

A CORPUS-BASED DIACHRONIC STUDY OF EPICENE PRONOUNS IN ENGLISH: A CASE OF ANAPHORA TO COMPOUND INDEFINITE PRONOUNS

Qiong Hu
National University of Defense Technology, Changsha 410073, Hunan, China.

Abstract: From old English to modern English, the gender category has shifted from a grammatical system to a natural one, with only pronouns retaining gender marking. *He* and *they* used as epicene pronouns represent a common phenomenon of gender disagreement. This study examines epicene pronoun usage with anaphors restricted to possessives (*his*, *their*) and antecedents to compound indefinite pronouns (*someone*, *somebody*), considering writing form, prefix and suffix semantics, and frequency. Results show that *their* is increasingly used as the epicene pronoun while *his* declines, a trend influenced by social factors as well as the semantics and frequency of antecedents. The use of plural *their* in anaphoric reference to singular indefinite pronouns can be viewed as number variation, and its replacement of *his* as a change in both gender and number category. Among candidates for the gender-neutral function, *their* proves the most promising based on diachronic evidence.

Keywords: Language variation and change; English personal pronouns; Gender category; Epicene pronouns

1 INTRODUCTION

1.1 The English Gender System: History and Present State

English has shifted from grammatical gender to natural gender between Old English and the present day. Old English, like Modern German, had grammatical gender on nouns, but the system disappeared completely in Middle English; Present-day English instead classifies nouns by natural gender: masculine, feminine, and neuter (non-humans) [1].

In Old English, pronouns, adjectives, numerals and other word classes changed form to agree with nouns in gender. Since nouns and their modifiers lost their inflections in Middle English, however, gender marking has survived only on pronouns — a “pronominal gender system” — and remains relatively intact in the third person singular (HE, SHE and IT) [1]. This matches the typological observation that personal pronouns are necessarily included in any gender system and are the word class that retains gender marking the longest as a gender system declines [2]. Gender moreover most often occurs in the third person, followed by the second and first, and singular pronouns show gender distinctions more readily than plural ones [3]; English conforms exactly: pronoun gender appears only in the third person, distinguishing masculine, feminine and neuter in the singular, with no distinction in the plural.

1.2 Epicene Anaphoric Uses of Third-Person Pronouns in English

The primary discourse function of third-person pronouns is anaphora [3]; an anaphoric pronoun must agree with its antecedent in gender, number and case, so antecedents with unclear referents create difficulties for gender choice. Corbett distinguishes three situations of indeterminate gender: (1) generic nouns; (2) non-specific reference; and (3) mixed-gender plural antecedents (only in languages where plural pronouns also distinguish gender) [2].

In English, plural antecedents pose no problem of gender: THEY serves anaphora uniformly. For singular antecedents, however, IT, HE or SHE must be chosen according to gender properties. Specific antecedents have a clear referent gender; but a human-denoting form with non-specific reference has an indeterminate referent, and hence indeterminate gender. Two cases are principal [4]: (1) generic singular common nouns denoting humans; and (2) indefinite pronouns such as *someone* and *somebody*.

English lacks a gender-neutral/epicene third-person singular form (hereafter uniformly called the epicene pronoun). Taking *everyone* as the antecedent, [5] sorted out the possible anaphoric forms: *his*, *her*, *his or her* (*her or his*), *his/her* (*her/his*) and *their* (Table 1).

Table 1 Common Third-Person Pronoun Forms in Anaphora to Compound Indefinite Pronouns [5]

i			%his	
ii			%her	
iii	Everyone had cast	{	his or her (her or his)	} vote.
iv			his/her (her/his)	
v			their	

Note: % indicates that the use is grammatical only in some varieties.

The long-standing absence of an epicene third-person singular pronoun in English has been extensively discussed, mainly in terms of prescriptive conventions and social factors such as the feminist movement, with proposals of various candidates [4, 6]; quantitative treatments are comparatively few. Corpus-based studies have concentrated on the synchronic level [7-12]; diachronic studies have covered mainly Early Modern English before 1800 or two isolated successive periods of Modern English [13-14]; and psycholinguistic experiments have examined diachronic changes in language users' mental representations of the candidate forms [15, 16].

In fact, the two principal candidates, HE and THEY, have long been in competition: research has found signs of THEY overtaking HE as early as 1500-1800 [13], but the diachronic development from 1800 to the present, which would reveal the full course of change in epicene pronoun usage, still awaits supplementation.

Using the Corpus of Historical American English (COHA) and taking anaphora to compound indefinite pronouns as a case study, this paper examines the diachronic change in the epicene anaphoric use of third-person pronouns in American English and attempts to analyse the factors that condition the choice of epicene pronouns [17].

2 SAMPLE SELECTION AND ANNOTATION

2.1 Antecedents: Compound Indefinite Pronouns

The principal human compound indefinite pronouns in Present-day English are *everyone/everybody*, *someone/somebody*, *anyone/anybody*, *no one* and *nobody* [5], and these are taken as the main objects of study in this paper. In preliminary searches, however, a certain number of separated forms were also found in the corpus, such as *some one/some body*, *any one/any body*, *every one/every body* and *no body*. For three reasons these were likewise included in the scope of the study (Table 2): (1) although some scholars point out that the difference between the two writing forms is not merely orthographic but also semantic [18], as the sources from which the compound forms grammaticalized, the separated forms themselves possess the core semantic property of "indefiniteness"; (2) there are many criteria for detecting grammaticalization, and orthographic fusion alone does not warrant the conclusion that the shift from indefinite phrase to compound indefinite pronoun has been completed - *no one*, for instance, is still written as two words yet is regarded as a single word and hence as a compound indefinite pronoun [5]; the boundary between the separated forms and the compound indefinite pronouns is in fact difficult to draw; (3) viewed diachronically, the compound indefinite pronouns were used relatively rarely between the 1810s and the 1930s, which may render the raw data for that period sparse, and adding the data of the separated forms can remedy this problem.

Table 2 Antecedents Searched in the COHA Corpus

Type of compound indefinite pronoun Suffix type	Universal	Assertive	Non-assertive	Negative
-one	everyone/ every one	someone/ some one	anyone/ any one	no one
-body	everybody/ every body	somebody/ some body	anybody/ any body	nobody/ no body

2.2 The Anaphoric Pronouns: Genitive Forms

It has been observed that in language production, subjects are more likely than objects to be pronominalized in subsequent anaphoric reference [19], whereas the anaphoric reference of objects may, besides third-person pronouns, resort to full NPs and other means. One consequence of the greater pronominalization of subjects is that third-person nominative forms, whose main function is anaphoric, are used more frequently in discourse than the corresponding accusative forms — a tendency confirmed by the COHA retrieval results. Accordingly, the accusative forms were excluded from the object of study in this case.

Furthermore, a previous statistical study of the epicene pronouns (HE and THEY) in the BNC found that, in terms of frequency of use, nominative > genitive > accusative > reflexive [11]. As a test, the nominative and genitive forms of HE and THEY - the forms most frequently used as epicene pronouns - were sampled in the COHA corpus: their co-occurrences with the compound indefinite pronoun *everyone* were retrieved, manually annotated, and the coreference rates calculated. When co-occurring with *everyone*, the genitive forms reach coreference rates as high as 53% (*his*) and 91.67% (*their*), whereas the nominative forms reach only 2% (*he*) and 14% (*they*); co-occurrence of genitive forms thus detects coreference with markedly higher accuracy. To improve the accuracy of the measure and the efficiency of manual annotation, the present study focuses on "genitive forms as epicene pronouns anaphoric to indefinite pronouns".

The basic genitive forms of the third-person pronouns in Present-day English are *his*, *her*, *its* and *their*. Searching for *everyone* in combination with *her* and *its* returned not a single entry with *its* anaphoric to *everyone* in an epicene use, and only three instances of *her*; the two were therefore excluded from the subsequent large-scale retrieval and statistics, while *his* or *her*, a candidate with a certain number of attestations, was added. Other possible epicene uses of *her* are touched upon in the results and analysis section.

2.3 Retrieval and Annotation Scheme

The study sample was finally defined as the entries in the COHA corpus in which a genitive pronoun form (*his*, *their* or *his or her*) occurs within a four-word window after a compound indefinite pronoun. The entries returned by batch retrieval of the co-occurrence of indefinite pronouns and candidate epicene forms cannot by themselves determine whether coreference obtains; the author therefore examined every retrieved entry in context to judge whether *everyone* and *his* (and the like) were coreferential. After the ineligible entries were excluded, quantitative analysis was performed on the valid data.

3 RESULTS

3.1 Overall Results of the Corpus Retrieval

A total of 6,532 entries were retrieved in which *his*, *his or her* or *their* occurred within a four-word window after the eight compound indefinite pronouns, of which 2,372 were coreferential, giving an overall precision (i.e., coreference rate) of 36.31%.

3.2 Diachronic Distribution of the Genitive Epicene Anaphoric Pronouns

The diachronic distribution of the co-occurrence of compound indefinite pronouns and epicene anaphoric pronouns is shown in Figures 1 to 8. As for absolute frequencies, almost all the distribution charts display data sparsity ($N < 5$) to varying degrees in certain periods, clearly concentrated in the 1810s-1910s, as in the early and middle portions of Figures 1 to 3; this is largely related to the issue of writing forms mentioned above. Since the separated forms held absolute dominance early on, the data retrieved for the compound indefinite pronouns in that period are extremely sparse, and this factor is taken into account in the subsequent data processing. The co-occurrence of *-body* compound indefinite pronouns with epicene anaphoric pronouns is somewhat more special: because there is no see-saw competition between two writing forms, Figures 5 to 8 do not show a complete absence of early data; nevertheless, since the absolute frequency of the *-body* forms themselves is low, a few extreme values still appear in certain decades, as in the 1810-1890 phase of Figure 8.

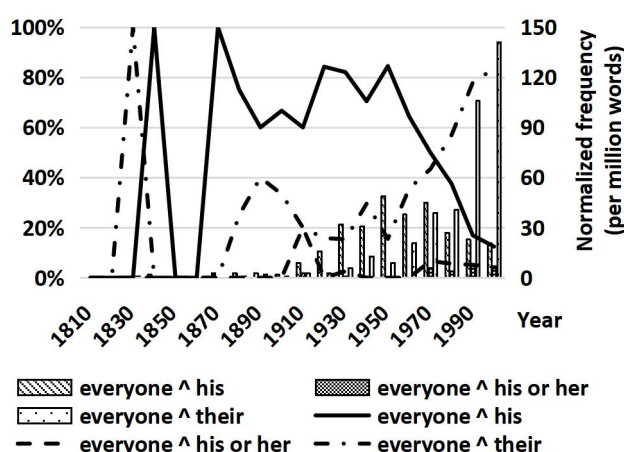


Figure 1 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *everyone* in the COHA Corpus

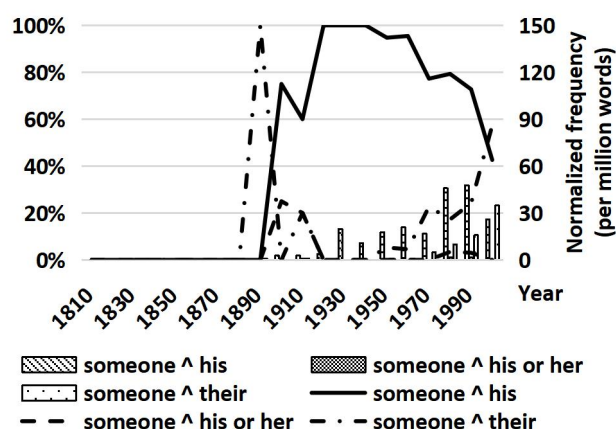


Figure 2 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *someone* in the COHA Corpus

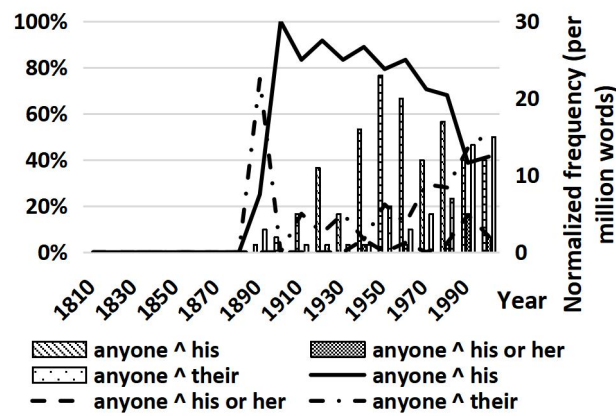


Figure 3 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *anyone* in the COHA Corpus

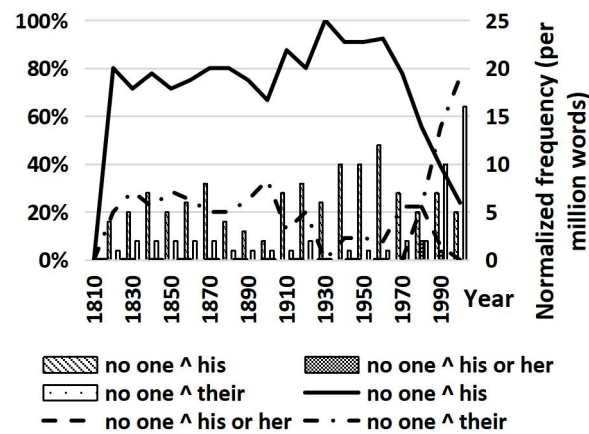


Figure 4 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *no one* in the COHA Corpus

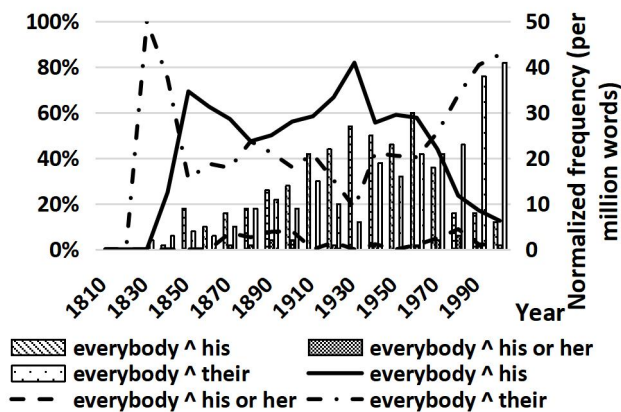


Figure 5 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *everybody* in the COHA Corpus

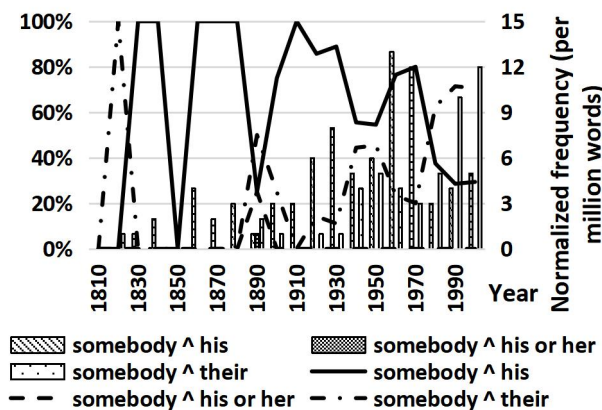


Figure 6 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *somebody* in the COHA Corpus

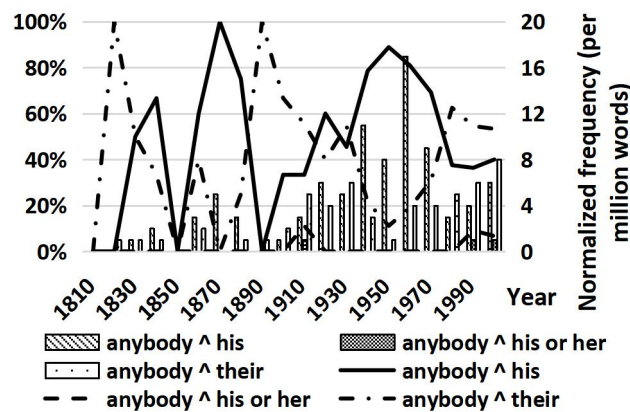


Figure 7 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *anybody* in the COHA Corpus

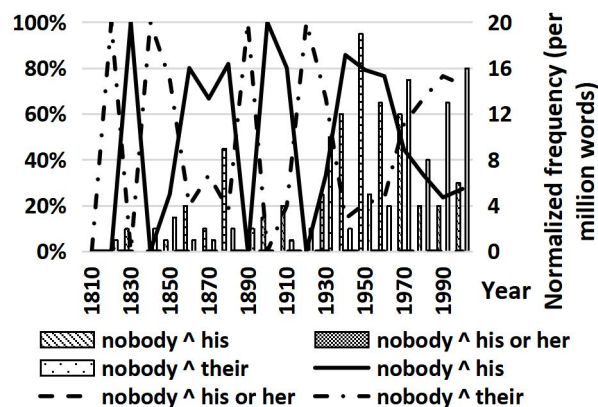


Figure 8 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *nobody* in the COHA Corpus

As for the proportions of the epicene pronouns, observation of Figures 1 to 8 shows that during the 1810s-2000s, when the antecedent is a compound indefinite pronoun, the co-occurring *his or her* within the four-word window, though invariably used as an epicene anaphoric pronoun, is generally low in frequency (absolute frequency per decade ≤ 7 , and zero in most decades), so that the proportion of *his or her* as an epicene anaphoric pronoun remains extremely low; and although the early data for *everyone*, *someone* and *anyone* are sparse or absent owing to the writing-form effect, it can still be observed that, on the whole and with the passage of time, the compound indefinite pronouns developed from using *his* more often as the epicene anaphoric pronoun to preferring *their*.

Statistical tests were carried out on the anaphoric data for these eight compound indefinite pronouns, using Kendall's tau-b correlation analysis. The results are shown in Table 3. In all eight sets of compound indefinite pronoun data, five sets show a significant diachronic increase in "compound indefinite pronoun ^ *their*".

Table 3 Statistical Tests on the Occurrence of Genitive Epicene Anaphoric Pronouns within a Four-Word Window after Compound Indefinite Pronouns in the COHA Corpus

Collocations	Kendall's tau b	p	Collocations	Kendall's tau b	p
Year-everyone ^ his	0.054	0.743	Year-everybody ^ his	0.048	0.770
Year-everyone ^ their	0.532**	0.001	Year-everybody ^ their	0.322*	0.048
everyone ^ his- everyone ^ their	-0.306	0.073	everybody ^ his- everybody ^ their	-0.505**	0.002
Year-someone ^ his	0.478**	0.006	Year-somebody ^ his	-0.155	0.355
Year-someone ^ their	0.560**	0.002	Year-somebody ^ their	0.490**	0.004
someone ^ his- someone ^ their	0.058	0.757	somebody ^ his- somebody ^ their	-0.659***	<.001
Year-anyone ^ his	0.374*	0.029	Year-anybody ^ his	0.215	0.192
Year-anyone ^ their	0.674***	<.001	Year-anybody ^ their	0.064	0.696
anyone ^ his- anyone ^ their	0.199	0.272	anybody ^ his- anybody ^ their	-0.617***	<.001
Year-no one ^ his	0.087	0.601	Year-nobody ^ his	0.027	0.870
Year-no one ^ their	0.049	0.768	Year-nobody ^ their	0.146	0.377
no one ^ his- no one ^ their	-0.734***	<.001	nobody ^ his- nobody ^ their	-0.837***	<.001

Note: Since the data for *his or her* are extremely sparse, the statistical tests exclude it and consider only the pairwise correlations among year, "compound indefinite pronoun ^ *his*" and "compound indefinite pronoun ^ *their*".

* $p < .05$, ** $p < .01$, *** $p < .001$

Because the data are sparse or even absent in some decades, the statistical tests in Table 3 may be affected; as Figures 1 to 8 show, the three sets of *everyone*, *someone* and *anyone* are likely to be the more strongly affected. The author therefore removed the extreme values of these three sets and reran the statistical tests (Table 4).

Table 4 Statistical Tests on the Occurrence of Genitive Epicene Anaphoric Pronouns within a Four-Word Window after Compound Indefinite Pronouns in the COHA Corpus (1910s-2000s)

Collocations	Kendall's tau b	p	Collocations	Kendall's tau b	p
Year-everyone ^ his	-0.644**	0.009	someone ^ his- someone ^ their	-0.905***	<.001
Year-everyone ^ their	0.733**	0.002	Year-anyone ^ his	-0.736**	0.004
everyone ^ his- everyone ^ their	-0.822***	<.001	Year-anyone ^ their	0.629*	0.012
Year-someone ^ his	-.506*	0.046	anyone ^ his- anyone ^ their	-0.837**	0.001
Year-someone ^ their	0.598**	0.018			

* p < .05, ** p < .01, *** p < .001

The extreme values of the *everyone*, *someone* and *anyone* sets are concentrated between the 1810s and the 1900s. Statistical testing of the 1910s-2000s data alone shows that when *everyone*, *someone* and *anyone* serve as antecedents, the choice of *his* as the epicene anaphoric pronoun displays a significant diachronic decrease, the choice of *their* a significant diachronic increase, and in all three sets the uses of *his* and *their* show a significant negative diachronic correlation.

3.3 Co-occurrence of Compound Indefinite Pronouns and Genitive Epicene Anaphoric Pronouns (Writing Form Taken into Account)

Taking the writing-form factor into account largely remedies the sparsity or absence of early data for some parameters. Originally, the three *-one* sets showed no statistically significant correlation between the diachronic uses of *his* and *their* over the 1810s-2000s; once the writing-form factor is included, however, the diachronic uses of these three *-one* forms with *his* and with *their* are all significantly negatively correlated (Figures 9, 10 and 11 and Table 5).

The corpus searches for *his/their* within a four-word window after *every body/some body/any body/no body* returned only 74 entries in total, of which only 28 proved coreferential after manual annotation; merged with the data in Figures 5 to 8, they exert no significant influence on the statistical results, and this subsection therefore does not report them separately.

Table 5 Statistical Tests on the Occurrence of Genitive Epicene Anaphoric Pronouns within a Four-Word Window after Compound Indefinite Pronouns in the COHA Corpus (Writing Form Taken into Account)

Collocations	Kendall's tau b	p
Year-(everyone + every one) ^ his	-0.691***	<.001
Year-(everyone + every one) ^ their	0.716***	<.001
(everyone + every one) ^ his - (everyone + every one) ^ their	-0.899***	<.001
Year-(someone + some one) ^ his	-0.039	0.817
Year-(someone + some one) ^ their	0.440*	0.010
(someone + some one) ^ his - (someone + some one) ^ their	-0.495**	0.005
Year-(anyone + any one) ^ his	-0.123	0.454
Year-(anyone + any one) ^ their	0.264	0.110
(anyone + any one) ^ his - (anyone + any one) ^ their	-0.738***	<.001

* p < .05, ** p < .01, *** p < .001

Table 6 Crossover Points at Which the Proportion of *their* Surpasses That of *his* under Different Compound Indefinite Pronoun Antecedents in the COHA Corpus

Form	Frequency	Crossover Time
everybody	519	1960-1970
nobody	188	1960-1970
anybody	145	1970-1980
somebody	131	1970-1980
everyone	719	1970-1980
no one	174	1980-1990
anyone	203	1980-1990
someone	293	1990-2000

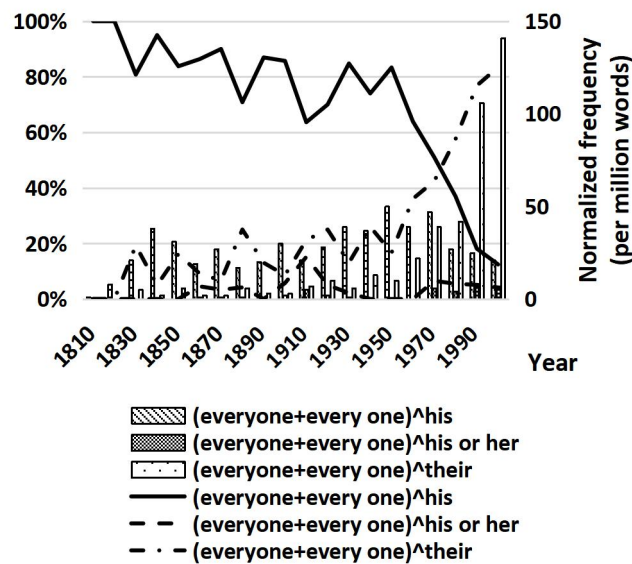


Figure 9 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *everyone* and *every one* in the COHA Corpus

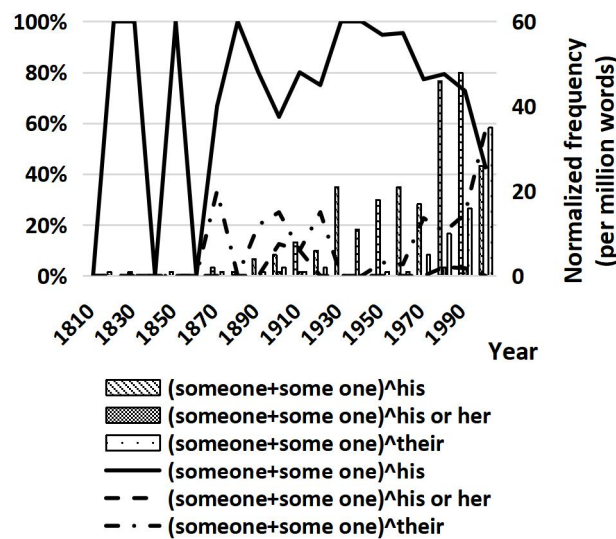


Figure 10 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *someone* and *some one* in the COHA Corpus

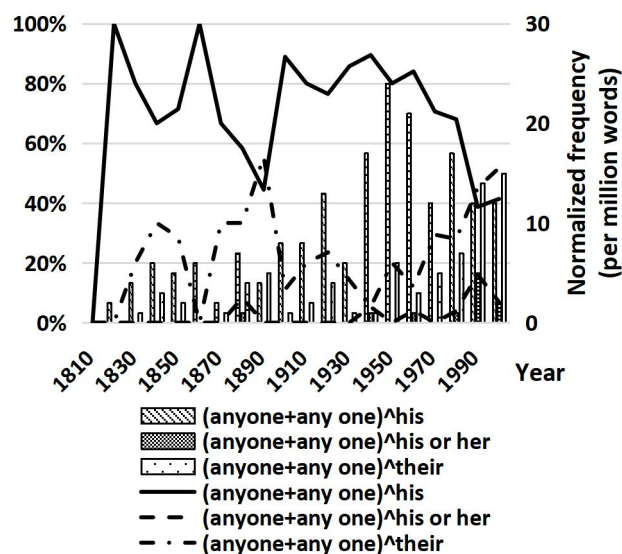


Figure 11 Diachronic Distribution of the Genitive Epicene Pronouns Anaphoric to *anyone* and *any one* in the COHA Corpus

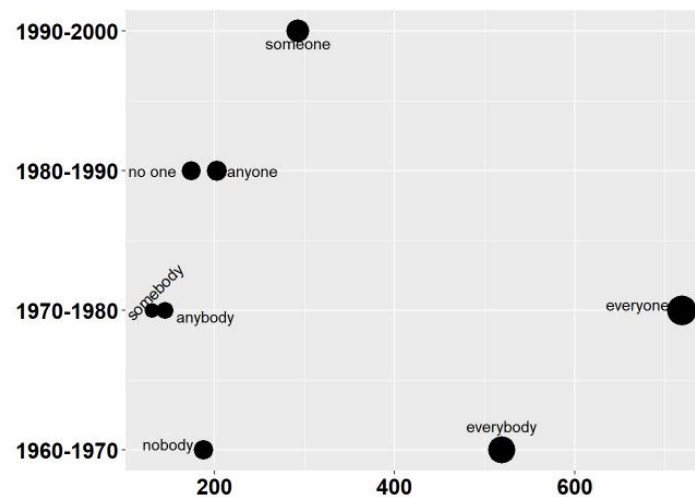


Figure 12 Crossover Points at Which the Proportion of *their* Surpasses That of *his* under Different Compound Indefinite Pronoun Antecedents in the COHA Corpus

4 DISCUSSION

Language change is jointly conditioned by a variety of factors: internal factors of the linguistic system such as syntax and semantics, as well as external forces such as socio-cultural factors [1], cognition and pragmatics. This section discusses the factors conditioning the variation and change in the use of epicene anaphoric pronouns.

4.1 The Conflict between Communicative Needs and the Lack of a Linguistic Form

English possesses content words expressing “indefiniteness” (e.g., indefinite pronouns and generic nouns) and therefore needs epicene pronouns for anaphoric reference to such expressions; yet it has never had a form dedicated to this function. This conflict between communicative need and the lack of an established form is the “fuse” of the coexistence and competition among several candidate forms — and of these candidates, neither HE nor THEY, the two most common choices, is in itself perfect.

Prescriptive convention dictates HE for such cases to preserve number agreement, but its effectiveness depends on the hearer recognizing the generic use: experiments show that generic *he* still biases hearers toward a masculine antecedent [20], because *he* is officially masculine, with generic uses vastly outnumbered by masculine ones, and because *he* outnumbers *she* by a wide margin in actual use [2].

The epicene anaphoric use of THEY long predates the heyday of the prescriptive convention. It avoids gender conflict with the antecedent but, being formally plural, cannot guarantee number agreement.

Other candidate forms in Table 1 fare no better. Feminine SHE mirrors HE: number agreement holds but gender agreement does not. Without a dedicated epicene anaphoric pronoun or a suitable “moonlighting” form, the various candidates have coexisted and competed diachronically under changing syntactic, semantic and socio-cultural factors.

4.2 Syntactic Agreement or Semantic Agreement

When a third-person pronoun serves as anaphora, it may agree either with the antecedent itself or with the entity the antecedent denotes: as in example (1), the verb may take the singular form, agreeing with the antecedent (syntactic agreement), or the plural form, agreeing with the entity it denotes (semantic agreement).

Example (1) a. The committee has decided. b. The committee have decided. [21]

Corbett proposed an “agreement hierarchy”: attributive > predicative > relative pronoun > personal pronoun [21]. For any controller that admits both syntactic and semantic agreement, the further right the target on this hierarchy, the more readily it manifests semantic agreement.

Compound indefinite pronouns are grammatically singular — they have no plural forms (*everyone* vs. **everyones*) — yet they can denote multiple entities and admit a plural semantic interpretation [22]. As the rightmost target on the agreement hierarchy, a personal pronoun readily takes semantic agreement; anaphoric THEY therefore agrees semantically, rather than syntactically, with these antecedents even though they are singular in form. Consequently, those compound indefinite pronouns that in context tend to be construed as semantically plural more readily select the morphologically plural THEY.

Anaphora to compound indefinite pronouns is indeed special: the referent of the antecedent is indeterminate to begin with, and being gender-indeterminate, it readily biases the choice toward THEY; considering the antecedent’s own suffix, however, the “singular” impression conveyed by *-one* also facilitates HE[23]. Although the rise of *their* with *-one* antecedents lags behind that with *-body* ones, in both groups the trend is toward *their* replacing *his* — evidence that in these construction types language users increasingly favor THEY, which carries no explicit gender.

4.3 Semantics and Frequency

Although the data (Figures 1 to 11) show that *their* overtook *his* after the 1960s for every antecedent, the crossover decade differs from antecedent to antecedent: the competition between *their* and *his* proceeded at different rates under different conditions.

Table 6 summarizes, for each antecedent, the decade in which *their* surpassed *his*, together with the frequency with which each compound indefinite pronoun occurs as an antecedent coreferential with an epicene anaphor; the data are also plotted in Figure 12.

In terms of the antecedent suffix, *their* overtakes *his* earlier with *-body* antecedents (crossovers concentrated in the 1960s-1980s) and later with *-one* antecedents (1970s-2000s).

In terms of the antecedent prefix, within the *-body* group the crossovers are earlier with *everybody* and *nobody* (1960s-1970s) than with *anybody* and *somebody* (1970s-1980s); that is, *every-* and *no-* cross over earlier, while *some-* and *any-* lag relatively behind. Although the four *-one* forms display three different crossover times, this does not contradict the observations on the *-body* group.

It can thus be observed that both the prefix and the suffix of the antecedent influence the choice of the anaphoric pronoun. Taking the suffix as an example, linguists differ on whether *-one* and *-body* are complete synonyms, some maintaining that the existence of any form must have its own rationale; in any case, “one” more readily evokes a “singular” construal, and hence *-body* is more receptive than *-one* to *they* as the anaphoric pronoun.

In terms of the frequency of co-occurrence, within the *-body* group the more frequently used *everybody* and *nobody* crossed over earlier (1960s-1970s) than the less frequent *anybody* and *somebody* (1970s-1980s); within the *-one* group, *everyone*, which has the highest coreference frequency, crossed over earliest, and the remaining three later.

Taken together, frequency, the indefinite prefixes of the compound indefinite pronouns, and their suffixes have all influenced the diachronic development of the epicene anaphoric pronouns.

4.4 Socio-Cultural Factors

The use of singular THEY can be traced back to Middle English [5], where it first occurred in informal registers. On the COHA data, however, up to the 1950s the use of THEY, though persistent, remained at a consistently low ebb, mainly because prescriptive grammarians wrote the rule “use HE in anaphora to indefinite pronouns” into their grammar books; flawed as HE was as an epicene anaphoric form, the spread of basic schooling and of print gave this prescribed usage wide currency. In recent years, with the decline of generic HE, the acceptability of generic THEY has risen sharply across registers and styles, which attacked the prescriptive use of HE fiercely; as a result, the competition between THEY and HE took a qualitative leap after the 1960s. The COHA data likewise confirm the role of this extra-linguistic factor: under all eight compound indefinite antecedent conditions the diachronic development of *his* and *their* is broadly congruent (Figures 1 to 11) — *their* rises abruptly around the 1960s while *his* falls abruptly, and within the following forty years the proportion of *their* overtook that of *his*.

Recent English grammars even treat THEY as the more general form, with HE grammatical only in certain varieties (see Table 1) — a sharp contrast with the prescription of HE in earlier grammar books.

4.5 The Broader Context of the Decline of the Gender System

As English inflectional morphology waned in Middle English, grammatical gender was lost; already in Old English, however, when a noun’s grammatical gender conflicted with the natural gender of its referent, anaphoric pronouns could agree with the latter [24] — usage that, while grammatical gender was still in operation, reflected all the more sharply speakers’ perception of natural gender. In Present-day English, natural gender is the sole determinant of noun gender.

A plausible reason for this shift is that grammatical gender rested on convention rather than the real world: speakers had to spend much effort learning it, and for hearers its classification of nouns hardly matched reality or supplied needed information. From both perspectives, its loss accords with the principle of linguistic economy.

In Present-day English, pronoun gender is agreement with the referent’s real-world sex and triggers no morphological or syntactic change elsewhere; it is thus relatively independent of the rest of the system, and its loss would “cost” little. The present trend toward THEY may therefore mark new change after the shift from grammatical to natural gender: even antecedents of unambiguous masculine, feminine or neuter gender are sometimes anaphorically resolved as THEY, though such usage is still far from scale.

Many Indo-European languages have seen a less dramatic but marked decline of grammatical gender [25]. Typologically, English lacks a dedicated epicene anaphoric pronoun, which some languages possess (Zande, for instance, uses *ni* for non-specific individuals [2]); instead, English has produced at least eighty epicene forms since the eighteenth century, none of which caught on [26], and dialects keep generating new candidates, including new third-person singular neuter pronouns such as *yo* [12] and *her* [4]. On the diachronic evidence assembled here, however, THEY is at present the most promising form — whether it will become the only epicene pronoun or be overtaken by other candidates remains to be seen.

5 CONCLUSION

Starting from the history and present state of the English gender system, this paper has focused on the variant uses and the change of English personal pronouns in the gender category, taking epicene anaphoric pronouns as the case in point. The data show that when the antecedent is an indefinite pronoun, the anaphoric pronoun may be chosen from among several forms; overall, the diachronic proportion of *their* has increased significantly with time, and *their* has replaced *his* as the epicene anaphoric pronoun most commonly used by language users. This process has been jointly conditioned by semantic, frequency, syntactic, functional and social factors. Weighing the internal and external factors of language together, the author holds that *THEY* currently displays strong momentum and is likely to replace the other forms in the course of future development, ultimately becoming the only epicene anaphoric pronoun. *THEY* will continue to face competition from other forms, and its “fate” remains to be tested.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by Hunan Provincial Social Science Fund (No. 21YBA034).

REFERENCES

- [1] Curzan A. *Gender Shifts in the History of English*. Cambridge: Cambridge University Press, 2003.
- [2] Corbett G G. *Gender*. Cambridge: Cambridge University Press, 1991.
- [3] Siewierska A. *Person*. Cambridge: Cambridge University Press, 2004.
- [4] Wales K. *Personal pronouns in present-day English*. New York: Cambridge University Press, 1996.
- [5] Huddleston R D, Pullum G K. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.
- [6] Newman M. A theoretical and descriptive study of epicene pronouns: The linguistics of a prescriptive problem. Columbia: Columbia University, 1993.
- [7] Stojanovska-Ilievska N. A corpus-based study of epicene pronouns used by Macedonian learners of English. *Open Linguistics*, 2024, 10(1): 281-308.
- [8] Yakut I, Genç B, Bada E. Epicene pronoun usage in the social sciences: The case of research articles. *Journal of English for Academic Purposes*, 2021, 52: 101005.
- [9] Balhorn M. The Epicene Pronoun in Contemporary Newspaper Prose. *American Speech*, 2009, 84(4): 391-413.
- [10] Baranowski M. Current usage of the epicene pronoun in written English. *Journal of Sociolinguistics*, 2002, 6(3): 378-397.
- [11] Laitinen M. HE and THEY in indefinite anaphora in written present-day English. *International Journal of Corpus Linguistics*, 2003, 7(2): 137-164.
- [12] Stotko E M, Troyer M. A New Gender-neutral Pronoun in Baltimore, Maryland: A Preliminary Study. *American Speech*, 2007, 82(3): 262-279.
- [13] Laitinen M. Indefinite Pronominal Anaphora in English Correspondence between 1500 and 1800. 2004.
- [14] Paterson L L. Epicene pronouns in UK national newspapers: A diachronic study. *ICAME Journal*, 2011, 35: 171-184.
- [15] Noll J, Lowry M, Bryant J. Changes Over Time in the Comprehension of *He* and *They* as Epicene Pronouns. *Journal of Psycholinguistic Research*, 2018, 47(5): 1057-1068.
- [16] Ma Z, Wu S, Xu S. Acceptance and Online Interpretation of “Gender-Neutral Pronouns”: Performance Asymmetry by Chinese English as a Foreign Language Learners. *Frontiers in Psychology*, 2022, 12.
- [17] Davies M. *The Corpus of Historical American English (COHA)*. 2010. <https://www.english-corpora.org/coha/>.
- [18] Haspelmath M. *Indefinite Pronouns*. Oxford: Clarendon Press, 1997.
- [19] Kehler A, Rohde H. Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, 2018: 23-2018.
- [20] Mackay D G, Fulkerson D C. On the Comprehension and Production of Pronouns. *Journal of Verbal Learning and Verbal Behavior*, 1979, 18(6): 661-673.
- [21] Corbett G G. *Agreement*. Cambridge: Cambridge University Press, 2006.
- [22] Quirk R, Greenbaum S, Leech G, et al. *A Comprehensive Grammar of the English Language*. New York: Longman, 1985.
- [23] Laitinen M. Sociolinguistic patterns in grammaticalization: *He*, *they*, and *those* in human indefinite reference. *Language Variation and Change*, 2008, 20(1): 155-185.
- [24] Baugh A C, Cable T. *A History of the English Language*. 5th ed. London: Routledge, 2002.
- [25] Ibrahim M H. *Grammatical Gender: Its Origin and Development*. The Hague: Mouton, 1973.
- [26] Baron D E. *Grammar and Gender*. New Haven and London: Yale University Press, 1986.