

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Language change in L2 academic writing: The case of epicene pronouns



Charlotte Stormbom

English Language and Literature, Åbo Akademi University, Tehtaankatu 2, FI-20500, Turku, Finland

ARTICLE INFO

Article history:

Received 2 May 2018

Received in revised form 27 November 2018

Accepted 1 February 2019

Available online 5 February 2019

Keywords:

Epicene pronouns

Singular *they*

Academic writing

EFL

Diachronic corpus study

ABSTRACT

With the increasing demand for gender-fair language, an important issue to take into consideration in academic writing is the use of so-called epicene pronouns, i.e. singular personal pronouns unspecified for gender. Yet, little research has focused on how L2 English academic writers use these pronouns. The present study examines the use of epicene pronouns since the 1970s in L2 academic writing and analyses their current use, starting from 2010. The data originate from two corpora with academic papers written by Swedish-speaking university students of English: (1) the Finland-Swedish *BATMAT corpus*, which is used to extract diachronic data, and (2) the Swedish subset of the *Varieties of English for Specific Purposes dAtabase*, VESPA-SE, which adds to the data of current use. The results show that although the use of generic *he* has decreased noticeably since the 1970s, there is considerable variation in and between the texts from the 2010s. Such variation is indicative of a language change that is still very much in progress. The findings also suggest that singular *they* is establishing itself as a third-person singular pronoun in L2 academic writing since this pronoun is used frequently with all types of epicene antecedents in the data of current use.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

In recent decades, feminist language planning efforts have resulted in the development and implementation of non-sexist language policies in English-language communities, for instance in the US print media (Fasold, 1987; Fasold, Yamada, Robinson, & Barish, 1990) and in the Australian Public Service (Pauwels, 1998). In the context of academic writing, too, style guides and writing manuals now tend to include recommendations for how to avoid gender-bias in language. For the user of English, a pertinent issue is the choice of 'epicene pronouns', i.e. third-person singular pronouns unspecified for gender (see e.g. Baron, 1981). The three most common types of epicene pronouns in native (L1) English today are generic *he*, coordinations like *he or she*, and singular *they* (see e.g. Huddleston & Pullum, 2002). The use of epicene pronouns has been problematised in prescriptive grammars of English since the end of the 17th century, when the only prescribed solution was to use generic *he*; other solutions were described as 'unwieldy', in the case of coordinated forms, or plain and simple 'ungrammatical', in the case of singular *they*. The widespread prescription of *he* lasted until the second half of the 20th century, when, supported by research findings, feminist linguists asserted that generic *he* is not generic at all because it tends to evoke male imagery (for a review, see Henley & Abueg, 2003). Although singular *they* is now widely used in English, explicit

E-mail address: charlotte.storbom@abo.fi.

endorsement of this pronoun is still rare in grammars and style guides, which instead tend to promote the use of avoidance strategies, such as pluralisation of the antecedent (see e.g. Paterson, 2014).

The complexity surrounding epicene pronouns in English raises the question of how non-native (L2) users approach the issue. In this context, it is highly relevant to consider academic writing because academic writers, in particular, are likely to strive for a language use that is perceived as standard at various levels. Yet, few studies have focused on the use of epicene pronouns in L2 writing. The following questions relating to the topic are addressed in the present study:

- (1) What changes, if any, can be seen in the use of epicene pronouns in L2 academic English writing since the 1970s?
- (2) What factors influence current, i.e. 2010 and onwards, use of epicene pronouns in L2 academic English writing?
- (3) How homogeneous is the use of epicene pronouns in a given group of L2 English academic writers?

These questions are examined using two corpora of academic texts written by L1 speakers of Swedish. The writers are university students of English language and literature in Sweden and in Finland.

2. Feminist language reform in English and Swedish

The second wave of feminism in the 1970s gave rise to a new kind of linguistic activism in English-language communities, which has been described as the ‘non-sexist language reform’ (Pauwels, 2000, p. 270). An important goal of such feminist activism was to raise awareness of gender bias inherent in language, the idea being that language contributes to and maintains gender inequalities in society. A prominent feature of feminist language reform in English has thus been corpus planning (see Kloss, 1969), that is, the proposal of new lexical items, such as *spokesperson* for *spokesman*, as well as new grammatical practices, such as gender-fair alternatives to generic *he*. As Cameron (1995) notes, critics of such non-sexist proposals do not usually argue that sexism in itself is tolerable, but instead claim that specific uses are not biased or that they are simply a matter of linguistic convention. As regards the spread of non-sexist terms, corpus research suggests that new terms tend to be more successful when they are based on already existing words or combinations of words, such as *police officer*, rather than being completely new inventions (Baker, 2010, p. 146).

Evidence of the success of feminist language reform can be seen in its endorsement by official language agencies in recent decades (see e.g. Liddicoat, 2011). As Pauwels (2000: 270) states, “[t]he adoption and implementation of gender-inclusive or non-sexist guidelines is now well advanced in most English language countries”. In general, the media have been quick at adopting feminist proposals for added equality. In terms of academic writing, influential style guides like the *Chicago Manual of Style* and the *Publication Manual of the American Psychological Association* (APA) now include sections on how to avoid gender bias in language, such as using pluralisation instead of generic *he*. There is also evidence from corpus studies that academic writing has become more gender fair since the second wave of feminism in the 1970s. For instance, in a diachronic study of androcentrism in four APA journals, it was concluded that the use of generic *he* decreased significantly between the decade 1965–1974 and the decade 1975–1984, and that no instances of *he* were found in the articles from 1985 until 2004 (Hegarty & Buechel, 2006). However, in a recent study of academic textbooks from the social sciences, the pronoun *he* still accounted for about one third of all epicene pronouns in the period from 1995 to 2005 (Parini, 2012). The effects of feminist language reform in academic writing are thus likely to be related to the field of research, as well as to how forcefully editors and publishers push the issue of non-sexist language use.

Attempts at implementing feminist language reform have been undertaken in many other language communities as well in the past few decades. One factor influencing the success of such attempts is the political climate in a given society (see e.g. Cameron, 1995; Milles, 2011). The debate about gender-fair language is also partly dependent on the particular features of a language, for instance how extensive grammatical gender agreement patterns are. Two common strategies for achieving gender-fair language are feminisation and neutralisation. Whereas feminisation aims at making the female visible in language, the focus of neutralisation is to eliminate gender-marking where it is not needed. In general, the former of these strategies is more common in gendered languages with extensive agreements, such as French and German, whereas the latter has been preferred in languages with less gender-marking, for instance English and the Scandinavian languages (see e.g. Pauwels, 2000).

In Swedish, the main resource for expressing gender is personal pronouns: similarly to English, Swedish has a set of pronouns that is used specifically for personal reference, *han* ‘he’ and *hon* ‘she’. The choice of a personal pronoun can be difficult in cases where a referent’s gender is unknown or otherwise unspecified. As in English, a commonly suggested solution in Swedish is to use coordinations or composites, such as *han eller hon* ‘he or she’ and *han/hon* ‘he/she’. Linguistic authorities in Sweden have generally been fast at adopting feminist language reforms, which, as suggested by Milles (2011), relates to the fact that Sweden is ranked among the world’s most gender-equal countries. The debate on gender-fair language in Swedish received new focus in the late 2000s, when the new gender-neutral pronoun *hen* was introduced into the language. Although pronouns tend to be regarded as a ‘closed class’, this neologism quickly received widespread recognition, and in 2015 it was included in *Svenska Akademiens ordlista* (SAOL, 2015), the glossary of the Swedish Academy. This development in Swedish is unusual from a linguistic typology point of view because it is rare for languages to use a special pronoun when the gender of the referent is indefinite (Corbett, 1991, p. 223).

3. Epicene pronouns in English writing

3.1. L1 English, L2 English and ELF

English has become a truly global language, and as far as professional academic writing goes, non-native writers now by far outnumber native writers of English. This calls into question the viability of native English as a norm in academic writing. Although native English is not considered normative as such in the present study, the use of epicene pronouns in L1 English writing is taken as a starting point. The reasoning behind this is that the corpus texts examined in the study were produced by university students of English in Sweden and Finland, who are advised to orient themselves towards a native English norm in their academic writing. The writers are considered to be advanced L2 learners of English, rather than users of English as a Lingua Franca (ELF). One reason for this is that the students are all L1 Swedish speakers, whereas ELF users by definition do not share a common language and socio-cultural background (see e.g. [Mauranen, 2012](#)). Another important factor is the context in which the corpus texts were produced: the corpora consist of exam papers and Bachelor's and Master's theses; that is, the texts have been read and assessed by university teachers and they do not serve any other communicative purpose as such.

3.2. L1 English writing

Numerous studies since the 1970s have examined the use of epicene pronouns in L1 writing. These studies unanimously show that the previously prescribed generic *he* is no longer the epicene pronoun of choice in L1 English; instead, the use of the previously proscribed singular *they* has increased dramatically, with the result that *they* is now the most frequently used pronoun with epicene antecedents of all kinds (see e.g. [Baranowski, 2002](#); [Laitinen, 2007](#); [Paterson, 2011, 2014](#)). Given the apparent variability of the pronoun *they*, [Paterson \(2014\)](#) suggests that there are, in fact, two homonymous *they* pronouns in English: one that is plural, and one that is singular. This view differs from the usual understanding of singular *they* as a manipulation of the plural form.

Although singular *they* is now exceedingly common in L1 English, some variation is still found in the use of epicene pronouns. Indeed, a certain level of heterogeneity is to be expected, given that the shift from *he* to *they* as the most common pronoun is a very recent change in the English language (for discussions of language change, see e.g. [Chambers & Schilling, 2013](#)). One factor that seems to influence L1 users' pronoun choice is the level of formality. Corpus studies suggest that generic *he* is more often used in very formal contexts, such as legal texts (see e.g. [Holmes, 1998](#); [Paterson, 2014](#)); the opposite is true for singular *they*, which appears to be particularly common in informal language (see e.g. [Newman, 1992](#)). Similarly, coordinated forms are rare in informal language (see e.g. [Curzan, 2014](#); [Newman, 1992](#)), whereas they seem to have a wider distribution in more formal genres, such as assessed university student writing ([Curzan, 2014](#); [LaScotte, 2016](#); [Meyers, 1990](#); [Mitchell, 1994](#)). Although prescriptive grammars and style guides no longer endorse the use of generic *he*, research shows that they still do not tend to promote singular *they* to any larger extent ([Paterson, 2014](#)). This lack of visibility may make language users shy away from singular *they* in writing that requires a relatively formal tone and style, such as academic writing. In addition, it could also explain why individual L1 users are not necessarily consistent in their use of epicene pronouns, but sometimes use singular *they* alongside another type of pronoun (see [LaScotte, 2016](#)).

L1 users' choice of epicene pronouns can also be affected by the type of antecedent that is referenced. Epicene antecedents can be divided into four types: definite noun phrases (NPs), e.g. *the child*; indefinite NPs, e.g. *a student*; quantificational NPs, e.g. *every person*; and indefinite pronouns, e.g. *anyone*. These types differ in terms of their degree of individuation, which, in this particular context, describes how easily an antecedent can be distinguished as a specific individual (see e.g. [Newman, 1998](#)). [Timberlake \(1977\)](#) lists the following features of individuation: proper/common, human/animate/inanimate, concrete/abstract, singular/plural, definite/indefinite. For all of these features, the form to the left is more individuated than the one to the right. In terms of epicene antecedents, one relevant feature is the distinction definite/indefinite. The most individuated type of antecedent is NPs that are definite. This is seen in L1 speakers' choice of epicene pronouns: Previous studies suggest that definite NPs have the highest co-occurrence with *he* out of the four antecedent types (see e.g. [Baranowski, 2002](#); [Gerner, 2000](#); [Newman, 1992](#); [Paterson, 2014](#)).

Another relevant distinction is that of singular versus plural. Although all antecedents of epicene pronouns are, by definition, syntactically singular, antecedents with quantifiers tend to have a plural meaning, as in *everyone* (indefinite pronoun) and *each person* (quantificational NP). Consequently, these types of antecedents are less individuated than definite and indefinite NPs, which are both clearly syntactically and semantically singular and they are therefore easier to conceptualise as specific individuals. This singular/plural distinction is reflected in L1 speakers' use of epicene pronouns in the sense that quantificational NPs and indefinite pronouns are more likely to take *they* than are the other two antecedent types (see e.g. [Baranowski, 2002](#); [Newman, 1992](#); [Paterson, 2014](#)). Research also shows that prescriptive grammars tend to give examples with indefinite pronouns as antecedents in cases where they actually promote the use of singular *they* ([Paterson, 2014](#), p. 127). Such a tendency suggests that the pronoun *they* is less marked with indefinite pronouns than with NPs.

Another aspect related to the antecedent is gender stereotyping: although epicene antecedents are lexically unspecified for sex, personal nouns can still be stereotypically biased; contrast, for instance, the male-stereotyped noun *mechanic* and the female-stereotyped noun *secretary* (for discussions, see e.g. [Abudaljuh, 2012](#); [Carreiras, Garnham, Oakhill, & Cain, 1996](#); [Doherty & Conklin, 2017](#); [Kennison & Trofe, 2003](#)). Stereotyping has been shown to affect the choice of epicene pronouns

in the sense that male-stereotyped nouns are more likely than neutral nouns to be referenced with generic *he* (see e.g. Newman, 1992; Paterson, 2014). Stereotyping can, in a sense, be subsumed under the umbrella notion of individuation as stereotypes make the referent more easily distinguishable as a specific individual.

Although individuation and stereotyping influence the frequency of singular *they*, it should be noted that *they* is now the most common epicene pronoun with all types of epicene antecedents (see e.g. Baranowski, 2002; Newman, 1992; Paterson, 2014). This development arguably supports Paterson's suggestion that *they* is a singular pronoun in its own right, since *they* also coindexes with singular antecedents that are high in individuation, such as definite NPs. In fact, singular *they* is sometimes used referentially, either when the gender of the referent is irrelevant, or, a more recent development, when the referent does not identify as either male or female (Bjorkman, 2017; Wales, 1996). Table 1 summarizes the factors that have been shown to influence the use of epicene pronouns in L1 English. For all pairs of features in the table, singular *they* is more common with the alternative to the right than with the one to the left, whereas the opposite is true in the case of generic *he*.

3.3. L2 English writing

Despite the status of English as a global language, only a limited number of studies have investigated the use of epicene pronouns in L2 English (see Abudaljuh, 2012; Lee, 2006; Stormbom, 2018); to the best of the author's knowledge, no previous research has focused specifically on L2 academic writing. The few studies that exist suggest that, in general, intermediate to advanced L2 writers are less prone to using singular *they* and more prone to using generic *he* than L1 writers. A tendency to use *he* seems to be particularly apparent in some L1 groups, which may be the result of influence from L1 sociolinguistic norms (Abudaljuh, 2012), and from the L1 gender system (Stormbom, 2018). Research has also suggested that, similarly to L1 writers, L2 writers can be affected by the type of antecedent in their choice of epicene pronouns; i.e. they are more likely to use singular *they* when the antecedent is notionally plural and thus low in individuation (see Abudaljuh, 2012; Lee, 2006; Stormbom, 2018).

For L2 learners of any language, a central factor in successful acquisition of L1 features is frequency in L1 input (see e.g. Ellis, 2012). Since singular *they* is now so common in L1 English, it can be assumed that intermediate to advanced L2 learners of English are exposed to singular *they* in various contexts. However, exposure alone does not necessarily mean that L2 writers will use the pronoun to the same extent as L1 writers. As Ellis (2012: 199) notes, L2 users' "computations and inductions are often affected by transfer, with L1-tuned expectations and selective attention". In practice, this can result in learners choosing generic *he* over singular *they* if they have highly gendered L1s with a traditional practice of masculine generics (see Stormbom, 2018). As Pauwels (2011) notes, learners can also transfer strategies for achieving gender-fair language, that is, feminisation versus neutralisation. For L1 Swedish speakers, the focus of the present study, the issue of choosing epicene pronouns in the L1 is very similar to that in English, and the preferred strategy for achieving gender-fair language has generally been neutralisation also in Swedish (see Section 2).

4. Materials and methods

The data were extracted from two corpora of L2 academic writing: The Finland-Swedish BATMAT corpus (Lindgrén, 2016), and the Swedish subset of the *Varieties of English for Specific Purposes dAtabase*, VESPA-SE (Larsson, 2015). The BATMAT corpus totals approximately 2.5 million words and includes 35 Bachelor's (BA) theses and 86 Master's (MA) theses written by students of English Language and Literature at Åbo Akademi University in Turku, Finland. The vast majority of the writers are Finland-Swedes, i.e. they have Swedish as their L1 (or one of their L1s). The thesis texts were produced between 1972 and 2016, and they can be divided into three main topic areas: linguistics, literature, and society. The corpus VESPA-SE, in turn, includes 69 term papers and BA theses written by students of English at Uppsala University and Stockholm University in Sweden. In total, the corpus amounts to some 280,000 words. Similarly to the BATMAT, the majority of these writers are L1 speakers of Swedish. All texts in the corpus VESPA-SE were written in 2014, and they treat topics within the field of linguistics. Table 2 shows the size, the time period, and the topic areas of the two corpora.

The corpus searches were conducted using the concordance function in WordSmith Tools 5.0 (Scott, 2008). Instances of the three most common epicene pronouns in English were extracted from the corpora: (1) generic *he*; (2) coordinations and composites, i.e. *he or she*, *she or he*, *he/she*, *she/he* and *s/he*; and (3) singular *they*. All forms of these pronouns were included in the data collection. A search for the pronoun *she* yielded no instances of generic use, and the use of *she* was therefore not further examined in the study. Corpus examples of the three pronoun types are shown in (1) to (3).

Table 1
Factors influencing the choice of epicene pronouns in L1 English.

	Features	Examples of studies
<i>Level of formality of language</i>	formal/informal	Holmes, 1998, Paterson, 2014
<i>Degree of individuation of the antecedent</i>	definite/indefinite notionally singular/plural	Baranowski, 2002, Gerner, 2000, Newman, 1992, Paterson, 2014
<i>Gender stereotyping of the antecedent</i>	male stereotyped/neutral	Paterson, 2014, Pauwels, 2001

Table 2
Corpora used in the study.

Corpus	No. of texts	No. of words	Time period	Topics
BATMAT	121	ca. 2.5 million	1970s–2010s	Linguistics, literature, society
VESPA-SE	69	ca. 280,000	2010s	Linguistics

- (1) When a Finnish learner of English omits articles in English, *he* is transferring mother-tongue forms, there being no articles in Finnish. (BATMAT)
- (2) Resistance writing is not only the product of a writer's need to express *him- or herself*. (BATMAT)
- (3) In addition, interlanguage is dynamic as *the learner* learns more rules and applies them into *their* system of grammar rules. (VESPA-SE)

The antecedents were defined as: (1) definite NPs, e.g. *the speaker*; (2) indefinite NPs, e.g. *a learner*; (3) quantificational NPs, e.g. *every child*; and (4) indefinite pronouns, e.g. *everyone*. Table 3 summarizes the types of pronouns and the types of antecedents relevant to the study. Two combinations of search words and context words were used to extract pronoun instances from the corpora: (1) indefinite pronouns (search words) + all forms of the epicene pronouns (context words), and (2) determiners (search words; underlined in Table 3) + all forms of the epicene pronouns (context words). Searching for determiners ensured that a larger number of gender-indefinite NPs could be found than if a specific, limited, list of personal nouns had been used. The maximum distance between antecedent (search word) and pronoun (context word) was set to 25 words, as most instances of pronouns are likely to occur within such a limit (see e.g. Gerner, 2000, p. 96, Laitinen, 2007, p. 109). The co-text of each pronoun instance was examined carefully, to ensure manually that no referential uses of the pronouns were included in the final data set.

The analysis of the data was divided into two parts: (1) a diachronic examination of the use of epicene pronouns since the 1970s, and (2) an in-depth investigation of current use of epicene pronouns, starting from 2010. The diachronic analysis focused on the BATMAT corpus as this corpus includes texts from four decades. For the investigation of current use, the BATMAT data from the 2010s were combined with the data from VESPA-SE, to get a broader data set to work with. To avoid the results being skewed by individual texts, condensed tokens were used in the analysis; in other words, multiple pronoun instances with the same antecedent were only counted as one token of the pronoun type (see Newman, 1992, p. 456, Paterson, 2014, p. 53). Example (4) shows multiple instances of generic *he* in reference to the antecedent *a person*.

- (4) When a person is talking *he* is at the same time using *his* language to create *his* identity. (BATMAT)

If a writer used more than one type of pronoun with the same antecedent, they were counted as one instance of each type (see e.g. Paterson, 2014). An example of this kind is given in (5).

- (5) *The prototypical buyer* could be expected to give reasons to why *s/he* should not buy a specific product, and *they* could also be expected to possess a certain degree of power over the seller. (VESPA-SE)

To examine the possible effects of gender-stereotyping in nouns, lists with ratings of nouns were used from four previous studies: Abudaljuh, 2012, Carreiras et al., 1996, Doherty & Conklin, 2017, and Kennison & Trofe, 2003 (see Paterson, 2014 for a similar approach). In these studies, L1 speakers were asked to rate gender-indefinite nouns in English according to whether they are: (1) associated with male individuals, (2) associated with female individuals, or (3) neutral in terms of gender expectations. The L1 studies did not include all nouns that occurred in the present data set; for this reason, clear near-synonyms from the studies were used in a limited number of cases. To give an example, the present data included the noun *individual*, which did not occur in any of the previous studies. This noun was viewed as a near-synonym for *person*, which was rated as gender-neutral in both Abudaljuh (2012) and Kennison and Trofe (2003). The decontextualized L1 English lists were used under the assumption that similar stereotypes exist in the students' L1 Swedish, but the possibility cannot be excluded that there may be socio-cultural differences between the languages; to the best of the author's knowledge, no comparable rating studies exist for the Swedish language. As pointed out above, Swedish authorities have been very supportive of feminist language planning, which Milles (2011) attributes to the high gender-equality in Sweden. Against this background, open-mindedness may be expected as concerns nominal references to traditional gender roles.

5. Results and discussion

5.1. Diachronic analysis

The corpus searches yielded 280 occurrences of epicene pronouns in the BATMAT corpus. In total, 86 of the 121 texts in the corpus included at least one pronoun instance. The three pronoun types were fairly equally distributed in the overall data: generic *he* occurred 103 times (36.8%), *he or she* occurred 90 times (32.1%), and singular *they* occurred 87 times (31.1%). This

Table 3

Types of antecedents and pronouns in the study.

Types of antecedents	Pronouns (lemmatized)
Definite NPs: e.g. <i>the speaker</i> , <i>that human being</i>	Generic <i>HE</i>
Indefinite NPs: e.g. <i>a learner</i> , <i>an individual</i>	Coordinations and composites, i.e. <i>HE OR SHE</i> , <i>SHE OR HE</i> , <i>HE/SHE</i> , <i>SHE/HE</i> , <i>S/HE</i>
Quantificational NPs e.g. <i>any person</i> , <i>each student</i> , <i>every child</i> , <i>no professor</i> , <i>some writer</i>	Singular <i>THEY</i>
Indefinite pronouns <i>anybody</i> , <i>anyone</i> , <i>everyone</i> , <i>everybody</i> , <i>someone</i> , <i>somebody</i> , <i>no one</i> , <i>nobody</i>	

picture changes considerably when the results are viewed diachronically. Table 4 shows the occurrences of epicene pronouns across decades.

Although the number of instances is limited for the earlier decades, a fairly distinct pattern can be discerned from these figures. As can be expected, the proportion of generic *he* decreases noticeably between the 1970s (97.0%) and the 2010s (14.1%), whereas the opposite trend is true for singular *they*, which does not occur at all in the 1970s, but accounts for 56.5% of all instances in the 2010s. With respect to coordinated forms, there seems to be no clear development in either direction, but irrespective of the variation, it can be concluded that these forms constitute a substantial proportion of all pronouns from the 1980s and onwards. The results show that feminist language planning has had an effect since the use of generic *he* decreases as time progresses, whereas *he or she* and *they* are preferred instead. As in English-language communities, there has been an intense debate in Sweden about gender-bias in language in recent decades. It can therefore be assumed that the writers are also familiar with gender-fair language from their L1, for instance in the use of *he or she* variants.

Another factor that proved relevant to the distribution of pronouns was the type of antecedent. Table 5 shows a breakdown of the corpus findings according to the four antecedent types: definite NPs (e.g. *the person*), indefinite NPs (e.g. *a child*), quantificational NPs (e.g. *any student*), and indefinite pronouns (e.g. *everyone*). The results support previous claims that the choice of epicene pronouns can be influenced by the antecedent's degree of individuation, both in L1 and L2 English writing (see e.g. Newman, 1998; Stormbom, 2018). As seen in the table, singular *they* is the most common pronoun in reference to indefinite pronouns (74.6%) and quantificational NPs (90.5%), whereas it is the least common pronoun when the referent is an indefinite NP (14.2%) or a definite NP (11.1%). The other two pronoun types show the opposite pattern; i.e. they occur more frequently with definite NPs and indefinite NPs than with the other two antecedent types. As quantificational NPs and indefinite pronouns tend to have a plural meaning, they are low in individuation and they therefore more readily take *they*, which is a better match than *he* with respect to both semantic number and semantic gender.

In terms of diachronic developments, there is an increase in the use of *they* with highly individuated antecedents in the later decades. As example (6) shows, the only instance of *they* from the 1970s and the 1980s is coreferential with an indefinite pronoun, *anyone*. Similarly, all instances of singular *they* in the 1990s are anaphors of either quantificational NPs or indefinite pronouns, i.e. antecedents that are low in individuation. The first instances of *they* with the other two antecedent types occur in the texts from the 2000s (6 instances, 26.1%), and in the 2010s, 21 out of 52 instances of *they* (40.4%) are coreferential with these types of antecedents. Examples (7) to (8) show *they* in reference to an indefinite NP, *a memoirist*, and a definite NP, *the person*.

- (6) *Anyone who has been discriminated against on racial grounds can take their case to a county court or, if the case is about employment, to an industrial tribunal. (BATMAT)*
- (7) *A memoirist usually chooses one single aspect or theme from their life to write about. (BATMAT)*
- (8) *One could even attempt to persuade the person to change their ways, as long as the behaviour was not prevented by law or other means. (BATMAT)*

This development is likely to have resulted from the increasing visibility of singular *they* in L1 English usage.

The findings further suggest that, overall, gender-stereotyping has a fairly limited effect on the choice of epicene pronouns. Table 6 lists all types of nouns that occur as heads in NP antecedents in the data. The figures in brackets represent the number

Table 4Distribution of epicene pronouns according to decade ($N = 280$).

	70s		80s		90s		00s		10s	
	<i>N</i>	%								
<i>he</i>	32	97.0	17	40.5	25	55.6	16	23.5	13	14.1
<i>he or she</i>	1	3.0	24	57.1	9	20.0	29	42.7	27	29.4
<i>they</i>	0	0	1	2.4	11	24.4	23	33.8	52	56.5
Total	33		42		45		68		92	

Table 5
Distribution of epicene pronouns according to antecedent type in the diachronic data.

	Definite NPs		Indefinite NPs		Quantific. NPs		Indefinite pronouns	
	N	%	N	%	N	%	N	%
<i>he</i>	32	50.8	62	44.0	2	9.5	7	12.7
<i>he or she</i>	24	38.1	59	41.8	0	0	7	12.7
<i>they</i>	7	11.1	20	14.2	19	90.5	41	74.6
Total	63		141		21		55	

of times that a given noun was used with each pronoun type. The nouns are grouped into three categories according to how speakers of English have rated them in previous research: neutral, male-stereotyped, and female-stereotyped (Abudalbh, 2012; Carreiras et al., 1996; Doherty & Conklin, 2017; Kennison & Trofe, 2003).

As can be seen, generic *he* occurs with the widest selection of nouns. This is to be expected since *he* is the most common pronoun with NPs in the data; *he* was used with a total of 96 NP tokens, compared to 83 NPs for *he or she*, and 46 NPs for *they*. The three most common head nouns overall were *person* (43 instances), *learner* (25 instances), and *speaker* (16 instances). Two of these – *person* and *learner* – were rated as neutral in previous research, and it can be expected that *speaker* is also neutral for most language users. It appears that generic *he* was used with the greatest proportion of male-stereotyped nouns, suggesting some influence of gender-expectancy. Apart from the noun *president*, however, none of these nouns occurred more than once in the data, which means that the high prevalence of generic *he* in the overall data is not simply a matter of stereotyping. It can also be concluded that most nouns in the final category (no previous data available) are likely to be neutral in terms of gender expectations, e.g. *reader*, *character*, and *test taker*. One potentially male-stereotyped noun in this group is *hero*, which occurs as an antecedent to both *he* and *he or she*. In general, it can be assumed that academic writing of this kind does not typically manifest a large number of the gender-stereotyped nouns present in the context-free L1 lists.

5.2. Current use: 2010 and onwards

Table 7 shows the distribution of epicene pronouns in VESPA-SE and in the BATMAT texts from the 2010s. The 222 instances of pronouns were found in 71 of the 102 texts: 33 texts in BATMAT and 38 texts in VESPA-SE. Overall, generic *he* had a limited distribution in the corpora, whereas both coordinated forms and singular *they* were used quite extensively. The writers in VESPA-SE appear to be more prone to using coordinated forms than the writers in BATMAT, who instead tend slightly more towards singular *they*. However, the differences in the proportion of the three types of epicene pronouns in the two corpora did not prove to be statistically significant when examined with the Mann-Whitney *U* test (alpha level set at $p < .05$), which takes into account writer variability within a corpus.

As shown in Table 8, the antecedent type proved to be a relevant factor in the choice of pronouns in the data from the 2010s. The figures indicate that singular *they* is the pronoun of choice for writers when the antecedent is an indefinite pronoun or a quantificational NP. Also, singular *they* is by no means uncommon with definite and indefinite NPs in the data from the 2010s, which differs from the diachronic data. However, these types of NPs are also frequently referenced by *he or she*.

Table 9 shows all types of head nouns in the antecedent NPs of the data, as well as the number of times they occur with each kind of pronoun. The nouns are grouped according to how they have been rated in previous research (see Abudalbh, 2012; Carreiras et al., 1996; Doherty & Conklin, 2017; Kennison & Trofe, 2003).

The largest variety of nouns is found with *he or she*, which is also the most common pronoun type with NPs: *he or she* was used 85 times in reference to NPs, compared to 21 times for *he*, and 77 times for *they*. The three most frequent head nouns overall were *person* (31 occurrences), *speaker* (20 occurrences), and *student* (18 occurrences). Both *person* and *student* were rated as neutral in previous research, and it can be assumed that *speaker* is also free from gender expectations for most language users. As seen in the table, the vast majority of the rated nouns were neutral in terms of gender expectancy, and no direct interaction can thus be observed between gender-stereotyping and pronoun choice. The nouns in the final category (no previous data available) also appear to be predominantly neutral in terms of gender expectations.

It cannot be ignored that there was still considerable variation in the use of epicene pronouns in the newer data. This can be seen in Table 8 with respect to definite NPs and indefinite NPs, which are frequent anaphors of both *he or she* and *they*. Variation was also found at the level of the individual texts: A total of 51 out of 71 texts included more than one instance of epicene pronouns, and in 33 of these texts, there were instances of more than one type of pronoun. Put in slightly different terms, more than half of the writers who used multiple anaphoric references were inconsistent in their use of epicene pronouns in the sense that they did not restrict themselves to only one of the three pronoun types. In four cases, the writers even used two pronoun types in reference to one and the same antecedent, as shown in (9) and (10).

- (9) Following an object during a longer period of time would offer a better way of really understanding how much *an individual* can change *his or her* identity according to the situation *they* are in. (BATMAT)

Table 6

Head nouns in NP antecedents in the diachronic data.

	<i>he</i>	<i>he or she</i>	<i>they</i>
Neutral	<i>author</i> (2) <i>child</i> (4) <i>individual</i> (3) <i>learner</i> (14) <i>one-year old</i> (1) <i>person</i> (13) <i>poet</i> (7) <i>student</i> (2) <i>writer</i> (5)	<i>author</i> (1) <i>human being</i> (1) <i>individual</i> (3) <i>learner</i> (8) <i>person</i> (15) <i>pupil</i> (1) <i>researcher</i> (1) <i>student</i> (3) <i>teenager</i> (1) <i>tour group leader</i> (1) <i>translator</i> (1) <i>writer</i> (5)	<i>author</i> (3) <i>child</i> (1) <i>human being</i> (1) <i>individual</i> (4) <i>journalist</i> (1) <i>learner</i> (3) <i>person</i> (15) <i>pupil</i> (1) <i>researcher</i> (1) <i>student</i> (3) <i>writer</i> (2)
Male-stereotyped	9 types 51 tokens <i>judge</i> (1) <i>magic user</i> (1) <i>magician</i> (1) <i>offender</i> (1) <i>officer</i> (1) <i>pilot</i> (1) <i>president</i> (3) <i>sports coach</i> (1) <i>ruler</i> (1)	12 types 41 tokens <i>coach</i> (1)	11 types 35 tokens <i>president</i> (1)
Female-stereotyped	9 types 11 tokens <i>teacher</i> (2)	1 type 1 token <i>teacher</i> (7) <i>victim</i> (1)	1 type 1 token
No previous data available	1 type 2 tokens <i>addressee</i> (1) <i>advertiser</i> (3) <i>Anglo-Indian</i> (1) <i>autobiographer</i> (1) <i>character</i> (4) <i>Finn</i> (1) <i>foreigner</i> (1) <i>hero</i> (2) <i>Marxist</i> (1) <i>nationalist</i> (1) <i>passport holder</i> (1) <i>proletarian</i> (1) <i>reader</i> (4) <i>slave</i> (1) <i>speaker</i> (8) <i>translator</i> (1)	8 tokens <i>American</i> (1) <i>character</i> (1) <i>Finn</i> (1) <i>hero</i> (1) <i>illiterate</i> (1) <i>interlocutor</i> (1) <i>language user</i> (1) <i>linguist</i> (1) <i>native speaker</i> (2) <i>non-native speaker</i> (4) <i>protagonist</i> (1) <i>sociologist</i> (1) <i>speaker</i> (6) <i>Swede</i> (1) <i>reader</i> (7) <i>test taker</i> (1) <i>visitor</i> (2)	<i>character</i> (3) <i>memoirist</i> (1) <i>reader</i> (2) <i>social actor</i> (2) <i>speaker</i> (2)
	16 types 32 tokens	17 types 33 tokens	5 types 10 tokens

- (10) Wright wonders whether complimenting a non-native speaker, assuring *such a person* that *he or she* speaks French well, is actually a way of thanking *them* for *their* efforts. (BATMAT)

The common features of these four cases are that all antecedents are relatively highly individuated NPs, and that the writers shift from using *he/she* or *he or she* to using singular *they*. This switch may be a result of the asserted unwieldiness of coordinations and composites (see e.g. Huddleston & Pullum, 2002). In the rest of the texts with more than one type of pronoun, the writers use different epicene pronouns in reference to different antecedents, as exemplified in (11). In these cases, too, the writers typically vacillate between using *he or she* and *they*.

- (11) How does *a translator* deal with this kind of nominalized adjective when *he or she* translates Swedish texts into English? [...] Author-centered translation is the strategy often used for literature - it means that *the translator* keeps close contact with the author throughout the translating process, and might translate things differently because of the information *they* receive from the author. (VESPA-SE)

Arguably, the variation found in the present data is indicative of a language change that is still very much in progress. The use of generic *he* is clearly not favoured in the more recent corpus data, but there is less consensus on what pronoun should take its place: *he or she* or *they*. The relatively high frequency of coordinated forms may be related to L1 Swedish influence, as

Table 7Distribution of epicene pronouns in the 2010s ($N = 222$).

	BATMAT		VESPA-SE		Overall	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>he</i>	13	14.1	9	6.9	22	9.9
<i>he or she</i>	27	29.4	60	46.2	87	39.2
<i>they</i>	52	56.5	61	46.9	113	50.9
Total	92		130		222	

Table 8

Distribution of epicene pronouns according to antecedent type in the 2010s.

	Definite NPs		Indefinite NPs		Quantific. NPs		Indefinite pronouns	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>he</i>	10	11.5	11	12.9	0	0	1	2.6
<i>he or she</i>	38	43.7	47	55.3	0	0	2	5.1
<i>they</i>	39	44.8	27	31.8	11	100.0	36	92.3
Total	87		85		11		39	

Table 9

Head nouns in antecedent NPs in the 2010s.

	<i>he</i>	<i>he or she</i>	<i>they</i>
Neutral	<i>author</i> (1) <i>learner</i> (2) <i>person</i> (6)	<i>author</i> (2) <i>child</i> (2) <i>contact person</i> (1) <i>human being</i> (1) <i>individual</i> (1) <i>informant</i> (4) <i>learner</i> (9) <i>parent</i> (2) <i>person</i> (10) <i>student</i> (11) <i>subject</i> (1) <i>translator</i> (2) <i>writer</i> (4)	<i>author</i> (4) <i>child</i> (2) <i>customer</i> (1) <i>human being</i> (1) <i>individual</i> (4) <i>journalist</i> (1) <i>learner</i> (6) <i>parent</i> (3) <i>patient</i> (1) <i>person</i> (15) <i>researcher</i> (1) <i>student</i> (7) <i>subject</i> (1) <i>translator</i> (1) <i>writer</i> (2)
	3 types 9 tokens	13 types 50 tokens	15 types 50 tokens
Male-stereotyped	<i>president</i> (3)		<i>assassin</i> (1) <i>president</i> (1)
	1 type 3 tokens		2 types 2 tokens
Female-stereotyped		<i>teacher</i> (2)	
		1 type 2 tokens	
No previous data available	<i>autobiographer</i> (1) <i>character</i> (2) <i>NNS</i> (1) <i>slave</i> (1) <i>speaker</i> (4)	<i>buyer</i> (1) <i>character</i> (1) <i>communicator</i> (2) <i>listener</i> (1) <i>orator</i> (1) <i>personality</i> (1) <i>poster</i> (1) <i>presidential candidate</i> (1) <i>protagonist</i> (1) <i>reader</i> (5) <i>recipient</i> (1) <i>respondent</i> (1) <i>seller</i> (1) <i>speaker</i> (10) <i>test taker</i> (4) <i>user</i> (1)	<i>bilingual</i> (1) <i>buyer</i> (1) <i>character</i> (3) <i>language user</i> (1) <i>listener</i> (1) <i>memoirist</i> (1) <i>native speaker</i> (1) <i>participant</i> (1) <i>poster</i> (2) <i>provider</i> (1) <i>recipient</i> (1) <i>sender</i> (2) <i>social actor</i> (2) <i>speaker</i> (6) <i>user</i> (1)
	5 types 9 tokens	16 types 33 tokens	15 types 25 tokens

well as to the nature of the texts, i.e. L2 academic writing by university students. Recent L1 research has shown that some writers perceive the use of *he or she* as more appropriate than *they* in fairly formal texts, the reason being the apparent ungrammaticality of singular *they* (LaScott, 2016). The power of language norms is obvious here: Although singular *they* is no longer proscribed, the acceptance of singular *they* in grammars and style guides has been slow. Thus, it can be expected that L1 and L2 users alike will strive for a language use that is perceived as standard in academic texts. Having said this, the fact that singular *they* is relatively frequent with all kinds of antecedents in the recent data suggests that *they* is becoming an established pronoun for third-person singular reference among these advanced learners of English, even in the context of academic writing. Tentatively, the versatility of *they* in current use can therefore be taken as support that these L2 users perceive *they* as a singular pronoun in its own right, rather than a manipulation of the third-person plural pronoun (see Paterson's (2014) Homonymy Theory). Given the nature of corpus data, however, a claim of this kind cannot be conclusively confirmed in the present study.

6. Conclusion

This study extends previous research by examining the use of the epicene pronouns *he*, *he or she*, and *they* in L2 English academic writing by Swedish-speaking university students. The diachronic analysis shows that the use of generic *he* has decreased since the 1970s, whereas the use of *he or she* variants and singular *they* has increased. In other words, feminist language planning has clearly had an effect on L1 Swedish speakers' advanced learner writing in English. Since the topic of language and gender has been discussed extensively in Sweden, these writers are likely to be familiar with the use of gender-fair language from their L1.

The findings also indicate that the antecedent's degree of individuation is relevant to the pronoun choice: Generic *he* is used more with antecedents that have a high degree of individuation, i.e. definite and indefinite NPs, whereas singular *they* is more frequent with antecedents that are low in individuation, i.e. quantificational NPs and indefinite pronouns. However, the influence of the antecedent type appears to be less pronounced in the later decades, most likely because singular *they* increases in frequency even with the more individuated types of antecedents. Coordinated forms remain common with definite and indefinite NPs, but they are rarely used with antecedents that are low in individuation.

The analysis of current use shows that there is still noticeable variation in and between the texts from the 2010s, with both coordinations and singular *they* being used extensively. This variation is arguably a sign of a language change that is still in progress. The frequent use of singular *they* in L1 English means that L2 users encounter this pronoun in L1 English input. Yet, writers may be hesitant to use singular *they* in relatively formal writing, such as academic texts, because its use is still not fully accepted in prescriptive grammar and style guides. The present study provides a starting point for further research, for instance of the use of epicene pronouns in L2 academic writing by professionals. As English has become the lingua franca of academia, most professional academic writing in English is now produced by L2 rather than L1 speakers. The question thus arises to what extent L2 users are involved in creating new norms as concerns epicene pronouns, and whether or not the development in the use of these pronouns will be the same as in L1 writing.

References

- Abudalbah, M. (2012). *Ideology, gender roles, and pronominal choice: A sociolinguistic analysis of the use of English third person generic pronouns by native speakers of Arabic*. Lawrence: University of Kansas (Unpublished doctoral dissertation).
- Baker, P. (2010). Will *ms* ever be as frequent as *mr*? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language*, 4(1), 125–149.
- Baranowski, M. (2002). Current usage of the epicene pronoun in written English. *Journal of Sociolinguistics*, 6(3), 378–397.
- Baron, D. (1981). The epicene pronoun: The word that failed. *American Speech*, 56(2), 83–97.
- Bjorkman, B. (2017). Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics*, 2(1), 1–13.
- Cameron, D. (1995). *Verbal hygiene*. London: Routledge.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49(3), 639–663.
- Chambers, J. K., & Schilling, N. (Eds.). (2013). *Handbook of language variation and change*. Oxford: Wiley Blackwell.
- Corbett, G. (1991). *Gender*. Cambridge: CUP.
- Curzan, A. (2014). *Fixing English: Prescriptivism and language history*. Cambridge: CUP.
- Doherty, A., & Conklin, K. (2017). How gender-expectancy affects the processing of 'them'. *The Quarterly Journal of Experimental Psychology*, 70(4), 718–735.
- Ellis, N. (2012). Frequency-based accounts of second language acquisition. In S. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 193–210).
- Fasold, R. (1987). Language policy and change: Sexist language in periodical news media. In P. Lowenberg (Ed.), *Language spread and language policy* (pp. 187–206). Washington: Georgetown University Press.
- Fasold, R., Yamada, H., Robinson, D., & Barish, S. (1990). The language planning effect of newspaper editorial policy: Gender differences in the *Washington Post*. *Language in Society*, 19, 521–539.
- Gerner, J. (2000). Singular and plural anaphors of indefinite personal pronouns in spoken British English. In J. Kirk (Ed.), *Corpora Galore: Analyses and techniques in describing English: Papers from the nineteenth international conference of English language research on computerised corpora (ICAME 1998)* (pp. 93–114). Amsterdam: Rodopi.
- Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals 1965–2004. *Review of General Psychology*, 10(4), 377–389.
- Henley, N., & Abueg, J. (2003). A review and synthesis of research on comprehension of the masculine as a generic form in English. *Estudios de Sociolingüística*, 4(2), 427–454.
- Holmes, J. (1998). Generic pronouns in NZ English. *Kotare: New Zealand Notes and Queries*, 1(1), 32–40.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: CUP.
- Kennison, S., & Trofe, J. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32(3), 355–378.

- Kloss, H. (1969). *Research possibilities on group bilingualism: A report*. Quebec: International Center for Research on Bilingualism.
- Laitinen, M. (2007). *Agreement patterns in English: Diachronic corpus Studies on common- number pronouns*. Helsinki, Finland: Société Néophilologique.
- Larsson, T. (2015). The Swedish subset of the *Varieties of English for specific Purposes dAtabase (VESPA-SE)*. Compiled at Uppsala University, Sweden. VESPA is administered by Dr. Magali Paquot at the Centre for English Corpus Linguistics at Université catholique de Louvain, Belgium <http://www.uclouvain.be/en-cecl-vespa.html>.
- LaScotte, D. (2016). Singular *they*: An empirical study of generic pronoun use. *American Speech*, 91(1), 62–80.
- Lee, J. (2006). Acceptability of sexist language among young people in Hong Kong. *Sex Roles*, 56, 285–295.
- Liddicoat, A. J. (2011). Feminist language planning. *Current Issues in Language Planning*, 12(1), 1–7.
- Lindgrén, S.-A. (2016). *The BATMAT corpus*. Finland: The Department of English Language and Literature at Åbo Akademi University.
- Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: CUP.
- Meyers, M. (1990). Current generic pronoun usage: An empirical study. *American Speech*, 65(3), 228–237.
- Milles, K. (2011). Feminist language planning in Sweden. *Current Issues in Language Planning*, 12(1), 21–33.
- Mitchell, F. (1994). Student writing and pronoun reference: A study of inclusive language in practice. *Women and Language*, 17(2), 2–6.
- Newman, M. (1992). Pronominal disagreements: The stubborn problem of singular epicene antecedents. *Language in Society*, 21(3), 447–475.
- Newman, M. (1998). What can pronouns tell us? A case study of English epicenes. *Studies in Language*, 22(2), 353–389.
- Parini, A. (2012). The use of the epicene pronoun in written English: Variation across genres. In O. Dontcheva-Navratilova, & R. Povolna (Eds.), *Discourse interpretation* (pp. 163–178). Newcastle upon Tyne: Cambridge Scholars.
- Paterson, L. (2011). Epicene pronouns in UK national newspapers: A diachronic study. *ICAME Journal*, 35, 171–184.
- Paterson, L. (2014). *British pronoun use, prescription, and processing. Linguistic and social influences affecting 'they' and 'He'*. Palgrave Macmillan.
- Pauwels, A. (1998). *Women changing language*. London: Longman.
- Pauwels, A. (2000). Generic pronouns in the marketplace or shopping centre. Exploring the impact of non-sexist language policy on everyday speech. In Anna Gunnarsdotter-Grönberg (Ed.), *Sett och hört. Festschrift for Kerstin Nordenstam*. Göteborg (pp. 270–276). Swedish Institute, Göteborg University.
- Pauwels, A. (2001). Non-sexist language reform and generic pronouns in Australian English. *English world-wide: A Journal of Varieties of English*, 22(1), 105–119.
- Pauwels, A. (2011). Planning for a global lingua franca: Challenges for Feminist Language planning in English(es) Around the world. *Current Issues in Language Planning*, 12(1), 9–19.
- SAOL. (2015). *Svenska Akademiens ordlista över svenska språket* (14th ed.). Stockholm: Svenska Akademien.
- Scott, M. (2008). *WordSmith Tools version 5*. Liverpool: Lexical analysis software.
- Stormbom, C. (2018). Epicene pronouns in intermediate to advanced EFL writing. *International Journal of Learner Corpus Research*, 4(1), 1–22.
- Timberlake, A. (1977). Reanalysis and Actualization in syntactic change. In C. N. Li (Ed.), *Mechanisms of syntactic change* (pp. 141–169). Austin: University of Texas Press.
- Wales, K. (1996). *Personal pronouns in present-day English*. Cambridge: CUP.

Charlotte Stormbom is a PhD student in English Language and Literature at Åbo Akademi University in Turku, Finland. Her current research focuses on the use of epicene pronouns in advanced L2 English writing.