	14	3	娇, 媚, 妖, 婀
Emotion & Psychology	3	19	4	嫉, 婪, 嫌, 媿
Morality & Conduct	1	17	3	奸, 娼, 妾, 媾

Leisure & Play	0	8	4	娛, 嫩, 嫖, 媪
Miscellaneous Positive	/ 8	0	0	妙, 嫵, 媪, 婉
<b>TOTAL</b>	<b>35</b>	<b>131</b>	<b>48</b>	

Three findings from Table 2 warrant particular emphasis. First, the domain of Character Traits exhibits the most extreme negative skew, with a negative-to-positive ratio of 17:1. When the 女 radical contributes a semantic dimension of 'character trait' to a compound character, that trait is overwhelmingly likely to be a negative one. Second, the domain of Social Status shows a similarly extreme distributional skew, with 28 of 31 characters (90.3%) coded as negative. The characters in this domain, 奴 (nú, 'slave/servant'), 妾 (qiè, 'concubine'), 婢 (bì, 'maidservant'), 嫔 (pín, 'imperial concubine of low rank'), collectively map a field of social positions characterized by servitude, sexual availability, and dependency. Third, the domain of Kinship and Family presents a contrasting pattern: neutral-denotative characters predominate (22 of 33), reflecting the primarily descriptive-relational character of kinship terminology. Even here, however, asymmetries persist: terms for female relatives acquired through marriage can carry subtly negative connotations in classical usage, which the modern standard dictionary tends to neutralize.

### Semiotic Analysis of Selected Characters

Table 3 presents selected 女-radical characters with negative connotation, and Table 4 presents those with positive connotation.

**Table 3.** Selected 女-Radical Characters with Negative Connotation: Semiotic Analysis

Character	Pinyin	Meaning	Semantic Category
奸	jiān	Wicked, treacherous; adultery, rape	Morality / Character Traits
妒	dù	Jealous, envious (esp. between women)	Emotion / Psychology
婪	lán	Greedy, avaricious, rapacious	Character Traits
妖	yāo	Goblin, demon (fem.); seductive, uncanny	Supernatural / Appearance
嫌	xián	Suspicion, dislike, resentment; grievance	Emotion / Psychology
嫉	jí	Envious (esp. of beauty or talent)	Emotion / Psychology
娼	chāng	Prostitute, female entertainer	Social Role / Morality
妾	qiè	Concubine, secondary wife	Social Status / Kinship
奴	nú	Slave, servant (女 + grasping hand)	Social Status
媾	gòu	Illicit intercourse; alliance by marriage	Conduct / Reproductive

The character 奸 (jiān) demands extended consideration as perhaps the most frequently cited example of explicit misogynistic encoding in the Han script. In its classical form, the character combines three instances of the radical 女, generating the meaning 'wickedness', 'treachery', or 'villainy'. The Shuowen Jiezi (说文解字), the foundational

second-century CE etymological dictionary compiled by Xu Shen, glosses the character with the implication that the aggregation of women produces moral disorder. The Shuowen's gloss is itself evidence of how Confucian ideological consolidation shaped the interpretive tradition of the writing system. Both Tran (2019) and Wang Binghua (2024) note that the Shuowen Jiezi's treatment of 奸 and the related character 姦 encodes a normative equation of femininity with moral and social disorder, a characterization that Wang Binghua (2024) situates within the broader Confucian ideological apparatus that also produced negatively valenced characters such as 妒 (jealousy), 婪 (greed), and 嫉 (envy), all of which direct moral pathology toward the female body.

The character 妖 (yāo) illustrates semantic shift across historical periods. In oracle bone and early bronze inscriptions, the character combined 女 (woman) with 夭 (yāo, young/premature), generating a meaning of 'young woman of unusual beauty', a reading not straightforwardly negative. However, by the Warring States period, the semantic field of 妖 expanded to encompass supernatural malevolence: the 'uncanny beauty' of the young woman was reinterpreted as the seductive appearance of a demonic entity. This trajectory, from 'extraordinarily beautiful woman' to 'dangerous supernatural being', encapsulates the broader cultural logic by which female attractiveness is constructed as a form of threat in the classical Chinese imaginary.

The character 奴 (nú, 'slave/servant') provides a structural analysis of particular clarity. The character combines 女 (woman) with 又 (yòu, a hand making a grasping gesture), visually depicting a woman being seized or held, a graphic representation of possession or capture. In its contemporary simplified form, the character retains both components, preserving the ideological encoding of the original composition despite its distance from the historical social practices that gave rise to it.

**Table 4.** Selected 女-Radical Characters with Positive Connotation: Semiotic Analysis

Character	Pinyin	Meaning	Semantic Category
好	hǎo	Good, fine, virtuous (女 woman + 子 son)	Value / Social Norms
妙	miào	Wondrous, exquisite, sublime	Aesthetics
娴	xián	Graceful, refined, skilled in the arts	Character / Ability
嫦	cháng	Lunar deity Chang'e (嫦娥)	Mythology / Proper Name
姝	shū	Beautiful, virtuous woman (archaic poetic term)	Appearance / Character
婵	chán	Graceful; beautiful (as in 嫦娥, 'the moon')	Aesthetics / Mythology

The character 好 (hǎo, 'good') is simultaneously the most commonly used 女-radical character in contemporary Mandarin and, on ideological analysis, one of the most revealing. The structural logic through which 'goodness' is generated, the combination of 女 (woman) and 子 (zǐ, son or child), locates the meaning of 'good' in the relationship between a woman and a male child. Boltz (1994) holds that 子 may refer to a child of unspecified sex, depicting a culturally universal image of warmth and care. The present study follows the patrilineal

reading of 好, interpreting 子 as specifically denoting a male child, as more consistent with the broader sociohistorical evidence for patrilineal bias in classical Confucian discourse, while acknowledging that the etymology of 好 remains a legitimate site of scholarly debate. Tran (2019) similarly highlights 好 as an exemplary case in which the Shuowen Jiezi encodes female virtue in terms of physical beauty and relational function, observing that the evolution of this character's semantic range, from a specific denotation of female attractiveness to a generalized value term, traces the process by which patriarchal ideological content becomes naturalized within the most ordinary vocabulary of daily life.

The character 妙 (miào, 'wondrous, exquisite') combines 女 with 少 (shǎo, 'few/young'), generating a composition that traditional philology glosses as 'the extraordinary quality of female youth'. Like 妖, this character encodes a presupposition that female value is tied to youth and physical appearance, though in this case with a positive rather than threatening evaluation. This form of 'positive' representation defines female excellence in terms of a transient physical quality rather than intellectual, moral, or social achievement, aligning with the broader feminist critique of beauty standards as a form of ideological control. Wang Binghua (2024) situates this aesthetic standard within the wider social apparatus that confined women to domestic roles, observing that prescriptive texts such as the Nü Jie (女诫) and Nü Lun Yu (女论语) systematized precisely this model of female virtue (softness, gentleness, and physical refinement) as the normative basis for women's education across the imperial period. The encoding of such standards within the compositional logic of individual characters thus participates in a broader disciplinary project that operated simultaneously at the level of script, conduct literature, and legal regulation.

### Diachronic Analysis

Analysis of the 50 characters for which well-documented oracle bone forms are available reveals significant diachronic variation that complicates any simple narrative of stable patriarchal encoding from the earliest period of the script's history. Table 5 presents a summary of the diachronic patterns observed.

**Table 5.** Summary of Diachronic Semantic Shifts in Selected 女-Radical Characters (n = 50)

Trajectory Type	Count (n)	Percentage (%)	Description
Stable negative	24	48.0%	Negative valence present from oracle bone through modern period
Intensified negative	11	22.0%	Neutral/mildly neg. in oracle bone; strongly neg. in classical and modern
Ameliorated	7	14.0%	Negative in classical usage; neutral or positive in modern standard
Stable neutral/pos.	8	16.0%	Neutral or positive valence throughout documented history

The pattern of 'intensified negative' (22% of the subsample) represents a historically significant process: the progressive semantic deterioration of characters that began with relatively neutral or only mildly negative meanings. The character 妾 (qiè, 'concubine') appears in oracle bone texts primarily as a descriptive term for a specific social institution without strongly negative evaluative loading; the negative connotation intensifies through the Han and Tang dynasties as Confucian prescriptive literature increasingly constructs the concubine as a moral threat to household harmony.

The pattern of 'amelioration' (14%), though quantitatively smaller, is theoretically significant as a corrective to any account of unidirectional ideological deterioration. Several characters bearing strongly negative connotations in classical usage have acquired more neutral uses in contemporary Mandarin, a process of semantic bleaching in which the historically specific negative connotation loses salience without the underlying structural encoding being altered. With respect to the simplification reforms, the diachronic analysis confirms Norman's (1988) judgment: the reforms modified the graphic form of characters but did not alter their semantic content. The radical 女 was preserved intact in all simplified forms.

The Spearman rank-order correlation between connotative valence and frequency of occurrence in the MARKUS classical corpus yields  $\rho = -0.31$  ( $p = .002$ ). This moderately negative correlation indicates that characters with more negative valence tend to appear more frequently in classical texts, consistent with the hypothesis that negative-valence 女-radical characters are not marginal lexical items but are woven into the fabric of classical Chinese literary and philosophical discourse.

Taken together, these three sets of findings converge on a coherent picture rather than three isolated observations. The overall distributional finding, that 61.2% of 女-radical characters carry negative connotations against a baseline expectation of 25–40%, establishes that the bias is statistically robust at the level of the corpus as a whole. The domain-specific findings sharpen this picture: the bias is not evenly distributed but is concentrated precisely in the domains most central to everyday vocabulary and moral evaluation, namely Character Traits (17:1 negative-to-positive ratio) and Social Status (90.3% negative). The frequency correlation ( $\rho = -0.31$ ,  $p = .002$ ) adds a third dimension, showing that these negatively valenced characters are not peripheral or archaic items but are disproportionately represented among the characters most frequently encountered in classical texts — and, by extension, in the literary and pedagogical canon through which Mandarin literacy is transmitted. The convergence of magnitude, concentration, and frequency is what grounds the theoretical and pedagogical discussion that follows: it is not merely that the script contains some negatively coded characters, but that this coding is statistically dominant, semantically concentrated in core domains, and historically embedded in the texts most central to the transmission of literacy itself.

### **The Han Script as Ideological Apparatus**

The aggregate findings invite interpretation through Althusser's (1971) concept of the ideological state apparatus (ISA). It is acknowledged that Althusser's original formulation

of the ISA refers to institutional apparatuses (schools, churches, legal systems), rather than to individual semiotic artifacts. The present application extends Althusser's framework to the writing system as a whole, following Pennycook's (2001) argument that language itself functions as an ideological apparatus insofar as it operates through the diffuse, decentralized reproduction of subject positions rather than through centralized institutional mandate. The Han writing system qualifies for this extended reading precisely because its ideological content is transmitted through the compulsory literacy curriculum itself an Althusserian ISA in the strict sense, making the script and the institution inseparable in practice. This extension is not without precedent: Kress and van Leeuwen (1996) similarly argue that semiotic systems carry ideological freight independently of the institutions that deploy them, and Street (1984) has documented how literacy practices reproduce ideological formations that exceed any single institutional site. The Han writing system operates as an ISA in exactly the sense Althusser describes: every instance of reading or writing a 女-radical character is an act of interpellation, the learner may be recruited into a subject position that tends to naturalize the associations between femininity and the semantic contents encoded in those characters. Because the writing system presents itself as a neutral technical system, its ideological content is built into the fabric of the script rather than explicitly displayed. The student who learns that 奸 means 'wicked' is not typically told that this meaning is encoded through the juxtaposition of three female-radical components, or that this compositional logic reflects a historical patriarchal worldview.

This analysis connects to Bourdieu's (1991) concept of linguistic habitus: the system of durable, transposable dispositions acquired through participation in linguistic fields. The child who acquires Mandarin literacy in a standard educational setting acquires a linguistic habitus shaped, among other things, by the distributional patterns of 女-radical characters documented in this study. The substantial body of research on lexical priming (Meyer & Schvaneveldt, 1971), implicit association (Greenwald et al., 1998), and language-attitude interaction (Slobin, 1996) is consistent with the hypothesis that systematic exposure to the ideological content of the writing system may plausibly influence the conceptual associations available to its users, a hypothesis consistent with the psycholinguistic literature on lexical priming and implicit association, though one that requires direct empirical testing before causal claims can be sustained.

The present study documents the distributional preconditions for such ideological effects; whether and to what degree these effects are realized in individual speakers remains an empirical question for future psycholinguistic research, as acknowledged in the Limitations section.

### **Comparative Perspectives**

Studies of English orthography and lexicography have documented a comparable, if differently structured, pattern of gender-ideological encoding notably the male-as-default principle by which 'man' or '-man' serves as the unmarked form for generic reference, analyzed as symbolic androcentrism by Spender (1980) and Cameron (1985). The dynamics in Han script are structurally different: the issue is semantic content encoded in graphic

composition rather than morphological marking. Nevertheless, the underlying ideological mechanism is comparable in both cases.

Japanese kanji, which largely inherited the Han character inventory, shares the 女 radical and many of the characters analyzed here, such that comparable ideological analyses would yield comparable findings. Korean hanja presents a historical instance of a writing system that carried the same ideological content, but where the phonological writing system (hangul) eventually displaced it in most domains, a transition that may have had implications for the everyday accessibility of Han script's ideological encoding. These cross-orthographic comparisons suggest that the phenomenon documented in this study is not specific to Chinese civilization but reflects a broader pattern of patriarchal inscription in writing systems associated with Confucian cultural spheres.

### **Implications for Chinese-Language Education in Indonesia**

The findings reported above carry direct pedagogical implications, particularly for Mandarin-as-a-foreign-language instruction in Indonesia. Mandarin-language education in Indonesia reemerged at significant scale following the post-Suharto democratization of the early 2000s, after three decades during which Chinese-language education was effectively prohibited under the New Order regime. This expansion has been rapid, with instruction now offered across hundreds of schools and dozens of universities (Suryadinata, 2008), and serves a student population that is not culturally homogeneous: it includes ethnic Chinese Indonesians reconnecting with a heritage script, as well as non-Chinese Indonesians acquiring Mandarin as a foreign language. For the former group, the encounter with 女-radical characters may carry affective dimensions tied to cultural belonging, which can mask their ideological content; for the latter, the script's compositional logic is entirely unfamiliar, leaving learners potentially more receptive to critical framing but also more likely to receive its content as natural and authoritative. In both cases, standard curricula currently introduce 女-radical characters early in the learning sequence without contextualizing the ideological logic embedded in their composition, a pedagogical silence that functions as a form of ideological naturalization by default.

A critical pedagogy of Han character literacy, consistent with the framework of critical language awareness (CLA) developed by Fairclough (1992) and applied to language education by Wallace (1992), would address this gap through three stages. The first, morphological transparency, teaches learners to decompose characters into their components and identify the semantic function of each, making the compositional logic of 女-radical characters visible rather than opaque. The second, historical contextualization, situates the ideological content identified through compositional analysis within its specific historical origins, the consolidation of Confucian patriarchal norms during the Han and Tang dynasties, so that learners understand this content as historically produced rather than natural or universal. The third, critical comparison, has learners compare the gender ideologies encoded in Chinese script with those of other linguistic systems they navigate, including Indonesian, where gender-neutral pronouns and largely non-gendered morphology present a contrasting case; this comparative dimension is particularly accessible in the multilingual

Indonesian context, where students routinely move between Indonesian, regional languages, English, and Mandarin.

This three-stage framework is presented as a theoretically motivated pedagogical proposal grounded in Indonesia's specific sociohistorical trajectory, rather than as an empirically validated intervention; its classroom implementation and effectiveness remain priority directions for future applied linguistic research in this context. Its transferability to other national contexts, where Mandarin occupies different historical and institutional positions, would require separate investigation.

### **The Script Reform Debate**

The findings of this study have direct relevance to an ongoing debate within Chinese linguistic scholarship about whether script reform should extend beyond formal simplification to address the semantic content of ideologically problematic characters.

The conservationist position holds that the integrity of the Han character system as a unified historical artifact takes precedence over the desire to expunge specific ideological contents, and raises the methodological concern that if ideological problematicity is the criterion for radical modification, the scope of potential revision is effectively unlimited. The present study does not adjudicate this debate. What it does demonstrate, with empirical rigor, is that the magnitude of ideological bias 61.2% negative versus 16.4% positive, concentrated in the domains of character traits, social status, and moral conduct, is sufficiently large and systematic to warrant serious and sustained policy deliberation.

### **Gender, Script, and Power: Intersectional Dimensions**

The ideological encoding documented in this study cannot be adequately analyzed through the lens of gender alone. A fully developed critical account must attend to the intersectional dimensions of the 女-radical inventory, that is, to the ways in which gender ideology articulates with class, ethnicity, and social position within the Han writing system. The distributional patterns documented in Tables 1 and 2 reveal not merely a generalized devaluation of femininity but a differentiated field in which different categories of women are positioned very differently. The Social Status domain (90.3% negative,  $n = 28$ ) encodes a hierarchy of female social positions that is simultaneously gendered and class-structured: characters such as 奴 (*nú*, 'slave'), 妾 (*qiè*, 'concubine'), and 婢 (*bì*, 'maidservant') define femininity in terms of social subordination and economic dependency that is also a mark of low class origin. By contrast, characters in the Mythology domain 嫦 (*cháng*, referring to the lunar deity Chang'e) and 嬋 (*chán*, associated with the moon and its graceful aspect), encode an elite, aestheticized femininity associated with supernatural status, aristocratic refinement, or divine elevation. The writing system thus does not produce a single, undifferentiated image of 'woman' but rather a hierarchically organized field of female subject positions, distributed along axes of class, sexual status, and social function.

This intersectional structure is not incidental but reflects the social organization of gender in imperial Chinese society as documented by historians of women and gender in China (Ko, 1994; Mann, 1997; Bray, 1997). The Confucian framework of *liè nǚ* (列女,

‘exemplary women’) constructed female virtue as a hierarchically ordered set of roles (wife, mother, daughter-in-law, widow) each with its associated obligations and behavioral constraints. This ideological architecture is directly legible in the distributional patterns of the 女-radical inventory: the Kinship domain is dominated by neutral-denotative characters (22 of 33) precisely because kinship terms encode female roles that are structurally necessary for social reproduction, while the Morality and Conduct domain concentrates the most intensely negative valuations, reflecting the Confucian preoccupation with female moral conduct as the guarantor of lineage purity and household order. The writing system, in this light, does not merely reflect pre-existing social structures but actively participates in their reproduction by encoding them in the most elementary level of the script, the morphological composition of individual characters.

### **Contemporary Gender Discourse and the Living Script**

A significant dimension of the contemporary relevance of this research lies in the intersection between the historical ideological content documented here and ongoing gender discourse in contemporary China and the broader Sinophone world. The past decade has witnessed the emergence of a substantial feminist public sphere in mainland China, concentrated initially on social media platforms such as Weibo and WeChat before facing increasing regulatory pressure after 2018. Within this discourse, the gender ideology encoded in Han characters has become an explicit site of critical contestation: feminist bloggers, activists, and scholars have circulated analyses of 女-radical characters, particularly 奸 and its compositional kin, as evidence of the patriarchal foundations of Chinese cultural heritage, generating both popular engagement and institutional resistance (Fincher, 2018; Yang, 2020).

The present study’s empirical findings provide a rigorous scholarly foundation for claims that have, in popular discourse, sometimes been made without systematic evidential support. The difference between a blog post asserting that Chinese characters encode misogyny and a peer-reviewed study demonstrating, with corpus evidence and inter-rater reliability testing, that 61.2% of 女-radical characters carry negative connotations is not merely a difference of tone or register: it is a difference of epistemic warrant. This scholarly grounding matters particularly in contexts, such as the Indonesian academic sphere, where Mandarin studies is a relatively young discipline, where the legitimacy of critical approaches to language and culture may be contested by more conservative pedagogical traditions.

It is also necessary to attend to counter-trends within contemporary Chinese linguistic creativity that complicate any account of the writing system as a static patriarchal apparatus. The digital media environment has stimulated the creation of new characters and character combinations, predominantly through the affordances of Unicode and digital input methods, that at times subvert or challenge the ideological logic of the classical script. Some online feminist communities have experimented with alternative character compositions that deliberately avoid the 女 radical in contexts where it would conventionally appear, or that introduce feminine-positive compositions as counter-lexicographic interventions. While these experiments remain marginal relative to the standard lexicon, they signal that the Han

writing system, like any living semiotic system, is not wholly determined by its historical ideological content but remains a site of contestable and productive meaning-making. The critical pedagogy proposed in this study should acknowledge this creative dimension, avoiding the risk of presenting the script as a wholly oppressive system that leaves no room for agency, resistance, or transformation.

### **Methodological Contributions and Directions for Future Research**

Beyond the substantive findings reported above, the present study offers methodological contributions that are applicable to the broader program of research on ideological encoding in writing systems. The combination of corpus-based frequency analysis, four-dimensional character coding with inter-rater reliability testing, semiotic decomposition, and diachronic trajectory analysis provides a replicable template for investigation of any radical, any writing system, or any ideological dimension. Several extensions of this template represent productive directions for future scholarship.

First, the methodology developed here is directly transferable to the investigation of other semantically charged radicals in the Han inventory. An obvious candidate is the radical 男 (nán, ‘man’), which presents a revealing contrastive case: preliminary analysis suggests that 男-radical characters are distributed with a significantly different valence profile, skewing toward positive and neutral rather than negative, a distributional asymmetry that, if confirmed through systematic study, would provide powerful complementary evidence for the gendered ideological logic of the script. Other radicals with ideologically significant semantic loads, such as 土 (tǔ, ‘earth’) and its associations with agricultural labor; 鬼 (guǐ, ‘ghost/demon’) and its associations with supernatural danger; 奴 (nú, ‘slave’) as a standalone radical component would reward similar treatment.

Second, the study calls for extension into psycholinguistic experimental design. The Spearman correlation ( $\rho = -0.31$ ,  $p = .002$ ) between negative valence and classical corpus frequency established here is a corpus-level finding: it demonstrates that negative-valence 女-radical characters are more frequently deployed in classical texts, but it does not establish the psycholinguistic effects of this exposure on contemporary speakers. Paradigms from the Implicit Association Test (IAT) literature (Greenwald et al., 1998) could be adapted to test whether native Mandarin speakers who display higher literacy and familiarity with 女-radical characters also display stronger gender-association may bias on IAT measures. Such a study would provide a direct empirical bridge between the corpus-linguistic evidence presented here and the broader claim about ideological effects on users of the writing system.

Third, a comparative cross-script study examining Japanese kanji and Korean hanja would substantially enrich the theoretical implications of the present findings. Japanese kanji retains the 女 radical and a significant proportion of the characters analyzed in this study, but the sociocultural context of their use, including Japan’s own trajectory of gender ideology and feminist critique, differs in important respects from the Chinese context. Korean hanja presents a particularly revealing case: historically the vehicle of Confucian literary culture and patriarchal governance, hanja was progressively displaced by hangul through the twentieth century, reaching a state of marginalization in contemporary South

Korean everyday literacy. A comparative study would address the question of whether the displacement of the ideologically marked script by a gender-neutral phonological script has measurable effects on speakers' implicit gender associations, a quasi-natural experiment of considerable theoretical significance for the linguistic relativity hypothesis.

Fourth, the discourse-analytic dimension of the research program merits development through genre-specific studies. The present study has demonstrated that negative-valence 女-radical characters appear with elevated frequency across the classical corpus as a whole; a finer-grained analysis examining the distribution of these characters across specific classical genres, philosophical treatises (such as the Four Books of the Confucian canon), legal codes (such as the Tang Code, *Táng lǚ*), historical chronicles (such as the dynastic histories of the *Shiji* and *Hanshu*), and literary fiction (such as the Tang *chuanqi* tale and Ming novel), would provide a richer account of the genre-specific contexts in which the ideological content of 女-radical characters is mobilized. Preliminary analysis of the MARKUS corpus subsets suggests that legal and philosophical texts exhibit the highest concentration of negatively valenced 女-radical characters, consistent with the hypothesis that these genres function as primary vehicles for normative gender ideology, but this finding requires systematic verification.

### **Limitations**

Several limitations require acknowledgment. First, the valence coding system, despite the inter-rater reliability procedure, involves judgments that are culturally situated and synchronically bounded. The three coders who participated in this study all hold postgraduate qualifications in Chinese linguistics from institutions in the Mandarin-speaking world; their cultural positioning may have shaped the valence assessments in ways that are not fully transparent. Future research might address this limitation by including assessors from a wider range of linguistic and cultural backgrounds, including non-Chinese scholars of gender studies and comparative linguistics. Second, the study analyzes characters in isolation from their contexts of use, whereas actual linguistic practice always involves characters in contexts where their connotations may be modified, ironic, or counter-hegemonic. A character coded as negatively valenced at the lexical level may be deployed in a literary or conversational context that inverts, parodies, or critically exposes that negative valence; the present study, by design, cannot capture this pragmatic dimension. Third, the study does not directly address whether and how exposure to 女-radical characters affects the gender attitudes or behavior of Mandarin speakers, an empirical question for psycholinguistics and social psychology that cannot be answered by corpus analysis alone. Fourth, the study's diachronic component is limited to 50 characters for which well-documented oracle bone forms exist; extending the diachronic analysis to the full 214-character corpus would require paleographic expertise and archival access that exceeded the scope of the present project. Fifth, the restriction of the frequency analysis to the MARKUS classical corpus, while methodologically appropriate for the historical framing of the study, means that the contemporary frequency profiles of 女-radical characters in digital media,

social networks, and contemporary fiction remain uninvestigated, a gap that future corpus-linguistic research should address.

## CONCLUSION

This study set out to answer three research questions: whether 女-radical characters are distributed with a statistically significant negative valence bias; through what structural and compositional mechanisms patriarchal ideology is encoded; and whether diachronic change has altered that encoding. All three are answered affirmatively and with greater precision than prior scholarship has achieved. The 女-radical inventory is characterized by a 61.2% negative valence rate that significantly exceeds the baseline negative rate for the Mandarin lexicon as a whole, is concentrated in the semantically charged domains of character traits, social status, and moral conduct, and is traceable through diachronic evidence to a historically specific process of Confucian ideological consolidation. The simplification reforms of 1956–1986 did not alter this encoding.

For language pedagogy, the evidence mandates a critical language awareness orientation in Mandarin instruction in Indonesia that makes the compositional logic of characters pedagogically visible rather than naturalized. For language policy, the empirical magnitude of the bias provides a substantive evidence base for deliberations about ideologically motivated script reform. For scholarship in East Asian linguistics, gender studies, and the cultural history of writing, the study offers a replicable mixed-methods model applicable to other radicals, other writing systems, and other dimensions of ideological encoding in script.

Future research should develop this foundation in at least three directions: psycholinguistic studies using implicit association test (IAT) paradigms to measure the effects of 女-radical character exposure on speakers' gender attitudes; discourse-analytic studies tracing the deployment of these characters in specific genres and registers; and comparative cross-script studies examining whether the distributional patterns documented here are replicated in Japanese kanji and Korean hanja.

## REFERENCES

- Althusser, L. (1971). Ideology and ideological state apparatuses: Notes towards an investigation. In B. Brewster (Trans.), *Lenin and Philosophy and Other Essays* (pp. 127–186). Monthly Review Press.
- Blum, S. D. (2002). *Portraits of 'Primitives': Ordering Human Kinds in the Chinese Nation*. Rowman & Littlefield.
- Boltz, W. G. (1994). *The Origin and Early Development of the Chinese Writing System*. American Oriental Society.
- Bourdieu, P. (1991). *Language and Symbolic Power* (G. Raymond & M. Adamson, Trans.). Harvard University Press.
- Bray, F. (1997). *Technology and Gender: Fabrics of Power in Late Imperial China*. University of California Press.
- Cameron, D. (1985). *Feminism and Linguistic Theory*. Palgrave Macmillan.

- DeFrancis, J. (1984). *The Chinese Language: Fact and Fantasy*. University of Hawaii Press.
- Fairclough, N. (1992). *Discourse and Social Change*. Polity Press.
- Fincher, L. H. (2018). *Betraying Big Brother: The Feminist Awakening in China*. Verso.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Idema, W. L., & Grant, B. (2004). *The Red Brush: Writing Women of Imperial China*. Harvard University Asia Center.
- Ko, D. (1994). *Teachers of the Inner Chambers: Women and Culture in Seventeenth-Century China*. Stanford University Press.
- Kress, G., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Routledge.
- Li, L. (2020). Gender representation in Chinese language. In W. Wang (Ed.), *Analysing Chinese Language and Discourse across Layers and Genres* (pp. 101–118). John Benjamins Publishing Company.
- Lu, S. (2018). 性别平等视角下的汉字改革 [Han character reform from the perspective of gender equality]. *语言政策研究 [Language Policy Research]*, 17(3), 301–318.
- Mann, S. (1997). *Precious Records: Women in China's Long Eighteenth Century*. Stanford University Press.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227–234. <https://doi.org/10.1037/h0031564>
- Norman, J. (1988). *Chinese*. Cambridge University Press.
- Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Lawrence Erlbaum.
- Street, B. (1984). *Literacy in theory and practice*. Cambridge University Press.
- Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking.' In J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge University Press.
- Spender, D. (1980). *Man Made Language*. Routledge & Kegan Paul.
- Suryadinata, L. (2008). *The rise of China and Chinese overseas: A study of Beijing's changing policy in Southeast Asia*. Institute of Southeast Asian Studies.
- Tran, V. L. (2019). 从《说文》里带有"女"字部的汉字看中国女性的形象 [Viewing Chinese women images via the Chinese radical 女 (female) in Xu Shen's Shuo Wen Jie Zi]. *Ho Chi Minh City University of Education Journal of Science*, 16(10), 673–678.
- van Dijk, T. A. (1998). *Ideology: A Multidisciplinary Approach*. Sage Publications.
- Wallace, C. (1992). Critical literacy awareness in the EFL classroom. In N. Fairclough (Ed.), *Critical Language Awareness* (pp. 59–92). Longman.
- Wang, B. (2024). 从"女"部字看中国古代"男尊女卑"的文化心理 [Looking at the cultural psychology of ancient China's "male superiority and female inferiority" from the perspective of characters with the "female" radical]. *中国高校人文社会科学信息网*

- [China Higher Education Humanities and Social Sciences Information Network].  
Zhengzhou University.
- Xiaoxuetang. (n.d.). 小學堂 [Database of ancient Chinese scripts]. National Taiwan  
University. <https://xiaoxuetang.org> (Accessed January 2026)
- Xu, Z. (1989). 甲骨文字典 [Jiǎgǔwén Zìdiǎn: Dictionary of oracle bone script]. 四川辞书  
出版社 [Sichuan Cishu Chubanshe].
- Yang, Y. (2020). The politics of Chinese feminism: Gender and state-society relations in  
China. *China Quarterly*, 242, 453–474. <https://doi.org/10.1017/S0305741020000661>
- 中国社会科学院语言研究所词典编辑室 [Lexicography Editorial Office, Institute of  
Linguistics, CASS]. (2016). 现代汉语词典 [Contemporary Chinese Dictionary] (7th  
ed.). 商务印书馆 [Commercial Press].