Generic personal pronouns in New Zealand newspaper English

Emilia Louhio
University of Tampere
School of Language, Translation and Literary Studies
English Philology
Pro Gradu Thesis
April 2015

_____

Tässä tutkielmassa tarkastellaan geneerisiä persoonapronomineja uudenseelanninenglannissa. Tarkoituksena on tutkia geneeristen pronominien *he, he or she, she* ja *they* käyttöä indefiniitti-pronominien kanssa ja selvittää, mitkä tekijät mahdollisesti vaikuttavat pronominin valintaan sekä onko pronominien käytössä havaittavissa eroja amerikanenglantiin ja brittienglantiin verrattuna.

Englannin kielessä ei ole sukupuolineutraalia yksikön kolmannen persoonan pronominia, mikä tuottaa ongelmia esimerkiksi indefiniittisiin korrelaatteihin viitattaessa. Tutkimuksissa on havaittu, että aiemmin yleinen maskuliinisen pronominin *he* geneerinen käyttö amerikan- ja brittienglannissa on vähentynyt 1960-luvulta lähtien merkittävästi erityisesti feministisen liikkeen vaikutuksesta. Monikollisen pronominin *they* geneerinen käyttö yksiköllisiin korrelaatteihin viitattaessa on näissä varieteeteissa sen sijaan yleistynyt, joskin käytäntö on yhä kiistanalainen ja varsinkin kirjoitetussa kielessä sitä saatetaan pitää kielioppisääntöjen vastaisena.

Tutkimuksen aineistona on käytetty Corpus of New Zealand Newspaper English -korpusta, joka sisältää tekstinäytteitä uusiseelantilaisista sanomalehdistä vuosilta 2010-2012. Uudenseelanninenglanti on aiheen kannalta kiintoisa varieteetti, sillä Uudella-Seelannilla on pitkä historia sukupuolisen tasa-arvon edelläkävijänä ja feministinen liike on myös vaikuttanut maan kielipoliittisiin uudistuksiin. Persoonapronominien käyttöä kirjoitetussa uudenseelanninenglannissa ei ole aiemmin tutkittu, ja ajantasaisen korpusmateriaalin avulla on mahdollista tarkastella pronominien esiintyvyyttä nykyhetken kontekstissa. Sanomalehdet toimivat lisäksi tärkeänä väylänä uusien kielellisten varianttien leviämisessä ja niiden voidaan olettaa heijastelevan vallalla olevia kielipoliittisia suuntauksia.

Tutkimuksen tulokset osoittavat, että uudenseelanninenglannissa indefiniittisten korrelaattien kanssa käytetään useimmiten pronominia *they,* kun taas muiden varianttien, yksikkömuotoisten geneeristen pronominien *he* tai *she* sekä *he or she* -yhdistelmäpronominin käyttö tässä yhteydessä on marginaalista. Myös siteeratussa tekstissä pronominien jakauman havaittiin olevan suurin piirtein sama, joskin variantti *he or she* esiintyi huomattavasti harvemmin siteerauksissa. Muilla tutkituilla lingvistisillä tai ekstralingvistisillä tekijöillä ei sen sijaan näyttänyt olevan suurta vaikutusta pronominien jakaumaan. Pääpiirteiltään tulokset heijastelivat odotetusti muissa englannin kielen varieteeteissa havaittuja trendejä, ja *they*-pronominin käyttö indefiniittipronominien kanssa vaikuttaa olevan uudenseelanninenglannissa jopa hieman yleisempää kuin amerikan- ja brittienglannissa.

Avainsanat: korpuslingvistiikka, uudenseelanninenglanti, persoonapronominit, sukupuolineutraali kieli

Table of Contents

1 Introduction

"God send everyone their heart's desire." This line from Shakespeare's *Much Ado About Nothing* illustrates one of the most profound issues in the English language, that of a gender-neutral singular third person pronoun. Traditionally the masculine pronoun *he* has been used in a generic sense in reference to singular antecedents and has been prescribed as the only grammatically correct generic pronoun since it agrees with the antecedent in number, whereas the clash with gender is often ignored. In the 1960s this generic usage came under attack from feminist linguists, however, who promoted the use of other alternatives instead, such as *he or she* and *they*. In present-day English the plural pronoun *they* is increasingly used in reference to singular antecedents and this use, dubbed by some linguists as "singular *they*", is also recognized by most modern dictionaries and grammars. Not everyone approves of this usage, however, and those with a strict sense of grammar still regard singular *they* as grammatically incorrect. The present study investigates the use of personal pronouns as anaphors of indefinite pronouns. The issue is topical even in present-day English, since the problem of the "correct" epicene pronoun still persists. Recently there has also been increasing awareness of the need to avoid binary gendered language, so the topic can be considered especially timely from this point of view as well.

Research on generic pronoun usage has mainly focused on the American English and British English varieties. The present study focuses on another variety situated in Kachru's (1985) inner circle of Englishes, New Zealand English (henceforth NZE). This variety has received less attention in the literature so far, and generic pronoun usage in written NZE has not yet been covered. Thus the aim of the present study is to fill this research gap. Generic pronoun usage in NZE is investigated by means of corpus linguistic examination and the material for this investigation is drawn from the Corpus of New Zealand Newspaper English. The corpus includes material from the 2010-2012 period and can be considered representative of current NZE usage. In previous research singular *they* was frequently found in American and British contemporary

newspapers (see chapter 2.6) and it is interesting to see if NZE usage conforms with these observations, especially since New Zealand can be considered a progressive country from gender-equal perspective (see 2.5). In this study I investigate the use of generic pronoun variants *they, he or she, he,* and *she* in reference to indefinite pronouns *someone/-body, everyone/-body, anyone/-body*, and *no-one/-body*. I aim to answer the following research questions:

1. What are the frequencies and distribution of generic third-person pronouns with indefinite antecedents in New Zealand English?

2. Which factors or features may influence the choice of pronoun?

3. Do these uses differ from those of American and British English?

The thesis is divided into two sections: the first part covers the linguistic background and the methodology used, and in the second part this theoretical framework is applied to a corpus examination of generic pronoun usage in NZE. The hypothesis is that the frequencies and distribution of the pronouns in NZE display similarities to American and British English. This study hopes to shed light on the present-day status of these pronouns in this particular variety and in the wider context also contribute to our general knowledge of the prevalence of singular *they* and its acceptance in ever wider contexts.

2 Theory and background

2.1 Anaphoric use of personal pronouns

Personal pronouns may be used for anaphoric (backward) or cataphoric (forward) reference. The present study is interested in the former, as indefinite pronouns and personal pronouns most typically have anaphoric relation. Anaphoric expressions can be said to derive their meaning from the surrounding linguistic material (Huddleston and Pullum 2005, 101), consider for example: "Cate said she was going home*"*. Here the anaphorically used personal pronoun *she* derives its meaning from the antecedent *Cate*. Personal pronouns and definite noun phrases are the main anaphorically used expressions in English, and the choice between a personal pronoun or an NP may be affected by, for example, register or anaphoric distance, i.e. the distance between the antecedent and the anaphoric expression (Biber et al. 1999, 237). In the *Longman Grammar of Spoken and Written English* Biber et al. (1999, 239) observe on the basis of their corpus findings that personal pronouns are characteristically used with a shorter anaphoric distance and appear more often in informal than formal contexts. As newspaper language represents a more formal register, these factors may affect the frequencies of generic pronouns in the present study to some extent.

Anaphorically used pronouns are usually expected to agree with their antecedents in terms of person, number and gender. In gender-specific contexts, such as in the example above where the referent's gender is known, the third-person singular pronouns *he*, *she,* and *it* are normally used. In gender-neutral contexts, however, it becomes necessary to choose between gender-specific pronouns even when the gender of the referent is unknown, unspecified, or "mixed", as in the case of indefinite personal pronouns, for example. This conflict of grammar and gender and the proposed solutions will be discussed in more detail below.

2.2 Generic pronouns

Gender can be considered a problematic category for English, since there is no singular gender-

neutral third person pronoun to be used in generic, or non gender-specific, contexts. Consider for example:

Everyone ought to do what _____ can

In sentences such as the example above, the choice of a pronoun has been the focus of "a long-standing conflict between the logic of grammatical 'authorities' and that of native speakers" (Wales 1996, 110). Historically, English had a dual pronoun system of gendered and sex-neutral pronouns, which gradually fell out of use (Baron 1986, 198). From the mid sixteenth century onwards, *they, he or she* and *he* have reportedly been used in reference to singular gender-neutral antecedents, and none of these alternatives were considered improper at the time (Baranowski 2002, 378; Baron 1986, 193). The lack of a common-gender pronoun started to get noted in grammars in the eighteenth century, and later that century prescriptive grammarians declared generic *he* as the only correct pronoun to be used in this context (Bodine 1975, 130-133). The dual pronoun *he or she* was rejected as stylistically awkward and too cumbersome, while singular *they* was criticized for violating number concord, i.e. not agreeing with its singular antecedent, and was thus proscribed as unacceptable and ungrammatical (Bodine 1975, 133). In the late 1960s this proscription was countered by the feminist linguistic movement, and although the discussion was most heated back then, the issue can still be considered topical in present-day English. The problem of the "correct" epicene pronoun has not been solved, and the need for a gender-neutral pronoun still persists today.

Since the 1960s, the generic use of *he* has come under heavy criticism and alternative strategies are now increasingly preferred (Biber et al. 1999, 316). According to Pauwels, who has studied the topic of feminism and language reform extensively, most commonly used strategies for avoiding sexist pronoun usage in English include (1) avoiding singular forms altogether by pluralizing the noun; (2) avoiding pronominal use by repeating the noun or noun phrase; (3) alternating the use of *he* and *she;* (4) using the dual pronoun *he or she*; and (5) using *they* as a singular pronoun (Pauwels 2001, 106). Avoidance strategies have their limitations, however, and

Pauwels argues they should mainly be used in addition to pronoun alternatives (ibid.). As strategies (1) and (2) are difficult to investigate by means of a corpus examination (cf. Adami 2009, 288), the present investigation will only focus on the pronoun alternatives.

One frequently cited example of the prescription of generic *he* is found in *A Dictionary of Modern English Usage* (Fowler 1926, 648) which condemns the use of both singular *they* and *he or she* in reference to indefinite pronouns, dismissing the former as simply a "mistake" and the latter as "cumbersome". For comparison, in the 3rd edition of the book, *The New Fowler's Modern English Usage* by Burchfield (1996, 776-779), it is stated that *they* is now often used with indefinite pronouns and that this usage is "passing unnoticed by standard speakers (except those trained in traditional grammar) and being left unaltered by copy editors". This example can be seen to demonstrate the shift from prescriptive to descriptive approach in grammars and dictionaries, i.e. they describe language as it is actually used today, rather than telling people how they should speak.

Modern grammars now commonly include a chapter on non-sexist pronoun usage, although their views differ to some extent. For example, Biber et al. (1999, 316) in the authoritative *Longman Grammar of Spoken and Written English* advise to use coordinated pronoun forms (*he or she,* etc.) or switch to plural forms instead, but do not recommend using singular *they*. Others display more condoning views of singular *they*, for example Carter and McCarthy in *Cambridge Grammar of English,* a usage guide intended especially for teaching purposes, briefly comment on the issue:

> In traditional formal usage, *he/him* may occur with reference to both sexes. Increasingly, however, gender-neutral pronoun forms are preferred, such as (subject forms) *he/she*, *he or she*, *they*, or (in writing) *(s)he*, or *s/he*; (object forms) *him/her*, *him or her*, *them* (2006, 378).

Huddleston and Pullum (2002, 493-494) in *The Cambridge Grammar of the English Language* discuss the issue and generic alternatives more extensively and they also dedicate an entire paragraph for singular *they,* which they describe as "very common in informal style" and widespread and neutral in style in reference to antecedents such as *everyone*, *someone* and *no one*.

The generic pronoun variants were chosen for the present corpus investigation based on the

discussion in the existing literature. The variants and their forms included in this study are:

1. singular *they: them, their, theirs, themself, themselves*
2. generic *he: him, his, himself*
3. dual pronoun *he or she: him or her, his or her, (s)he, s/he* (conversely *she or he*)
4. generic *she: her, hers, herself*

These variants will be presented in more detail in the following chapters (2.2.1 – 2.2.4), while 2.2.5 will briefly discuss other pronouns that can sometimes be used as generics but that were left out of the present investigation. In previous research, generically used pronouns have also been called *epicene*, *gender-neutral* or even *gender-fair* pronouns, among others. The present study will principally employ the term *generic*, although other terms may be used interchangeably.

2.2.1 Generic *he*

Traditionally the pronoun *he* could be used with reference to both men and women, and it was argued to cover both gender-specific as well as gender-neutral reference. This neutral use has since become under question, however, and the issue has received a great deal of attention in the literature, especially from feminist linguists (see also 2.4). It has been shown in numerous empirical studies (e.g. Martyna 1980) that majority of people interpret the referent as male and do not unproblematically perceive *he* as a truly generic pronoun. Martyna (1980, 72-77), for example, found that her subjects responded to male words by thinking of predominantly male images, and points out that generic *he* was "far from adequate in covering generic ground". Wales (1996, 114) similarly comments that it is difficult to imagine the pronoun being used as a true generic with no male connotations. Other reasons for avoiding generic *he* have also been presented, including the exclusiveness of the expression, and the social injustice evoked by it (Frank and Treichler 1989, 147-153; Martyna 1980, 69-70).

Generic *he* was prescribed by the eighteenth-century grammarians as the only grammatically correct anaphoric pronoun that can be used in reference to singular gender-neutral antecedents, basing their argument on the fact that it agrees with the antecedent in number while they ignored the

clash with gender (Wales 1996, 112; Balhorn 2009, 394). However, the expression is limited even in terms of agreement, and keeping strict concord is problematic in some contexts, such as across clause boundaries. Consider, for instance, this example from American speech, quoted in Neubauer:

> Do you think *anyone* here is left alive? UNIDENTIFIED MALE: No, I don't think *he's* alive. (2008, 154)

Neubauer (ibid.) notes that the use of *he* here seems semantically strange, while others regard similar sentences as ungrammatical (e.g. Baron 1986, 195; Wales 1996, 127). Besides such grammatically compromised constructions, some linguists opposing the masculine generic have also presented ludicrous or bizarre examples of its use, such as this line from a law text (quoted in Meyers 1990, 228): "No person shall be forced to have an abortion against his will." Huddleston and Pullum (2005, 104) in *A Student's Introduction to English Grammar* note that avoiding *he* is not only a matter of political correctness, but the generic use is also problematic in the sense that it "unquestionably colors the interpretation". They also point out that nearly all speakers would consider *he* "impossible" with some coordinated NPs, such as in the example "The husband or the wife will have to give up his seat on the board" (ibid.). Furthermore, the conflict of number is generally considered by speakers to be less important than the conflict of gender. When faced with a choice of either maintaining the genericity of the antecedent or the structural number agreement, native speakers of English have traditionally preferred the logic of meaning over that of form and used alternative terms such as the plural pronoun *they* (Balhorn 2004, 86; 2009, 395; Wales 1996, 113).

Some contemporary dictionaries of English do not comment on the inadequacy of generic *he* in any way. The *Oxford English Dictionary Online* (2013), for example, simply states that the personal pronoun may be used "in anaphoric reference to a singular noun or pronoun of undetermined gender". Learner's dictionaries, on the other hand, usually take a more "educative" approach than general ones and more readily give advise on usage. Many of them now also recognize the controversial nature of this pronoun, the *Cambridge Advanced Learner's Dictionary*

(2008), for example, points out that using generic *he* is now regarded as "slightly old-fashioned" and that many people do not like this use and prefer to use *they* instead. Similarly, in the *Oxford Dictionary of English* (2003) it is pointed out that "until recently *he* was used uncontroversially to refer to a person of unspecified sex", but now this use has "become problematic and is a hallmark of old-fashioned or sexist language".

2.2.2 Singular *they*

The plural *they* may also be used in the sense of 'people in general' which is not to be confused with the generic sense of the pronoun at issue here, i.e. when *they* is used as a third-person singular pronoun in reference to singular antecedents, also dubbed by some linguists as "singular *they*". Using *they* and its inflected forms in reference to a singular pronoun or noun has a long history, and the oldest records of this usage date back to the fourteenth century (Burchfield 1996, 779; Balhorn 2004, 82). Reportedly the use of *they* and *their* was well-established by the early sixteenth century and there is both anecdotal and statistical evidence of the ubiquity of singular *they* in both written and spoken Middle English (Balhorn 2004, 80; Holmes 1998, 31). The pronoun is commonly found in the literature of previous centuries, such as in the works of Austen, Fielding, and Shakespeare (Baron 1986, 193; Balhorn 2004, 80).

Although the early 16th century grammarians, as Bodine (1975, 134) reports, regarded the masculine "worthier" than the feminine, they never proscribed the use of singular *they*. This use only came to be opposed in the late 18th century, when prescriptive grammarians claimed that singular *they* used in reference to singular nouns is grammatically incorrect as it does not agree with its antecedent in number (Bodine 1975, 132-133). This use of *they* has never been completely eradicated from speech, however, and as Wales (1996, 126) points out, the form has been well established in informal usage for centuries. The viability of singular *they,* both in informal and formal registers, has been discussed to great extent in previous literature (e.g. Bodine 1975; Baron

1986; Pauwels 1998; Wales 1996; Balhorn 2004). The main motivation for choosing *they* is that it is not marked for gender, i.e. it conveys a gender-neutral image and has unbiased connotation. Due to its gender-neutralness, singular *they* is especially useful in contexts where the gender of the referent is unknown or unimportant, or in cases where the speaker or author does not wish to assign gender to the referent, even if the referent is known (Wales 1996, 129). The construction has other advantages as well, including simplicity and naturalness of expression, for example (ibid., 126).

Contemporary dictionaries and grammars now commonly include a comment on the use of *they* as an epicene pronoun, and many grammars now also use *they* for gender-neutral reference (e.g. Carter and McCarthy 2006, Huddleston and Pullum 2002). The *Oxford English Dictionary Online* (2013) notes that *they* is used especially "in relation to a noun phrase involving one of the indefinite determiners or pronouns *any*, *each*, *every*, *no*, *some*, *anybody*, *anyone*, etc." but adds that this use "has sometimes been considered erroneous". *Oxford Dictionary of English* (2003), on the other hand, mentions that *they* is "becoming more and more accepted both in speech and in writing" and add that it is now "generally accepted" in reference to indefinite pronouns, although with singular nouns in formal contexts it is still regarded ungrammatical. Biber et al. (1999, 317) take a more critical view and regard the plural pronoun as ungrammatical even when the antecedent is used in reference to all members of a group rather than to a single entity, as with *everyone/-body*, and argue that *they* is least likely to be adopted by academic writing as it is "a register much concerned with correctness". Indeed, Adami's (2009) findings in academic corpora confirm that this view still seems to be prevalent today (see chapter 2.6).

The conflict in agreement is most apparent in the reflexive construction, which has been regarded as a "problem area" for generic *they* (Huddleston and Pullum 2005, 104; Wales 1996, 127). While *themselves* seems more acceptable when used for example with the notionally plural *everybody,* it may sound odd with a singular antecedent such as *someone* (Huddleston and Pullum 2005, 104). For instance, the *Longman Dictionary of Contemporary English* (2009), a dictionary for

advanced learners, points out that "many teachers think this is not correct English". A possible solution would be to use the "singular" form *themself* instead of *themselves*. Far from a new introduction to the language, *themself* in fact preceded *themselves* in standard English and it is reported in the OED to have been used as a standard reflexive form from the 14th century to the 16th century (Burchfield 1996, 777; Wales 1996, 127). Later on the form fell out of use but it re-emerged in the 1980s as a by-product of the search for gender-neutral pronouns (Burchfield 1996, 777). There are somewhat contradicting accounts of the present status of the form. According to Burridge and Mulder (1998, 160), the form is "making something of a comeback" and is commonly used for example in spoken and written Australian English. On the other hand, Huddleston and Pullum (2005, 104) observe that the form is rare and can not yet be regarded as standard. This view is shared by many dictionaries as well, for example in the *Oxford Dictionary of English* (2003) it is noted that the form is not widely accepted in standard English. *Longman Dictionary of Contemporary English* (2009) is cautious about this form as well, commenting that "many people think this use is incorrect", while the *Collins Cobuild Advanced Learner's Dictionary* (2006) advises that *themself* is sometimes used instead of *themselves* when it "clearly refers to a singular object", but similarly adds that "some consider this use to be incorrect". In the corpus investigation I will examine whether *themself* and *themselves* can be considered standard usage in NZE.

Singular *they* has anything but disappeared despite its stigmatization as ungrammatical since the eighteenth century, and although some still consider it incorrect, it is now preferred in spoken English, in non-formal written English, and increasingly even in non-informal written registers, as Wales (1996, 126) points out (see also chapter 2.6). There are a number of academics who believe that *they* will eventually be accepted as the standard gender-neutral pronoun, for example Bodine (1975, 144) compares the adoption process of *they* to that of second person pronouns in the past, and predicts that the third person pronouns may as well be expected to change to "reflect the new ideology and social practices" of the times.

2.2.3 Dual pronoun *he or she*

The dual pronoun *he or she*, also called the coordinated pronoun, is often mentioned in the literature as the most widely accepted generic pronoun, as it agrees with the antecedent in both number and gender and may in this sense be regarded as the most "correct" alternative (Baranowski 2002, 394; Stringer and Hopper 1998, 219; Wales 1996, 121). Baron, for instance, argues it is the only possible alternative if one wishes to maintain strict number and gender concord (1986, 191). Frank and Treichler, on the other hand, point out that the variant is especially useful when it is important to keep the focus on the individual, like in the example: "Every reader responds to this scene in his or her own way" (1989, 158). Usually focus is placed on more informationally significant items, however, and the coordinated form is seen to interfere by drawing attention to itself, especially if the pronoun is used repeatedly or in tag-questions (Wales 1996, 121). It is also criticized as awkward, cumbersome, and clumsy, particularly in the reflexive form *himself or herself* (ibid.; Miller and Swift 1977, 29). Thus the variant is perhaps not the most viable option for spontaneous conversation, but in the written medium, which needs planning and is more concerned with accuracy, it may appear more frequently. Most grammars also seem to agree on this point. Biber et al. (1999, 317) observe that the coordinated forms are more often used by writers and editors who are "aware of the need to avoid gender bias". Huddleston and Pullum (2002, 493) also agree that the form is more common in relatively formal style, and recent studies have reported that the usage of *he or she* has in fact increased during the last decades especially in the more formal registers (e.g. Laitinen 2007, 98; Adami 2009, 293). Pauwels (2003, 564) also interestingly observes that educators and academics display a greater use of *he or she,* while other educated speakers seem to prefer *they.*

The construction *he or she* is, however, criticized by many because it seems to confirm the priority of men (Wales 1996, 121). Some linguists challenge this custom and use *she or he* instead

(e.g. Pauwels 1998) or alternate between the two (Miller and Swift 1977, 29). Contracted forms and blends such as *he/she*, *s/he* and *(s)he* have also emerged as viable alternatives, normally appearing in written texts (Huddleston and Pullum 2002, 493). Quirk et al. (1985, 343) call them experimental forms, but Wales (1996, 120) argues that they are in fact "well established in many written registers". In fact, examples of their usage in the *Oxford English Dictionary Online* (2013) appear to date back to at least the late 19th century. They may not be very common, however, since the disadvantage with these forms is that they are difficult to pronounce and can be considered unsuitable for speech, which reduces their viability to some extent (Pauwels 1998, 132-133). Some find them unsatisfactory, Frank and Treichler (1989, 161-162), for example, recommend avoiding these forms, warning that they do not provide a "quick gender-neutral cure for sexist language" and may even discourage writers from seeking better alternatives. Modern dictionaries that mention these forms are generally less critical, for example *Collins Cobuild Advanced Learners Dictionary* (2006) simply notes that *s/he* is a possibility when "referring to someone who might exist but has not been identified".

2.2.4 Generic *she*

Traditionally the pronoun *she* has been used primarily in contexts where the antecedent was clearly female but in recent times it has come to be used also in a generic sense (Frank and Treichler 1989, 162-163). For example, *Collins Cobuild Advanced Learners Dictionary* (2006) notes it may be used to "refer to a person who is not identified as either male or female". *Oxford Dictionary of English* (2003) comments that *she* may be used in the same way *he* has been used, and point out that it has become common in some types of writing, such as in books on childcare or child psychology", but also notes that in most contexts it is "likely to be distracting" and that other alternatives are preferable. Huddleston and Pullum (2002, 493) note that it represents "a new and very much minority usage" and state that its purpose is to introduce "linguistic discrimination in favour of

females to counterbalance the effects of the long tradition of linguistic discrimination in favour of males". For example, some feminist linguists (e.g. Cameron 1985) use it politically to challenge the convention of using generic *he*. Replacing *he* with *she* is problematic from the linguistic equality perspective, however, and can be criticized of simply reversing the bias by making one sex visible at the expense of another (Pauwels 1998, 130). Although the reason for favouring generic *she* is well-founded, it is not usually recommended and, for example, Wales (1996, 123) points out that generic *she* is not usually mentioned as an option in style-sheets or handbooks because of its strong ideological connotation. In writing, alternating between *he* and *she* has been suggested as a possible strategy, although if this method is employed it is important to avoid displaying stereotyped usages or confusing or distracting the reader (Wales 1996, 122; Frank and Treichler 1989, 137).

2.2.5 Other alternatives

Other pronouns mentioned as generic alternatives include *it* and *one*. *It* is sometimes used in reference to babies or children, but it is primarily associated with inanimate nouns and this may, according to Pauwels, "hinder its extension to human referents" (1998, 134). *One*, on the other hand, has fairly limited use as a generic pronoun. It is restricted mainly to formal contexts, usually considered viable only in reference to the nouns *one* and *everyone* (Pauwels 1998, 133; Quirk et al. 1985, 388), and, according to Wales (1996, 120), its "egocentric connotation" rules it out as a useful epicene solution. Furthermore, according to Burridge and Mulder (1998, 160) most New Zealand speakers regard *one* as an indefinite pronoun rather than a personal pronoun and forms such as *one's* and *oneself* are rarely heard in this variety.

The most appropriate solution according to Pauwels (2001, 106) would be to invent a completely new pronoun. Numerous neologisms have in fact been proposed throughout the last few centuries, with the earliest suggestions dated around 1850 (Baron 1981, 87). The most widely publicized epicene pronoun, as Baron (1986, 200-201) notes, is the word *thon,* a combination of *the*

+ *one,* which was coined in 1884 and provoked discussion and further word creation at the time, and reportedly was still adopted by some linguists in the 20th century (cf. Wales 1996, 119). More recent gender-neutral pronoun solutions include *ze* and *e* (with the forms *hir* and *hirself*, and *em, er, and erself*, respectively) which may be preferred by persons who identify themselves as genderqueer or transgender, for instance. A preliminary investigation of the present corpus material showed no instances of these pronouns, however. The suggested pronoun alternatives do not appear to very widely adopted as of yet, and speakers perhaps reject neologisms as too "alien" and turn to existing alternatives instead (cf. Baron 1986, 220; Wales 1996, 120). Since the alternatives listed in this chapter do not seem to be very viable in reference to indefinite pronouns, they will not be inspected further in the present study.

2.3 Indefinite personal pronouns

Indefinite pronouns, also sometimes called quantifier pronouns, can be used in reference to an indefinite person or thing, and here the focus is on personal reference. The two main types of indefinite pronouns are compound pronouns, such as *someone*, and *of*-pronouns, such as *some of.* The compound pronouns can be divided into four groups:

> the *some* group: *someone, somebody*
> the *every* group: *everyone, everybody*
> the *any* group: *anyone, anybody*
> the *no* group: *no one/no-one, nobody*

The notional plurality of these pronouns is often noted in grammars. Quirk et al. (1985, 378) in *A Comprehensive Grammar of the English Language* point out that when indefinites appear as subjects, they normally take the singular verb form, though notionally they may denote more than one person. Similarly, Carter and McCarthy (2006) note that if personal pronouns are used to refer to these words, plural forms are commonly used, for example: *somebody lost their wallet.*

The indefinite pronouns are divided into four classes based on their meaning: universal, assertive, non-assertive, and negative (Quirk et al. 1985, 376). The non-assertive *anyone/-body* may

refer to an individual or to a group of people, while the assertive *someone/-body* normally refers only to one individual (ibid.). The assertive and non-assertive pronouns, also sometimes called existential indefinites, can be used generically or specifically. *Everyone/-body* is considered to have universal meaning and denotes more than one person even though morphologically it appears singular (ibid.; Wales 1996, 128). Similarly, the negative pronoun *no one/-body* may be used for individual or collective reference, and the pronoun can be regarded as the negative equivalent of the universal indefinite (Quirk et al. 1985, 376; Wales 1996, 128). It has been noted (e.g. Neubauer 2008, 146; Wales 1996, 128) that the universal and negative pronouns are most easily interpreted by speakers as notionally plural, and this quality may motivate speakers to use plural personal pronouns in reference to them.

The *-one* and *-body* forms are generally thought to be equivalent in function and meaning, with no significant differences in meaning between the two sets. Quirk et al. (1985, 378), however, see slight differences in their usage and argue that *-one* compounds are regarded more elegant than *-body* compounds, and Carter and McCarthy (2006) also note that *somebody, anybody* and *nobody* are used more frequently in informal contexts. Similarly, Biber et al. (1999, 353) report on the basis of their corpus findings that *-body* forms are regarded more casual in style and are slightly more common in conversation, while *-one* forms are preferred in writing. This preference has also been observed elsewhere, in Laitinen (2007) for example, who also suggests that *-one* forms have "more individualised" reference and likely appear with singular anaphors more often than the *-body* compounds do (see also chapter 2.6).

In this study I decided to focus on indefinite personal pronouns as they form a well-defined set and are gender-neutral morphologically and in meaning, denoting neither male nor female subjects. Indefinites and anaphoric pronouns have also been covered to some extent in previous research (see 2.6) and thus lend themselves to comparison with other varieties. The hypothesis is that *they* is chosen more often for the notionally plural antecedents *everyone/-body* and *no-one/-*

body than the notionally singular antecedents *someone/-body* and *anyone/-body*. It is also interesting to see if the frequencies of epicene pronouns differ in reference to *-body* and *-one* forms.

2.4 Feminist linguistic reform

The question of language and gender, especially the linguistic treatment and representation of the sexes, has been an important issue for the feminist movement, and feminists have also played a major role in the introduction and promotion of gender-neutral pronoun alternatives (Frank 1987, 118; Pauwels 1998, 17). Many feminist activists now subscribe to an interactionist view of language and reality, which is based on the idea that language shapes and reflects social reality (Pauwels 2003, 554). This and many other views adopted by feminist linguists have their roots in the Sapir-Whorf hypothesis, either in one of the more radical versions that language determines thinking, or in the more liberal approaches, which are now generally more accepted (ibid.; Mucchi-Faina 2005, 190). Especially those subscribing to liberal views have directed attention towards eliminating gender bias and inequalities from language through, for example, language planning and reform (Pauwels 1998, 97). According to Pauwels, the main motivations for feminist language change include

> 1. a desire to expose the sexist nature of the current language system;
> 2. a desire to create a language which can express reality from a woman's perspective;
> 3. a desire to amend the current language system to achieve a symmetrical and equitable representation of men and women. (2003, 555)

Exposing and eliminating sexist language has attracted a considerable amount of attention from feminist linguists, as gender inequity in language is a means by which a society can maintain its sexist attitudes. Sexist language is argued to manifest itself in a variety of ways, such as in the vocabulary and in the use of masculine generics, which contribute to the androcentric world-view depicting men as the norm (McConnell-Ginet 1989, 35; Penfield 1987, xiii). When male is seen as the norm, Miller and Swift (1977, 34) argue, it becomes natural to think of women as a subordinate class and from there to considering women as "a minority or special interest group". Some linguists

even see a direct link between women's subordinate status in society and the androcentrism in language, although most would not go so far as to say that sexual inequality exists because of language use, or that language causes oppression rather than being a symptom of it (Cameron 1985, 75; Frank 1989, 108-109; Pauwels 1998, xii). Sexism in language is far from trivial, however, especially judging by the vast amount of criticism directed at feminist language planners (see Henley 1987, 5-11; Pauwels 1998, 66-70 for overviews). Critics of language change have been observed to claim that sexist language does not exist, or that "attempting linguistic change is impossible because language is too deeply ingrained, slow to change and shaped by forces other than social movements" (Henley 1987, 9-11). However, similar linguistic change has already occurred, for example, in the swift change from the prevalence of *negro* to *black* in the late 1960s, in the use of *none* with a plural verb, and in the rapid spread and increasing adoption of the title *Ms* towards the 21st century (Frank & Treichler 1989, 142; Henley 1987, 9; Pauwels 2003, 565-566). Furthermore, Bodine (1975, 141) points out that in the wider context of language change, "pronominal systems are particularly susceptible to alteration in response to social change". This was the case with second person pronouns, for example, which changed in the fourteenth century to reflect the new ideology and social practices of the time (Baron 1986, 193).

Feminist language planning is essentially concerned with the promotion of non-sexist practices, alternatives, and forms as replacements for sexist ones (Pauwels 1998, 10). The attempts at achieving linguistic equality have involved proposing amendments to the forms, rules, and uses of language, and the approaches have been diverse. The alterations were first focused on linguistic aspects such as the pronouns and vocabulary, as especially pronouns are seen to have an important subliminal influence on people's perception of the sexes (Cameron 1985, 84; Pauwels 2003, 556; Penfield 1987, xv). Thus it logically follows that one of the main targets of language change has been the prescription of the masculine generic pronoun. Feminist linguists strongly advise against using this pronoun in generic contexts and suggest using various alternative strategies instead, most

notably those of pronoun avoidance and pronoun replacement (Pauwels 1998, 127) (see also chapter 2.2). The usefulness of non-sexist language policies in attempting language change has also been questioned, however, and some are sceptical as to whether linguistic equality can even be achieved. Henley (1987, 6), for example, argues that male supremacy "is fundamentally encoded in our masculist language and will be defended vehemently by those who will fight to retain male privilege", whereas Cameron (1985, 88-90) points out that non-sexist language merely replaces overtly offensive terms with covert ones, and instead she favors practising positive discrimination through positive language, such as using feminine terms for generic reference. Supporters of non-sexist language, on the other hand, argue that care should be taken to avoid a bias in favor of either sex (Frank and Treichler 1989, 16). Regardless of these pessimistic views, feminism and gender-egalitarianism have "no doubt" had an impact on the representation of gender in present-day English and research has shown that feminists' efforts have greatly increased awareness of gender bias in many English-language communities (Cameron 2006, 739; Pauwels 2003, 561). Pauwels observes that a growing number of people display behaviour that points to greater awareness, such as self-correcting generic *he* constructions or commenting about title use and gender stereotypes (2003, 561-562). Multiple studies have also reported a steady decrease in the use of masculine generic nouns and pronouns over the years, and this trend now seems to be widespread in both written and spoken registers as well as in different varieties of English (see chapter 2.6). Changes tend to be slow, however, and Cameron (2006, 739) points out that old usages are not instantly replaced by new innovations, but rather they coexist and oblige speakers to make choices which may result in new patterns of variation.

Another, growing field closely related to feminist linguistics is queer linguistics, which aims to reconceptualise the dominant discourse of gender and sexuality, targeting especially heteronormativity and gender binarism (Motschenbacher 2011, 150-151). One of the major objectives for queer linguistics is the study of how non-binary gender identities are constructed and

performed through use of personal pronouns. People who identify as non-binary gendered may, for example, prefer to use gender-neutral *they* instead of gender-specific pronouns to refer to themselves, and recently there has also been growing awareness and increasing recognition of non-binary gender identities in the media. Generic pronoun usage can thus be considered a timely topic from this perspective as well.

As the present study investigates newspaper language, it is important to note in this context that the mass media also plays a significant role in language reform. The print and electronic media, in particular, are believed to be key agencies in influencing and regulating the language behaviour of speech communities (Pauwels 1998, 14). Informing people about alternative terms is a crucial factor in their adoption and although the media acts as a gatekeeper in institutionalising new terms in the official and public domains, it also doubles as a powerful tool for disseminating new words and meanings (Cameron 1985, 82). Publications such as newspapers and magazines also provide a necessary means for facilitating and spreading change through speech communities (Pauwels 2003, 560; Penfield 1987, xv). Thus for feminist advocates an important and widespread strategy for language change has been to press those working in the media, such as editors and journalists, to adopt language guidelines and recommendations (Pauwels 1998, 14; 144). A more detailed discussion of the guidelines in the New Zealand context is presented in chapter 2.5.2.

2.5 New Zealand English

In this chapter I will first briefly discuss the origins and the general features of present-day New Zealand English, and then turn the focus to the issue of language and gender in New Zealand. The purpose here is to examine how the features of NZE or the language situation in New Zealand today might affect generic pronoun usage. Finally, factors bearing on New Zealand newspaper style will be discussed.

2.5.1 The origins and features of New Zealand English

In Kachru's (1985) three concentric circles of English, an influential model of World Englishes, New Zealand is placed within the inner circle of countries where English is the primary language. New Zealand is in fact regarded in the literature as an unusually monolingual country, as English is practically the only language spoken in the public domain and around 98 per cent of the population can speak it (Bell and Kuiper 2000, 12-13; Hay et al. 2008, 11). New Zealand English is said to have developed in a so-called linguistic melting pot, in which different varieties of English blended together and formed a new homogeneous language, which was still distinct from other existing varieties (Burridge and Mulder 1998, 277). Due to its similarities with Cockney and Australian English, it has been suggested that New Zealand English originated from either of these varieties, but now the widely held belief is that NZE developed separately from other Englishes and it is today regarded as its own independent language variety (Hay et al. 2008, 86).

Before the arrival of European settlers in the eighteenth century the only language spoken in New Zealand had been Maori, but after New Zealand became a colony of Great Britain in 1840 the population of English-speaking immigrants started to rise rapidly and soon outnumbered the native speakers of Maori (Bauer & Warren 2008, 39; Gordon et al. 2004, 39). During the gold rushes in 1850-1900 many more immigrants from Britain, Ireland, and Australia arrived in New Zealand and brought different varieties of English with them, greatly influencing the formation of New Zealand English (Gordon et al. 2004, 77). The populations of these immigrants according to place of origin are presented in Table 1.

Table 1: Birthplace of immigrants in 1881, adapted from Gordon et al. (2004, 45)

	Number	**%**
England	119,224	45.0
Scotland	52,753	19.9
Ireland	49,363	18.6
Australian colonies	17,277	6.5

The English settlers mostly dominated the Wellington, New Plymouth and Canterbury provinces,

while the Scots were settled in the Southland and Otago areas and the Irish populated Westland (Gordon et al. 2004, 51). Australian influence, then, was strongest in Auckland and in the gold mining areas of Westland, Otago, and Southland (ibid. 61). Thus many of the linguistic features of NZE have their roots in pre-settlement British, Irish, and Australian usage. Maori, the second official language of New Zealand, has also been influential particularly for the NZE lexicon (Gordon et al. 2004, 70). In recent decades American influence has been strong in New Zealand as well. The cultural links between the two countries have been greatly reinforced by the import of American culture in New Zealand, such as films and TV programs, which have left their mark most importantly on the NZE vocabulary (Burridge and Mulder 1998, 281; Hay et al. 2008, 75).

Research on New Zealand English has for the most part focused on the phonological and lexical features, as NZE has been said to differ most obviously from other varieties in these aspects (e.g. Hundt et al. 2008, 305), while syntactic phenomena have been studied less frequently. Few studies have so far investigated anaphoric pronoun usage in New Zealand English (see chapter 2.6), and to my knowledge none of them have focused on newspaper language. The present study thus aims to fill this research gap. New Zealand English may show similarities with the varieties discussed above and this should be borne in mind when analysing the findings.

2.5.2 Language and gender in New Zealand

New Zealand adopted non-sexist language guidelines in the 1970s following the feminist language reform, and the linguistic equality approach was especially strong and influential in the country during the 1980s and the early 1990s (Pauwels 1998, 148;173; Pauwels & Winter 2005, 108). Many public and government agencies as well as private corporations have adopted non-sexist language guidelines and policies, for example, all major educational publishing houses have issued guidelines for using non-discriminatory language (Pauwels 1998, 148). According to Pauwels (1998, 162) the dominant strategy promoted in New Zealand language guidelines is that of gender

neutralisation, i.e. minimising or neutralising the gender reference in generic contexts and in occupational nouns, for example. These guidelines are only recommendations, however, and their purpose is first and foremost to raise awareness and to provide writers with non-sexist alternatives. Still, according to Pauwels (1998, 167), these guidelines have become a "major vehicle for implementing non-sexist language reform in the context of public, official, and written forms of communication". Especially news media has been a major target for language reform (see also chapter 2.4), and thus these recommendations can be expected to have an effect on newspaper language as well.

Gender-neutral pronoun usage in NZE is briefly mentioned in Hay et al. (2008, 58-59), who claim that the rates of singular *they* are "very high" in New Zealand. Deverson and Kennedy (2005) also comment on the generic use of *they* in the *New Zealand Oxford Dictionary*, a reference book prepared especially for New Zealand users. They state that the construction is "particularly useful when the sex of the person is unspecified or unknown and the writer wishes to avoid the accusation of sexism that can arise from the use of *he*" and they also note that although some people may still consider *they* incorrect, it is now commonly used instead of *he or she* in spoken English and increasingly in written English as well (ibid.). In the global context, New Zealand appears to be following closely behind North America and Australia which, according to Pauwels and Winter (2005, 124), are trailblazers in the change towards gender-neutral generic pronoun alternatives.

New Zealand can be said to be a progressive country from gender-equal perspective. New Zealand was a site of early female suffrage success and the first country to grant women the right to vote (Grimshaw 2001, 32). New Zealand has worked towards creating a gender-equal environment in a wide range of disciplines and for many years it has also been a global leader in media gender equity (Comrie and McMillan 2013, 197). However, in recent years the progress has in some respects come to a halt and reports have emerged that there appears to be a continuing gender imbalance in the New Zealand media (Comrie and McMillan 2013, 197). The Global Media

Monitoring Project (GMMP), the world's largest study on gender and media, released its latest report on New Zealand media in 2010, in which also four Fairfax Media newspapers (i.e. *Dominion Post, Manawatu Standard, The Press* and *Waikato Times*) were monitored. The report shows that women are under-represented in the print media, particularly in the categories of sports, celebrity stories, and politics (Comrie and McMillan 2012). Other recent studies (e.g. French 2013) have also reported that the gender disparity is especially apparent in the sports department. Women rarely appear as the authors or topics of the news articles in this section, and according to French (2013, 46), "male domination still exists in the everyday print media coverage of sports in New Zealand". Although the results of the present study can in general be expected to display gender-neutral pronoun usage, the observed imbalance in print media could affect the results to some extent, particularly since sports is a regular section in all of the newspapers and thus forms a substantial part of the data.

2.5.3 New Zealand newspaper style

A number of factors may affect the features of written corpus material, which is subject to editorial interventions and publisher's guidelines, for example. Although the newspapers under investigation are all owned by Fairfax Media, they may not follow the same guidelines. Unlike in the US, where there is the Associated Press Stylebook that is widely used by media personnel as a writing and editing reference, there are no specific style guides or guidelines for the New Zealand press to follow. According to librarian Lesley Longstaff of Fairfax Media (personal correspondence February 21, 2014), who is closely acquainted with Fairfax journalists, each newspaper has their own style preferences, for example *The Dominion Post* uses honorifics while *Sunday Star Times* does not. She notes that at some point in the past a quick introductory course on grammar was held for journalists during their training, but then points out that these have now largely been forgotten as the times have changed. A recent change in the publication strategy, dating a couple of years back,

has meant that nowadays it is important to publish articles as fast as possible and there is great emphasis on putting the news online immediately without much editing. This development has meant that there are no longer editors dedicated to each publication, but newspapers now use "hubs" of sub-editors, who may move from one publication to another. This often results in articles being copied across different newspapers, and thus the different styles may be transferred and become mixed in the process. This may be reflected in my corpus results as well, with no great variation in pronoun use between the different newspapers.

2.6 Findings in previous research

The majority of the previous studies on the topic of generic pronouns have taken a feminist linguistics point of view, which will be discussed in more detail below, while others have followed a historic linguistics approach, for instance (e.g. Balhorn 2004, Laitinen 2007). A number of studies have focused on generic pronouns in reference to indefinite antecedents as well, mainly in British and American English. In this chapter I will discuss the findings in previous literature that bear significance to the present study, the emphasis being on recent corpus-based research on anaphoric expressions.

As early linguists did not have electronic corpora at their disposal, they relied on other methods to acquire empirical evidence of pronoun usage. Martyna (1980), for example, used elicited data, while Meyers (1990) examined American university students' writing. The early studies typically investigated the mental images produced by the epicene pronouns, focusing on disproving the neutralness of generic *he*. Baron (1981, 1986) and Wales (1996), then, discuss various epicene pronouns extensively and present examples of their usage, but no exact frequencies are given.

Corpus-based studies on the subject have most often focused on spoken English. Gerner (2000) investigated the use of generic pronouns in relation to indefinite personal pronouns using the

10-million-word spoken sub-corpus of the *British National Corpus* (BNC). He found that singular *they* after indefinite pronouns can be regarded as "the norm" in spoken English and that singular anaphors were extremely infrequent with indefinite pronouns. He also notes that the coordinated pronoun *he or she* was surprisingly rare, as there was only one instance (0.1%) recorded in the corpus (ibid., 96-112). Pauwels (2001), then, studied Australian English, a variety that has historically been very close to New Zealand English, and compared the use of gender-neutral pronouns in two corpora of spoken Australian English. She found that the use of singular *he* had drastically decreased from the 1960s to the 1990s, whereas the use of *they* had increased considerably. Other generic pronouns still remained marginal, however, and similarly to Gerner (2000), Pauwels notes that *he or she* was almost completely absent in the data and points out that its use had not increased dramatically in the post-reform period despite heavy promotion of this alternative through non-sexist language guidelines (2001, 112). Furthermore, Strahan (2008, 19) also observes that using *they* with a singular antecedent is not considered by Australian speakers in any way unusual, while earlier on Stringer and Hopper (1998, 217) also reported a similar tendency in American speakers, who regarded singular *they* as "unmarked and unremarkable". In a more recent study, Neubauer (2008) examined transcripts of spoken American English from the 2000-2007 period and found that the plural pronoun was the predominant choice with indefinite pronouns *any, every* and *some* groups, while *he or she,* on the other hand, was less frequently used than singular pronouns with all three antecedents (2008, 153).

Similar trends have been observed in written English, and across different varieties as well. Laitinen (2007) studied the distribution of generic *he* and singular *they* using the written part of the BNC. He found that writers heavily favoured plural epicenes over masculine singulars in reference to indefinite pronouns, but similarly to Gerner (2000), Laitinen reports that epicene *he* is still used in more literate registers and suggests that within these registers the change towards epicene *they* seems to still be in progress (2007, 121). As for variation between the two pronouns, Laitinen shows

that the indefinite pronoun antecedents clearly affect the choice of an epicene pronoun: *they* appeared more often with universal and negative indefinites than with assertive and non-assertive indefinites (ibid.). Furthermore, he also found differences in the distribution of *-body* and *-one* type nouns: while *they* was the preferred pronoun with both types, it was slightly more frequently used with *-body* pronouns, whereas *he* appeared more often with *-one* forms than with *-body* forms (ibid., 113-115). Although the findings in Gerner (2000) and Laitinen (2007) offer useful background information for the present study, it should be noted that the data in the BNC that they used was collected in the 1990s and may not be entirely representative of present-day situation.

More recent reports perhaps provide more viable comparison points, considering the timely nature of the topic. Baranowski (2002), for instance, compared one American and one British newspaper using self-compiled corpora and found that in both varieties singular *they* was the preferred epicene pronoun in reference to indefinite pronouns. The findings reflect, according to Baranowski, the view of "most modern usage guides suggesting that singular *they* is the only reasonable choice of pronoun for an antecedent specified by *some*, *every*, *no*, *any* and *each*" (2002, 386). In contrast, generic *he* was rarely used, and Baranowski suggests it is now likely to appear only in very formal contexts (ibid.). He also compared the two varieties and found that American writers were more likely to use *he or she* than British authors (ibid. 394). Similarly, Balhorn (2009) studied generic pronoun use in five contemporary US newspapers and reported that *they* was the most often used generic in reference to indefinite pronouns. Furthermore, he found that *everyone/-body* coreferred with *they* significantly more often than *anyone/-body* and *someone/-body* did (2009, 397). These findings are especially of interest since the present investigation also focuses on newspaper language.

In contrast, Adami's (2009) findings seem to confirm Baranowski's (2002) suggestion that *he* is still prevalent in more formal contexts. She investigated the use of generic pronouns in several academic corpora and found that although generic *he* still appeared to be the preferred choice for

singular generic reference, its use had decreased substantially from the 1960s (Adami 2009, 291-292). Interestingly, she suggests that the "gap" left by the decrease of *he* was in British English filled by the increased use of *he or she*, while Americans may prefer other strategies such as pronoun avoidance and pluralisation (ibid.). Use of singular *they* had not increased significantly, however, and the pronoun was equally infrequent in both varieties which, according to Adami (2009, 292-295), seems to suggest that the proscription of *they* still affects academic writers and editors. Although the present data may not show exactly the same preferences as in the formal academic writings, Adami's observations still offer an interesting additional perspective to the topic and show that the use of singular *they* is not without controversies even today.

Thus far, only a handful of studies on generic pronouns have focused on New Zealand English. For example, Holmes has studied this variety extensively and has also looked at generic pronouns with indefinite antecedents in the Wellington Corpus of Spoken New Zealand English (Holmes 1998). Her findings indicate that in NZE, as seems to be the case with other varieties discussed above, *they* is used far more often than other generic pronouns, and that the conservative "pseudo-generic" *he* is only found infrequently. Furthermore, she also points out that New Zealand seems to be ahead of Australia in "avoiding the use of pseudo-generic *man*" (ibid., 32). Thus the change towards more gender-inclusive language observed in other varieties seems to be under way in New Zealand as well, at least in spoken discourse. The present study investigates whether the same trends can be observed in present-day written New Zealand English.

The influence of contextual factors such as the register or speaker-related variables may also play a meaningful role in generic pronoun use. Besides the influence of register, extralinguistic features are not often covered in corpus studies, however, perhaps due to the fact that many corpora lack the necessary information about writers and/or speakers (cf. Lindquist 2009, 150), while sometimes the frequencies turn out to be simply too low to draw any major conclusions, as some linguists have noted (e.g. Pauwels 2001). In Gerner (2000), Laitinen (2007), and Balhorn (2009) the

influence of sociolinguistic features is covered to some extent. All three found differences in pronoun usage between men and women: women were more likely to use epicene *they* and significantly less likely to use singular *he* than men. Gerner (2000, 95) suggests that the acceptance of *they* is "highly sensitive to contextual factors, like the medium and the level of formality". However, contrary to expectations he found that the register did not seem to significantly affect the frequency of *they* (2000, 111). Laitinen also interestingly points out that the social characteristics of writers or speakers would appear to override other external factors, such as register (2007, 268). Confirming these findings need be left for future studies, however, as unfortunately an in-depth investigation of the influence of such extralinguistic factors is beyond the scope of the present study.

All in all, on the basis of these previous findings the use of singular *they* in reference to singular gender-neutral antecedents now seems to be accepted in spoken language as well as in the less formal written registers. In contrast to the spreading of *they*, generic *he* seems to have been reduced to a marginal position in all but academic registers. The dual pronoun *he or she* is mostly found in the more formal contexts. Generic *she*, on the other hand, was recorded in some studies (e.g. Gerner 2000) but only rarely. The focus has most often been on the American and British varieties, but the same trends have also been observed in Australian English, a variety that is quite similar to New Zealand English. Pronoun usage in NZE has not yet been as widely covered, however, and the existing literature for this variety has concentrated on spoken language. The present study can be seen to fill a research gap in this sense, while previous research on other varieties and especially on newspaper language provide a basis for comparing the findings. I expect to find similar frequencies in the CNZNE as in the previous studies, especially considering the increasing acceptance of singular *they*.

3 Corpus linguistics

This chapter discusses the theoretical background of corpus linguistics, the main methodology used in this study. Present-day corpus linguistics makes use of electronic corpora which are collections of texts made available in computer-readable form for purposes of linguistic analysis (Lindquist 2009, 1; Meyer 2002, xii; McEnery and Wilson 2001, 29). Corpus linguistics is considered by many to be a methodology rather than a branch of linguistics, offering a set of methods for investigating a large number of different research questions across various fields of linguistics (Lüdeling and Kytö 2008, v; Meyer 2002, xii; McEnery and Wilson 2001, 2). Present-day studies on language use are generally expected to present empirical evidence to support their claims, and by using corpora linguists can make observations based on actual examples of spoken or written texts instead of relying on their own intuitions (Biber et al. 1998, 9; McEnery and Wilson 2001, 103; Meyer 2002, 102). Indeed, reliability is one of the advantages of corpus linguistics since corpora offer authentic data on linguistic constructions as well as accurate statistical information on how frequently these constructions occur, and the results of a corpus study can also be replicated and verified by others (Lindquist 2009, 5; Lüdeling and Kytö 2008, ix; Meyer 2002, 102).

Linguistic studies on variation may take a corpus-based or a corpus-driven approach. In corpus-based approach linguistic items are pre-selected on the basis of traditional linguistic criteria and their usage is analysed based on how they appear in a corpus, whereas in a corpus-driven approach the patterns emerge from the corpus data itself (Biber 2009, 276). This present study takes a corpus-based rather than a corpus-driven approach. The purpose of such research is to assess the extent to which a linguistic pattern or a construction is found, and to analyse the contextual factors that may influence variability (Biber et al. 1998, 3). One of the major issues to consider before conducting a corpus-based study is defining the key terms and concepts of analysis, which may be words or other units usually derived from existing linguistic theory (Biber et al. 2007, 11; Meyer 2002, 107). The next step is then to choose a corpus that is most suitable for investigating these

linguistic phenomena. The size of a corpus is often a decisive factor: bigger, "balanced" corpora are generally regarded as more representative of a language than specialised corpora, ideally including multiple texts collected from various sources and covering many different registers (McEnery and Wilson 2001, 103; Meyer 2002, xiii; Reppen et al. 2002, vii). Depending on the variable or linguistic unit under investigation, smaller, specialised corpora may also be sufficient, such as when focusing on one language variety or register as in the present study (Meyer 2002, xiii). Another issue to consider is that some corpora also provide extralinguistic information on contextual factors, such as the age or gender of the writer or speaker, which may in turn affect linguistic variability (Lüdeling and Kytö 2008, ix; Reppen et al. 2002, vii).

Modern corpus linguistics makes great use of computers and analytical software. By means of automatically analysing and processing data, electronic corpora allow carrying out large-scale language investigations that would have been impossible to conduct using traditional methods. Corpora also make the task of data-gathering faster and less laborious, and they offer more material immediately available for research (Lindquist 2009, 5). This is vital especially for investigations where large amounts of data are needed to make sure that the results are representative of the language community at large and that the conclusions are not based on a few speakers' idiosyncrasies (Reppen et al. 2002, viii). However, larger corpora also present issues for the collecting and handling of data. When dealing with automated text analysis in a corpus study it is crucial to aim for optimal precision and recall: precision refers to the number of relevant hits as opposed to irrelevant hits retrieved in the search, while recall is a measure of how many of the relevant hits in the corpus are actually retrieved in the search (Lindquist 2009, 44). While poor precision can often be dealt with by manual analysis, recall is a more problematic factor as it is often difficult to know if something has been missed in an automatic corpus search (Ball 1994, 295). This may in turn affect the accuracy and reliability of the results, and therefore it is advisable that a corpus investigation employs manual methods in addition to automated techniques (ibid.).

It is often emphasized in the literature that corpus-based studies should employ both qualitative as well as quantitative methods, i.e. use statistical counts or examples to test a linguistic hypothesis. Classifying and counting linguistic features forms the quantitative basis of a corpus investigation, and presenting quantitative findings is also important for the accountability of the linguistic analysis. However, corpus linguists have sometimes been criticized of regarding the gathering of evidence and presenting statistical information as their primary goal, thus producing "trivial" results (Biber 2012, 11; Lindquist 2009, 25; Meyer 2002, 102). Therefore great emphasis should be placed on the qualitative analysis as well, which essentially involves making observations on the basis of the quantitative findings and providing explanations for the patterns that emerge from the data (Biber et al. 1998, 9; Meyer 2002, 124). This way linguists can provide descriptive information about the corpus results that cannot be presented strictly quantitatively and account for any surprising frequencies that may emerge from the data (Lindquist 2009, 25; Meyer 2002, 124). The present investigation thus combines both quantitative methods, which provide statistically reliable and generalisable results and allow comparisons with other corpora to be made, and qualitative methods, which provide greater richness and precision (Lindquist 2009, 25; McEnery and Wilson 2001, 76-77; Meyer 2002, 102).

When findings from different corpora or sub-corpora are compared, the total word counts in the corpora need to be taken into account. If the text collections differ in length, simply comparing the raw frequencies will not produce accurate results (Biber et al. 1998, 263). In order to compare, for example, the findings retrieved from a 1-million-word corpus to those from a 5-million-word corpus, the results first need to be normalised. Biber et al. (1998, 263) stress the importance of normalisation and also provide a formula for norming frequency counts: if the figures are normalised per 1 million words, for example, the raw frequencies should first be divided by the total number of words in the corpus or sub-section, and then multiplied by 1,000,000. It is considered common practice to normalise towards a figure close to the size of the smallest corpus, though with

larger corpora it is usually better to use a higher figure (Meyer 2002, 126; Lindquist 2009, 42). In

the present study frequencies are normalised towards 1 million words.

4 Material and methods

This chapter will first introduce the corpus material used in this study, after which the methodological grounds for gathering the data and the pruning of the results will be discussed in detail.

4.1 The Corpus of New Zealand Newspaper English

Primary data for this thesis is drawn from The Corpus of New Zealand Newspaper English (CNZNE). The corpus was released in 2013 and it was compiled by Paul Rickman from the School of Language, Translation and Literature Studies of the University of Tampere. The corpus was chosen for this study since it was readily available, offers current data and is large enough for investigating the topic at hand.

The CNZNE is divided into two sections: the first section covers the 1995-1998 period and the second section covers the 2010-2012 period. The two sections combined include 100 million words in total. As the entire corpus would have been too large to use in the present study, I decided to use only the second sub-section since it provides more recent data and is thus more suitable for studying time-sensitive topics.

The 2010-2012 sub-corpus includes material from 10 different newspapers and consists of one-year or two-year samples of each paper. The corpus is fairly large with 58.4 million words in total and can be regarded as a good representative sample of New Zealand English. Due to its large size the corpus seemed suitable for this study, while it still produced a manageable amount of data. The up-to-date corpus samples also provide reliable data for studying NZE as it is used today. Specifically focusing on newspaper language, on the other hand, conveniently allows comparison with previous studies on American and British English (see chapter 2.6). Furthermore, as Garrett and Bell (1998, 3) note "media usage influences and represents people's use of and attitudes towards language in a speech community". Thus newspapers can be expected to display

contemporary language use in New Zealand.

The breakdown of the newspapers, periods sampled and the word counts in the CNZNE are presented in a table below.

Table 2: New Zealand Newspapers (Rickman, forthcoming)

Magazine	**Sample period**	**Size** (million words)
The Dominion Post	Jan 2011-Dec 2011	10.1
Manawatu Standard	Jan 2011-Dec 2011	5.0
Nelson Mail	Jan 2010-Dec 2010	4.7
Press	Jan 2012-Dec 2012	11.0
Southland Times	Jan 2012-Dec 2012	5.9
Sunday News (tabloid)	Jan 2011-Dec 2011	3.4
Sunday Star Times	Jan 2012-Dec 2012	3.6
Taranaki Daily News	Jan 2010-Dec 2010	4.8
Timaru Herald	Jan 2011-Dec 2011	4.1
Waikato Times	Jan 2010-Dec 2010	5.8

The sections vary greatly in size and therefore, as was mentioned in chapter 3, the corpus results will be normalised per one million words in order to allow comparing the frequencies between the different newspapers.

It is also worthwhile to investigate whether other attributes such as the publishing region and circulation size, presented in the table below, may affect the distribution of pronouns.

Table 3: Newspaper region and circulation (News Works NZ 2014)

	Region	**Circulation**
The Dominion Post	Metropolitan (Wellington)	70,211
Manawatu Standard	North	12,754
Nelson Mail	South	11,997
Press	Metropolitan (Christchurch)	65,379
Southland Times	South	23,742
Sunday News	North	29,603
Sunday Star Times	North	118,465
Taranaki Daily News	Central	19,215
Timaru Herald	South	10,970
Waikato Times	North	27,740

Although the modern trend of sharing articles between different publications may limit the variation between the newspapers to some extent (see chapter 2.5.3), it will nevertheless be interesting to see if they display any differences in pronoun use. According to a report by World Trade Press (2010), some of the papers have conservative political leaning, i.e. *Dominion Post, The Press* and *Sunday Star Time*s, which are also the three magazines with the widest circulations. These newspapers might prefer more "grammatically correct" forms such as *he or she* over *they,* for example. As this information came from a commercial company, it is considered here as a useful additional perspective rather than a crucial insight, however.

The composition of CNZNE follows that of the newspaper sub-section in the British National Corpus and the corpus material features the broadsheet newspaper sub-genres arts, commerce, editorial, miscellaneous, report, science, social and sports, and also a small percentage of tabloid material (Rickman, forthcoming). Although studying variation in the different sections would be useful, it was not possible to analyse these categories separately in this study, as the corpus did not allow genre-specific searches at the time of writing. Furthermore, the original newspaper material, provided by the Fairfax Archives, includes only very basic information, such as the title of the article and the author's name if available (Rickman, forthcoming). Therefore it is

unfortunately not possible in the present study to investigate the influence of sociolinguistic features such as writer gender or age, since this type of data was not available in the original material. Due to these limitations I will not be commenting extensively on the extralinguistic features that may or may not affect the choice of pronouns. The corpus examples may occasionally include such information and will be mentioned where appropriate, but these are too infrequent in number to allow a proper quantitative analysis.

4.2 Methods

The primary methodology used in this investigation is that of corpus linguistics (see chapter 3) and both quantitative and qualitative methods are employed in this study. In this chapter the methodological issues and choices behind the corpus investigation are discussed, and the pruning of the results will be explained in detail. The identification tags assigned to the corpus examples (shown in brackets) are based on the text files in which they appear, consisting of the name of the newspaper, date of publication and file number.

The primary data was retrieved from the CNZNE using the Antconc concordance software tool. The objective of the corpus search was to find words occurring near each other, i.e. instances of indefinite personal pronouns *someone/-body*, *everyone/-body*, *anyone/-body* and *no-one/nobody* followed by a generic pronoun in the subjective, objective, possessive, or reflexive form, e.g. *someone forgot their umbrella*. There are various possible approaches for doing this, of which two were considered here: searching for context words and using regular expressions. A regular expression is a string, or a pattern of characters, that is used to match sequences of characters in a text (Baker et al. 2006, 138). Regular expressions can also include special characters, or wildcards, and match more than one string of words at a time (ibid.). The advantage of this method is that it is often easier to define a regular expression search that matches a set of words, instead of searching for them individually (ibid.). On the other hand, context, or collocate, search is perhaps a more

approachable strategy for finding words that occur near each other and these searches are usually carried out with the help of a concordancing program (ibid. 37; 43). A preliminary search was done to test which of the two approaches was more suitable for the present study. Curiously, the results showed that the regular expression query returned a greater number of hits than the context search and, more importantly, included relevant instances that for some reason were not found in the context search. The regular expression search seemed to produce more accurate results and was thus deemed the most suitable approach for this present study.

The regular expression was formulated so that it finds (ideally) all "sets" of an indefinite pronoun followed by a personal pronoun within a certain number of words. The complete regular expression is given below:

\b(word|word)(?:\W+\w+){0,16}?\W+(word|word|word)\b

The *-one/-body* forms of the indefinite pronouns were placed between the first set of parentheses, while the second set featured the personal pronouns. Since the searches appeared to produce a large number of hits, the results were initially grouped according to grammatical function to make the manual analysis more organized. Thus one possible search string would be:

\b(anyone|anybody)(?:\W+\w+){0,16}?\W+(they|he|she)\b

The string was then modified to find all possible pronoun combinations. Conveniently, not all forms needed to be spelled out as, for example *he* would also find all instances of *he or she, s/he,* and so on, with the exception of the negative paradigm which included the indefinite pronoun written as two lexical units, i.e. *no one/no-one/nobody.*

The numbers in the curly brackets are used to determine the number of w-units, or words, allowed to appear between the search terms. This limit was set at 0 words minimum and 16 words maximum. In some previous studies that employed this method, the maximum was set higher (e.g. Laitinen 2007), while in others it was set lower (e.g. Balhorn 2004). Although a wider context could in theory produce more hits in total, it was uncertain if the number of relevant hits would increase

accordingly. A comparison test was thus conducted to investigate this matter by extending the limit to 20 words. Surprisingly, the results indicated that the additional relevant hits obtained this way would have been marginal, approximately 1 in 10 ratio. While the manual work would have increased considerably, the percentage of relevant hits would in fact have decreased. Thus the present study would not have benefited in any significant way from setting the word limit higher and therefore the 16 word limit seemed justified.

Some studies (e.g. Gerner 2000) have employed a slightly different method of searching inside <s> -elements, or sentences. This method has both advantages and disadvantages compared to the near operator (i.e. the w-unit method) used in this study: while the method used here may produce less hits overall, both relevant and irrelevant, restricting the search to <s> -units would exclude all instances that span across more than one sentence, like in this example:

> (1) 'Double Tap' is what mobsters do when they put *somebody* down. One bullet in the heart, one in the head. That way *they* stay down. (southland_times_5_10_2012_72)

This restriction would potentially exclude a significant amount of relevant hits. Naturally I wanted to include these instances in this study to improve the recall, while at the same time precision needed to be reasonably high as well, and thus the near operator was deemed more suitable. Finding all relevant tokens in the corpus would, however, be an extremely laborious task and beyond the scope of this study, and therefore it needs to be emphasized that the present data is only a representative sample, if a large one at that, of all possible instances.

The automatic search produced 15,740 corpus hits altogether. However, this data still included a large amount of hits irrelevant to this study. These "false hits" typically contain no anaphoric relationship between the indefinite pronoun and the following personal pronoun. In (2), for example, the possessive pronoun *his* does not refer to *someone*.

> (2) Seemed like everywhere he turned *someone* slammed a door in *his* face. (dominion_post_20_5_2011_132)

Since no simple automatic method was available for distinguishing these false hits, the most reliable

method was to analyse the tokens manually (cf. Laitinen 2010, Gerner 2000, Neubauer 2008). The ratio of relevant hits was ca. 3 out of 10 instances, which means the precision was not ideal, but the amount of manual work was still reasonable. Recall can be assumed to be satisfactory as well, as similar ratios were reported in earlier studies (cf. Gerner 2000, Neubauer 2008).

The problem of duplicated material in the corpus has been addressed and taken care of during the compilation stage of the corpus:

> All 2010-12 data from publications sharing the same sample year [was run] through plagiarism detection software. This resulted in the detection of around 13,000 articles that were more or less copies of other articles. These were removed from the 2010-12 section, reducing the word count to 58.4 million words. (Rickman, forthcoming)

Despite these efforts, the preliminary corpus search still produced quite many hits that appeared to be identical. The precision of the present corpus findings is not affected by this, however, since the automatically returned results needed to be edited manually in any case, and therefore any duplicates were detected and removed in the process. Although in other circumstances duplicates might count as authentic instances, I chose to exclude completely identical examples as they do not contribute any valuable additional information to this study, and thus only unique hits are included in the data.

During the manual editing phase an unexpected problem occurred when it became apparent that all pronouns starting with a capital letter were missing from the corpus results. It turned out that the search query was accidentally formulated so that it would only match lowercase words. Another search was then carried out to retrieve the missing tokens, and considering that this search produced around 670 additional relevant instances, amounting to nearly 20% of all tokens, this incidence certainly highlights the importance of constructing the search string correctly.

It should be noted that not all of the different forms of pronouns were counted, but rather each set of quantifier + personal pronoun.

> (3) […] we have made it very clear to *anyone* attending the afterball that if *they* are under 18 and wish to drink *they* need a parent present with *them* at all times. (timaru_herald_15_6_2011_95)

As in example 3, only the first instance of *they* was counted and any further forms of the same pronoun were left out, as counting all forms would have been unnecessarily complex and since the purpose here is to study primarily the division of the different personal pronouns and only secondarily the possible influence of the different word forms. However, in the rare case that the indefinite pronoun was followed by another anaphoric pronoun, for example if *they* and *his or her* both referred to the same antecedent, each pronoun (but not each form) was counted.

Besides the corpus instances that were clearly irrelevant, as in example 2 above, there were also cases where a pronoun was used anaphorically but did not have a generic meaning. The reference needed to be non-specific in order for the instances to be counted as relevant data, i.e. the reference should not be to any specific individual. Especially in the case of s*omeone,* there were several hits where the referent would in fact appear to be specific, and instances of this type were removed:

> (4) I knew *someone* who tried it and *he* said it was disgusting. (daily_news_22_1_2010_34)

In some cases it was somewhat difficult to determine if the reference was specific or not, for example:

> (5) The actions he took on the morning were consistent with *someone* who is aware *he* has made a mistake. (daily_news_2_6_2010_44)

In this particular instance I argue that even though *someone* here seems to refer back to *he*, it is in fact used in a generic sense and does not directly refer to the actions of the person in question, but to something that is "consistent with" those actions, and in this case the example was counted towards relevant corpus data.

Similarly, it was sometimes problematic to decide whether *they* was used in a plural or singular sense, for example when the pronoun followed *everyone/-body*:

> (6) *Everyone*'s feeling a bit different but *they*'re all pretty devastated. (nelson_mail_26_4_2010_60)

In sentences such as this it had to be individually decided whether *they* had specific or non-specific

reference, i.e. if it could be interpreted as referring to a specific group of people. In cases where *they* was followed by *all,* as in example 6, the reference often appeared to be specific. Following Laitinen's example (2007, 110) I have attempted to remove any such instances from the data. However, in the present study an exception was made in cases where, mainly with *someone/-body*, the plural pronoun was used even when the referent was in fact known, if the writer (or speaker) has chosen to not reveal information about the referent's gender and thus the reference remains ambiguous to the addressee (7), or *they* can be seen to have a generic rather than individuated interpretation.

> (7) *Somebody* Facebooked me the other day when *they* heard I was leaving. (dominion_post_17_9_2011_98)

While instances that spanned across sentence boundaries were generally counted towards relevant data, examples including reported speech could not always be reliably interpreted as relevant hits. Any instances where the coreference could be considered ambiguous were left out, for example when the anaphoric pronoun could refer to an antecedent other than the indefinite pronoun:

> (8) In normal circumstances, she would go up to *someone* in the street and just ask. "Usually, *they* are cool with it. --" (nelson_mail_17_3_2010_52)

5 Corpus results

This chapter introduces the frequencies and the distribution of the epicene pronouns in the assertive (*someone/somebody*), universal (*everyone/everybody*), non-assertive (*anyone/anybody*) and negative (*no one/nobody*) paradigms, in the order of most frequent to least frequent. Both raw numbers and normalised figures per 1 million words are included in the results and the frequencies for each newspaper will also be given (see chapter 4.1 for the number of words in the whole corpus and in each sub-section). Reflexive forms are counted separately, since it was felt that they deserve closer inspection, and the results are presented in chapter 5.5.

The automatic searches produced a total of 15,740 unedited corpus hits, but as was described in chapter 4.2, this raw data was then subjected to manual analysis. Out of these raw hits, 3,692 were genuine instances of anaphorically used pronouns, which can be considered a sufficient amount of data within the scope of this study. Altogether the relevant corpus hits accounted for 24% of the automatically retrieved data:

Table 4: Actual corpus hits (% of unedited data)

	someone/body	*everyone/body*	*anyone/body*	*no-one/body*	ALL
Retrieved hits	5137	4746	3559	2298	15740
Actual hits	**1307 (25%)**	**1246 (26%)**	**815 (23%)**	**324 (14%)**	**3692 (24%)**

The figures appear to be similar across the indefinite pronouns, except for *no-one/no one/nobody* which had the least amount of tokens with only 1.4 relevant hits out of 10. These numbers are consistent with previous studies however (e.g. Gerner 2000), and would not seem to indicate a problem with recall. Table 5 shows the total frequencies of the anaphoric pronouns, and their distribution among the four indefinite pronouns is presented in Figure 1 below (the subjective forms represent all forms).

Table 5: Total frequencies of anaphoric pronouns, N/1,000,000 (raw frequencies)

they	*them*	*their(s)*	*he*	*him*	*his*	*he or she*	*him or her*	*his or her*	*she*	*her*
25.53 (1491)	6.51 (380)	29.55 (1726)	0.45 (26)	0.12 (7)	0.36 (21)	0.22 (13)	0.02 (1)	0.38 (22)	0.02 (1)	0.02 (1)
61.64 (3600) (97.5%)			**0.92 (54)** (1.5%)			**0.62 (36)** (1.0%)			**0.03 (2)** (0%)	

Figure 1: Distribution of anaphoric pronouns, percentages
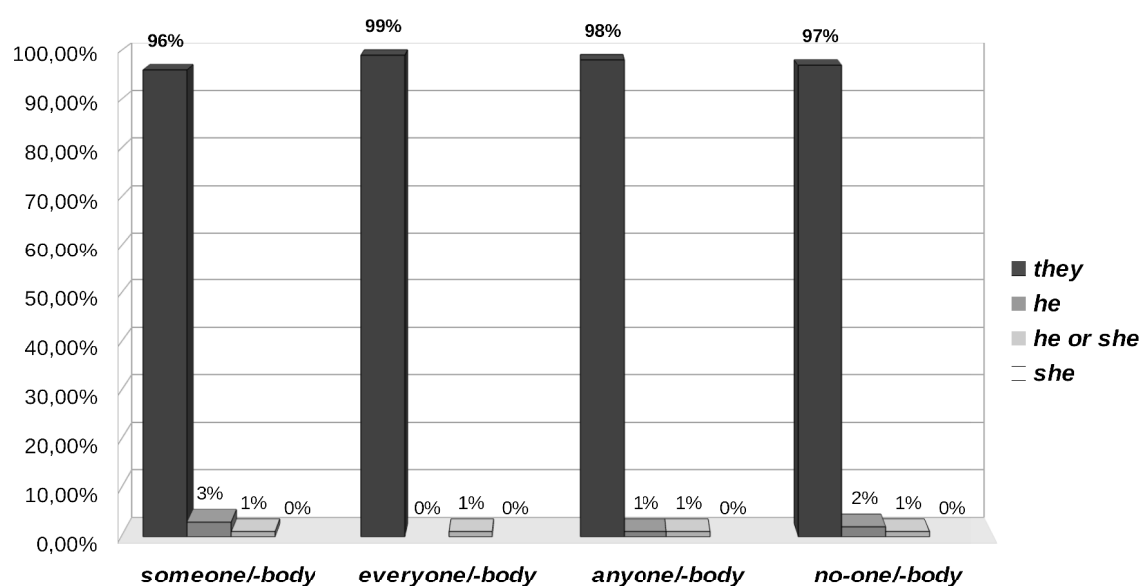
Regardless of grammatical function, the plural pronoun *they* was clearly the most frequent epicene pronoun in the data, whereas singular generic pronouns were relatively infrequent and amounted to only 2.5% of the data. The plural and coordinated pronouns most often appeared in the possessive case, while for the masculine generic the subjective and possessive forms were more even. Objective forms, on the other hand, were much less frequent overall and only accounted for 10.5% of all data. Generic *she* appeared only 2 times in the corpus, once in the subjective and once in the objective form (both are discussed in more detail in 6.5). The distribution of pronouns across the four paradigms was also very similar: as could be expected, the plural pronoun was the most popular pronoun in reference to the indefinite pronouns, and there was very little difference in

variation between the paradigms.

5.1 *Someone/somebody*

In the assertive paradigm a total of 1,307 relevant hits was retrieved from the corpus. The frequencies found in each newspaper are presented in Table 6. For the sake of clarity, only the subjective forms of the anaphoric pronouns will be displayed in this table, and the frequencies for each grammatical category can be found in a separate table below.

Table 6: Frequencies of anaphoric pronouns with someone/-body, N/1,000,000 (raw figures)

	they	*he*	*he or she*	*she/her*
Dominion Post	28.71 (290)	0.79 (8)	0.69 (7)	-
Manawatu Standard	19.40 (97)	0.60 (3)	0.2 (1)	-
Nelson Mail	22.13 (104)	0.43 (2)	-	-
Press	21.55 (237)	0.27 (3)	0.27 (3)	0.09 (1)
Southland Times	14.58 (86)	0.51 (3)	-	-
Sunday News (tabloid)	30.59 (104)	0.29 (1)	-	-
Sunday Star Times	17.78 (64)	0.56 (2)	0.28 (1)	-
Taranaki Daily News	20.00 (96)	1.04 (5)	0.42 (2)	-
Timaru Herald	16.83 (69)	0.98 (4)	0.49 (2)	-
Waikato Times	18.79 (109)	0.34 (2)	0.17 (1)	-
ALL	**21.51 (1256)** (96%)	**0.56 (33)** (3%)	**0.29 (17)** (1%)	**0.02 (1)** (0%)

Table 7: Pronouns with someone/-body per grammatical form, raw figures

they	*them*	*their(s)*	*he*	*him*	*his*	*he or she*	*him or her*	*his or her*	*she*	*her*
558	238	460	17	5	11	9	0	8	0	1

The plural pronoun retrieved the highest number of corpus hits and it appeared most often in the subjective form, with 558 hits in total.

> (9) If *someone* can't get a job, *they* ought not have children until *they*'re employed (dominion_post_9_12_2011_75)

(10) The best jobs occupy the sweet spot between what *someone* likes and what *they* are good at, she says. (the_press_6_9_2012_102)

(11) If *someone* was not familiar with the road and a speed advisory sign was missing, *they* would probably come off the road. (southland_times_11_12_2012_68)

Second most often the pronoun was found in the possessive form with 460 hits, while in the objective form it appeared 238 times.

(12) Could you imagine *someone* doing that at, say, the TSB, setting up *their* own little booth just inside the door. (daily_news_4_8_2010_66)

(13) I'd hate to think that we would never pick *somebody* because of *their* misdemeanours in the past. (sunday_news_18_4_2010_117)

(14) Squeezing *somebody* into a piece of equipment that's not big enough is embarrassing for *them*. (sunday_news_9_10_2011_103)

(15) If *someone* screws up, let *them* know about it. (the_press_9_11_2012_151)

In total, 51 tokens (4%) featured a personal pronoun other than *they*. Generic *he* accounted for 33 of them, appearing in the subjective form 17 times, in the possessive form 11 times, and 5 times in the objective form.

(16) If *someone* gets sick or does not look well *he* is expected to stay away from everyone else. (timaru_herald_6_1_2011_39)

(17) Leaseholders all know stories about *someone* who has had to walk away from *his* home because of huge increases in leases […] (daily_news_27_11_2010_58)

(18) When *somebody* tells you nothing is impossible, ask *him* to dribble a football. (the_press_6_10_2012_110)

The coordinated pronoun *he or she* appeared 17 times in total. It was found in the subjective form 9 times and 8 times in the possessive form, but no instances of the objective form were found.

(19) Today if *someone's* great grandfather is, or was, part Maori *he or she* can claim to be Maori. (timaru_herald_26_2_2011_56)

(20) If *someone* carries out a copycat suicide, *s/he* was on the brink of it, and it's delusional to think that banning media discussion of the topic is necessary to save someone's life. (dominion_post_28_5_2011_115)

(21) How responsive are you when *someone* else reveals *his or her* inner self? (dominion_post_30_11_2011_143)

Interestingly, one instance of generic *she* was also found in this paradigm:

> (22) If you spilled your drink on *someone* else's silk blouse, however, apologise immediately, do what you can to help *her* clean up [...] (the_press_5_12_2012_34)

This example will be discussed in more detail in chapter 6.5.

5.2 *Everyone/everybody*

In the universal paradigm a total of 1,246 hits was found in the corpus. The frequencies of generic pronouns are shown in the tables below.

Table 8: Frequencies of anaphoric pronouns with everyone/-body, N/1,000,000 (raw figures)

	they	he	he or she	she
Dominion Post	22.87 (231)	0.10 (1)	0.20 (2)	-
Manawatu Standard	18.20 (91)	0.20 (1)	0.20 (1)	-
Nelson Mail	23.19 (109)	-	-	-
Press	18.64 (205)	0.18 (2)	0.18 (2)	-
Southland Times	21.02 (124)	-	-	-
Sunday News (tabloid)	27.06 (92)	-	-	-
Sunday Star Times	23.06 (83)	-	0.27 (1)	-
Taranaki Daily News	19.58 (94)	-	0.21 (1)	-
Timaru Herald	20.00 (82)	-	0.24 (1)	-
Waikato Times	21.03 (122)	-	0.17 (1)	-
ALL	**21.11 (1233)** (99%)	**0.07 (4)** (0%)	**0.15 (9)** (1%)	**0**

Table 9: Pronouns with everyone/-body per grammatical form, raw figures

they	them	their(s)	he	him	his	he or she	him or her	his or her	she	her
401	73	759	0	1	3	3	0	6	0	0

Nearly all of the tokens with *everyone/-body* occurred with the plural pronoun. The pronoun appeared most often in the possessive form with 759 hits in total.

> (23) We want *everyone* to know how much we appreciate *their* thoughts and *their*

kindness. (dominion_post_6_10_2011_65)

(24) It was all harmless fun and *everyone* left with a smile on *their* face. (manawatu_standard_17_9_2011_73)

Second most often the pronoun was found in the subjective form with 401 hits, while the objective form appeared 73 times.

(25) But *everybody* just used what *they* had to get ahead. (dominion_post_18_10_2011_42)

(26) *Everybody* has some creativity inside *them*. (nelson_mail_6_10_2010_34)

Pronouns other than *they* were less abundant here than in the assertive paradigm; only 13 tokens (1%) in total were found. The masculine generic occurred 4 times; 3 times in the possessive form and once in the objective:

(27) *Everyone* is as God has made *him*, and oftentimes a great deal worse. (the_press_2_6_2012_115)

(28) In medieval Europe, where *everybody* from warlords to peasants was on *his* own when it came to defending *his* property, *his* rights and *his* "honour", the murder rates were astronomically high: 110 people per 100,000 in 14th-century Oxford, for example. (tagged_the_press_19_12_2012_59)

The coordinated pronoun *he or she* was found in 9 tokens in total; 6 times in the possessive form and 3 times in the subjective form, making the universal paradigm the only one where this variant was more frequent than the masculine generic.

(29) Of course, not *everyone* being put forward by *his or her* government is a good candidate. (timaru_herald_31_5_2011_60)

(30) As Cantabrians know better than most, when disaster strikes, *everyone* must do what *he or she* can, and the vulnerable must be protected. (the_press_23_1_2012_70)

No examples of the feminine generic were found with *everyone/everybody*.

5.3 *Anyone/anybody*

In the non-assertive paradigm, a total of 815 hits was found in the corpus. The frequencies are presented in the tables below.

Table 10: Frequencies of anaphoric pronouns with anyone/-body, N/1,000,000 (raw figures)

	they	*he*	*he or she*	*she*
Dominion Post	13.86 (140)	-	0.10 (1)	-
Manawatu Standard	15.00 (75)	-	0.40 (2)	-
Nelson Mail	11.70 (55)	-	-	-
Press	12.18 (134)	0.45 (5)	0.18 (2)	-
Southland Times	15.08 (91)	0.17 (1)	0.17 (1)	-
Sunday News (tabloid)	11.18 (38)	-	-	-
Sunday Star Times	13.06 (47)	-	-	-
Taranaki Daily News	16.04 (77)	0.42 (2)	-	-
Timaru Herald	14.15 (58)	0.24 (1)	0.24 (1)	-
Waikato Times	14.31 (83)	0.17 (1)	-	-
ALL	**13.66 (798)** (98%)	**0.17 (10)** (1%)	**0.12 (7)** (1%)	**0**

Table 11: Pronouns with anyone/-body per grammatical form, raw figures

they	*them*	*their(s)*	*he*	*him*	*his*	*he or she*	*him or her*	*his or her*	*she*	*her*
380	62	356	4	1	5	0	0	7	0	0

As with *some*, the majority of the plural pronouns found with *any* compounds were in the subjective form, with 380 hits in total.

> (31) Whether *anybody* is prepared to buy now depends on whether *they* believe the company's growth targets can be achieved. (dominion_post_4_5_2011_39)

> (32) It's painful for *anyone* to see the struggle and decline of someone *they* love. (the_press_9_10_2012_93)

The plural pronoun was found second most often in the possessive form with 356 hits, while in the objective form it appeared 62 times in total.

> (33) Comfort is important for *anyone* who spends a lot of time on *their* feet, and police are no exception. (dominion_post_13_5_2011_78)

> (34) You can't be horrible to *anyone*, because you know you're bound to pass *them* in the street the next day. (waikato_times_10_11_2010_56)

In 17 cases (2%) a pronoun other than *they* was recorded in the data. The masculine generic

amounted to 10 hits in total, appearing 4 times in the subjective form, once in the objective form

and 5 times in the possessive form.

> (35) Meals were held in an orderly fashion and if *anyone* was late, *he* missed out. (daily_news_24_4_2010_26)

> (36) A barbecue ban should be placed on *anyone* who cooks with the lid of *his* barbecue open so *he* can constantly poke and prod the meat -- (the_press_24_10_2012_107)

The coordinated pronoun was recorded 7 times in the data, all of them in the possessive form.

> (37) Has *anyone*, ever, liked *his or her* passport photo? (the_press_20_9_2012_162)

As with *everyone/-body*, no examples of generic *she* were found with *anyone/-body*.

5.4 No-one/nobody

Instances in the negative paradigm were the least abundant, with a total of 324 hits found in the

corpus. The frequencies are shown in the tables below.

Table 12: Frequencies of anaphoric pronouns with no-one/-body, N/1,000,000 (raw figures)

	they	*he*	*he or she*	*she*
Dominion Post	4.55 (46)	-	-	-
Manawatu Standard	5.60 (28)	0.40 (2)	0.20 (1)	-
Nelson Mail	7.02 (33)	-	-	-
Press	6.09 (67)	-	-	-
Southland Times	5.08 (30)	-	-	-
Sunday News (tabloid)	5.88 (20)	-	-	-
Sunday Star Times	5.28 (19)	-	0.28 (1)	-
Taranaki Daily News	5.21 (25)	0.42 (2)	0.21 (1)	-
Timaru Herald	6.10 (25)	0.49 (2)	-	0.24 (1)
Waikato Times	3.45 (20)	0.17 (1)	-	-
ALL	**5.36 (313)** (97%)	**0.12 (7)** (2%)	**0.05 (3)** (1%)	**0.02 (1)** (0%)

Table 13: Pronouns with no-one/-body per grammatical form, raw figures

they	them	their(s)	he	him	his	he or she	him or her	his or her	she	her
152	10	151	5	0	2	1	1	1	1	0

As with *some* and *any*, the plural pronoun most often appeared with *no-one/nobody* in the subjective form, with 152 hits in total.

> (38) The jobs just aren't there for people and *nobody's* employing unless *they* have to. (nelson_mail_10_8_2010_56)

> (39) *No-one* should be in Parliament unless *they* have been voted in by the people. (dominion_post_10_6_2011_61)

The pronoun was found second most often in the possessive form with 151 hits. As with the other three paradigms, the plural pronoun appeared considerably less often in the objective form than in the subjective or possessive form; *them* appeared 10 times in total in the negative paradigm.

> (40) My view was *nobody* should be discriminated against because of *their* occupation. (daily_news_25_6_2010_84)

> (41) 'We decided that we would go totally gluten free, then *no-one* could go home with anything by mistake, that would disagree with *them*,' says Sarah. (the_press_3_10_2012_163)

A total of 11 tokens (3%) other than *they* occurred in the data. The masculine generic appeared 7 times; 5 times in the subjective form and twice in the possessive form.

> (42) *Nobody* who holds the record for major championships in *his* chosen sport does it without a supersized ego. (manawatu_standard_1_9_2011_51)

> (43) After all, *no one* washes *his* hands before doing up *his* zip – the basins are inevitably too far away. (daily_news_14_8_2010_43)

The coordinated pronoun was found 3 times in total, once in each of the three forms and including the only objective form of this pronoun found in the corpus:

> (44) Just like three years ago, the risk for Labour in 2011 – if Mr Goff quits– is that *nobody* is really forced to rise to the challenge of proving that *he or she* has the numbers. (manawatu_standard_28_11_2011_71)

> (45) He is so popular that he could choose a complete *nobody* as his successor and get *him or her* elected. (daily_news_12_10_2010_46)

(46) Until then *no-one* would have imagined that *his/her* contributory overseas superannuations would be deducted from NZ Super. (sunday_star_times_22_7_2012_121)

Interestingly, another example of generic *she* was found in this paradigm, and will likewise be discussed in detail in 6.5.

(47) *No-one* wants a lover *she* feels *she* has to mother [...] (timaru_herald_8_12_2011_47)

5.5 Reflexive forms

The reflexives were counted separately from the other forms for the sake of clarity, as they may function slightly differently. The main interest in this section is firstly to look at the distribution of generic reflexives, and secondarily to investigate the present-day status of *themself*.

In total, 209 instances of reflexive pronouns were found in the corpus data and the frequencies are presented in Table 14. Only 5 (2%) of the instances featured a pronoun other than the plural epicene.

Table 14: Reflexive pronouns, N/1,000,000 words (raw figures)

	someone/-body	*everyone/-body*	*anyone/-body*	*no one/-body*	ALL
themselves	0.86 (50)	1.35 (79)	0.87 (51)	0.39 (23)	**3.48 (203)** (98%)
themself	0	0	0.02 (1)	0	**0.02 (1)** (0%)
himself	0	0.03 (2)	0.03 (2)	0	**0.07 (4)** (2%)
himself or herself	0	0.02 (1)	0	0	**0.02 (1)** (0%)
herself	0	0	0	0	**0**

As can be seen from the table above, the proportions appear to be in line with those of the subjective, objective, and possessive forms discussed above, and the proportions are nearly identical. The plural form *themselves* accounted for almost all of the tokens, appearing 203 times (98%) in total. The pronoun was popular with *everyone/-body* as expected due to its semantically

plural nature, but the form was also seemingly unproblematically used alongside *someone/-body* and *anyone/-body* almost as often.

> (48) *Everyone* had to have a good look at *themselves* and see where they needed to improve. (dominion_post_29_12_2011_61)

> (49) There were well-cut pieces of wood stacked there as though *someone* was trying to keep *themselves* warm. (manawatu_standard_10_10_2011_34)

> (50) When you have *anyone* in the group who see *themselves* as bigger than the team you're doomed to fail. (sunday_star_times_7_10_2012_139)

The most frequent singular epicene pronoun was *himself* with 4 tokens (2%) found in the non-assertive (51) and universal paradigm.

> (51) *Anyone* who survives the wet can classify *himself* a good driver. (manawatu_standard_29_10_2011_64)

No instances occurred in the assertive paradigm, which is rather surprising considering that the masculine generic was otherwise most frequently found with *someone/-body*. For comparison, in Laitinen (2007, 111) the division of reflexive *he* and *they* was more even, while here the distribution reflects that of the other forms, with *they* clearly the predominant pronoun.

The coordinated reflexive pronoun *himself or herself* appeared only once. This example is discussed in chapter 6.4 below.

> (52) *Everyone* is liable to imprisonment for a term not exceeding seven years who personates or represents *himself or herself* to be any person, living or dead [...] (sunday_star_times_16_12_2012_198)

There were no instances of generic *herself*, which accordingly reflects the rareness of the feminine pronoun throughout the data.

While *themselves* appeared quite often and was used seemingly unproblematically with the indefinite antecedents, only one instance of the "singular" reflexive form *themself* was found in the corpus (for comparison, a quick corpus search showed that *themself* appeared in reference to other than indefinite pronoun antecedents only 6 times in total).

> (53) Ask *anyone* in a similar position to mine and they'll tell you the last thing they wanted to do was bring attention to *themself* and win the term 'TF tosser'.

(waikato_times_7_8_2010_41)

This result was perhaps not surprising since the pronoun was similarly rare in previous studies: Abbott (1984, 47-48), for instance, found two examples of its use but was sceptical of it becoming a widely accepted form in the future. Indeed, more recently Baranowski (2002) found no examples of *themself* in American or British newspapers. The findings of this study would thus seem to confirm that *themself* is not in standard use yet, at least in the language of the press. The total number of reflexives found is too small, however, to make any concluding remarks on the current status of the form.

6 Analysis and discussion

In this section the corpus results will be analysed, and I will be looking at which features might influence the choice of generic pronoun. First I will compare the results from the different newspapers and discuss the findings. The bulk of this chapter, then, is dedicated to the qualitative analysis of each of the epicene pronouns and their usage. The influence of quoted versus non-quoted text and of *-one* and *-body* indefinites is also discussed. Finally, I will compare the findings to those in previous studies.

Overall, no great differences could be observed in pronoun usage between the different newspapers. The combined frequencies for each newspaper are presented in the table below.

Table 15: Distribution of personal pronouns across the newspapers, N/1,000,000 (raw figures)

	they	*he*	*he or she*	*she*
Dominion Post	70.00 (707)	0.89 (9)	1.00 (10)	0
Manawatu Standard	58.20 (291)	1.20 (6)	1.00 (5)	0
Nelson Mail	64.04 (301)	0.43 (2)	0	0
Press	58.45 (643)	0.91 (10)	0.64 (7)	0.09 (1)
Southland Times	56.10 (331)	0.68 (4)	0.17 (1)	0
Sunday News (tabloid)	74.70 (254)	0.29 (1)	0	0
Sunday Star Times	59.17 (213)	0.56 (2)	0.83 (3)	0
Taranaki Daily News	60.83 (292)	1.86 (9)	0.83 (4)	0
Timaru Herald	57.07 (234)	1.71 (7)	0.98 (4)	0.24 (1)
Waikato Times	57.59 (334)	0.69 (4)	0.34 (2)	0

The plural pronoun was consistently the most popular anaphorically used epicene pronoun in all 10 newspapers. In the tabloid magazine *Sunday News* the plural pronoun was slightly more frequent than in the other newspapers, whereas the other pronoun variants were conspicuously absent. This may be explained by the slightly different content material, as tabloids tend to emphasize topics such as sensational crime stories and astrology in which the neutral and ambiguous *they* is especially useful (see also 6.2 below). Furthermore, the physical size may also affect the choice, as tabloids are typically smaller than broadsheet newspapers, and perhaps the more brief plural

pronoun is therefore favoured over the coordinated pronoun in this format.

Variation does not appear to have any obvious connection with the publishing regions, and neither does there appear to be regional preference of generic pronouns. For instance, generic *he* was slightly more frequently used in *Manawatu Standard*, *Taranaki Daily News* and *Timaru Herald*, published in the northern, central and southern part of the country, respectively. Neither are the other pronoun variants preferred in any particular area. Thus the choice of pronoun does not seem to be influenced by geographical location to any significant extent.

It is also interesting to note that political standing did not seem to greatly affect the choice of pronoun, since the newspapers that were labelled as conservative, *Dominion Post*, *Press* and *Sunday Star Times*, did not display any notable differences in pronoun usage. Neither the singular pronouns nor the "more formal" pronoun *he or she* were found to be significantly more frequent in these newspapers.

6.1 Quoted and non-quoted text

One interesting feature to investigate is the division of the generic pronouns in non-quoted texts and quoted texts. Studying the differences between the quoted and non-quoted instances can also be regarded as a small-scale comparison of spoken and written English, which allows investigating if and how the register affects the use of pronouns, i.e. whether speakers and writers favour different constructions. Tokens were counted as quoted text if they were clearly identifiable as quotations or reported speech, for example if they appeared inside quotation marks. Figure 2 shows the distribution for each pronoun and the frequencies are presented in Table 16.

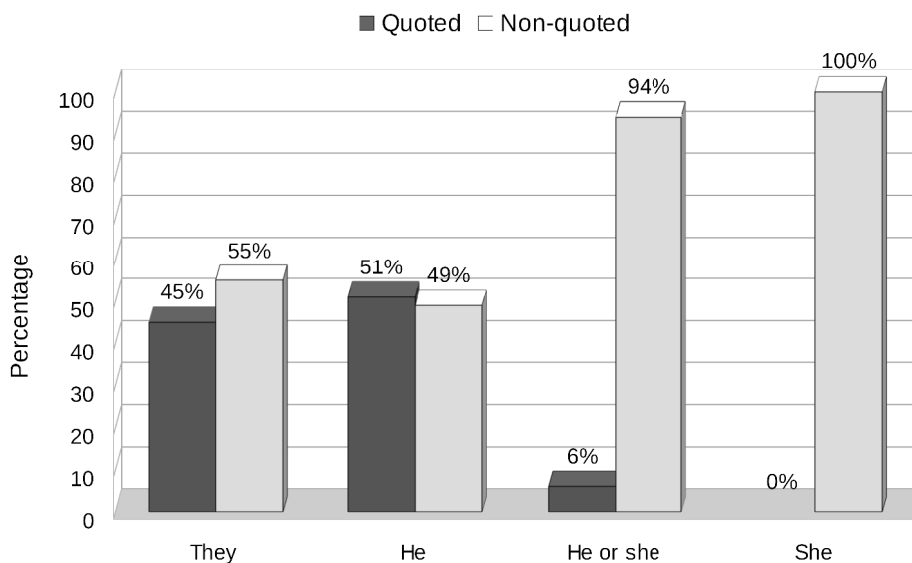*Figure 2: Distribution of generic pronouns in quoted and non-quoted texts*

Table 16: Frequencies of anaphoric pronouns in quoted/non-quoted text, raw figures

	some	*every*	*any*	*no*	TOTAL (% of all)
they	565/691	652/581	265/533	121/192	**1603** (98%) **/ 1997** (99%)
he	15/18	3/1	5/5	4/3	**27** (2%) **/ 26** (1%)
he or she	1/16	0/9	1/6	0/3	**2** (0%) **/ 34** (2%)
she	0/1	0	0	0/1	**0** (0%) **/ 2** (0%)

Overall, 1632 (44%) and 2059 (56%) instances were found in quoted and non-quoted texts, respectively. The proportion of quoted instances is quite high, perhaps due to the fact that indefinites tend to be more common in conversation than in formal registers such as news (Biber et al. 1999, 353). *They* was the preferred pronoun in both registers, although there were some differences between the paradigms: the universal paradigm included slightly more quoted instances, whereas in the other three the non-quoted examples were more frequent and distinctively more so in the non-assertive paradigm. The percentages with *they* appear to be slightly different to Balhorn's (2009, 399-401) findings in US newspapers, which showed that in AmE *they* was more common in quoted than non-quoted texts (59% / 41%). The singular pronouns, on the other hand, appeared in Balhorn's (ibid.) data more often in non-quoted texts (76%). In the NZE results, the numbers were

almost even for singular pronouns, with 27 quoted and 26 non-quoted instances of generic *he*. Generic *she* only appeared in non-quoted texts, which might suggest it appears more often in writing than in speech, but the sample size is too small to reliably confirm this.

The co-ordinated pronoun *he or she* was considerably more frequent in the non-quoted texts, appearing 34 times in non-quoted but only 2 times in quoted context. The low frequency in quoted speech seems to confirm many linguists' observations in previous research (e.g. Pauwels 2001) that this alternative is rare in spoken English. It was then perhaps not surprising that it mostly appeared in the written register, which requires planning and is more concerned with grammatical correctness. Similar observations were also made by Balhorn (2009, 399), who argues that the prevalence of non-quoted *he or she* in newspapers indicates its use is "subject to extralinguistic pressures" and suggests that "some writers or editors are considering number agreement and/or avoidance of sexism" by using *he or she*. This might be the case in the present data as well, since this pronoun was almost exclusively found in non-quoted contexts. Conversely, writers might be avoiding using generic *he* due to awareness of politically correct language. Agreement in number, however, seems to be less of an issue to most writers, considering that *they* was far more popular than *he or she* and the majority of the plural pronouns appeared in non-quoted texts nonetheless. On the basis of these findings it seems safe to conclude that *they* is preferred regardless of register.

6.2 *They*

Adversaries of *they* have sometimes claimed that the pronoun is too ambiguous and confusing, but if readers did not understand whether it expresses singular or plural meaning then writers would likely not use it. Looking at the present data, this is clearly not the case since singular *they* is flourishing in the newspapers. There was very little variation in pronoun use throughout the data and epicene *they* was clearly the predominant pronoun in all four paradigms, occurring in 97.5 % of all hits. The plural pronoun was most frequent (99%) with *everyone/everybody*. This result was

somewhat expected since the reference of the universal indefinite pronoun is often readily interpreted as semantically plural, as was noted in background literature (e.g. Quirk et al. 1985, 376; Wales 1996, 128). Another reason for the prevalence of *they* might be that especially in this paradigm a singular pronoun would seem odd in some contexts, for example:

> (54) *Everyone* wants to be a fashion designer but *they* don't know what it means. (dominion_post_14_9_2011_102)

The tendency of the universal indefinite pronoun to occur more often with plural than singular generics has been noted in many previous studies as well: Gerner (2000, 106), for instance, found that singular anaphors were "virtually non-existent" in the universal paradigm, and the same tendency was also observed by Laitinen (2007) and Neubauer (2008). In the present study *they* was also nearly as popular (97%) with *no-one/nobody,* the "negative equivalent of *everyone*"(cf. Wales 1996, 128). The hypothesis that *they* is chosen more often for the universal and negative paradigms thus seems to be at least partially confirmed, although perhaps by a smaller margin than expected since the frequencies are high in the other two paradigms (96% and 98%) as well. In other studies the frequency of *they* was usually considerably lower with *someone/-body,* whereas in the present data the difference is not as distinctive. Thus in NZE *they* seems to be preferred over singular pronouns even with this more individualised indefinite, and also more so than in other varieties.

Interestingly, especially in the non-assertive paradigm many of the examples occurred in conditional *if* -clauses, often requesting people to take action like in this example:

> (55) If *anyone* had any information about what happened, *they* should contact Cromwell police. (southland_times_30_1_2012_72)

It could be argued that using a singular pronoun in this context could potentially confuse the reader. Here using *he,* for instance, would imply that the police already knows that someone has the information they need. The coordinated pronoun, however, would have been a viable choice in most cases, but no instances of this pronoun were found in a similar context. This seems to suggest that the issue of brevity is important to consider in the newspaper genre, and it may be that the more

condense pronoun *they* is chosen in order to save space.

Another interesting use of *they* observed in the data, as was also noted in Wales (1996, 129), is that the pronoun is especially favoured by astrologists. These type of hits were more abundant in *Sunday News*, although they appeared to some extent in other newspapers as well.

> (56) Communication and conversation are highlighted today, Libra, as you enjoy catching up with *someone* special. Slow down and take the time to really listen to what *they* have to say. (the_press_2_7_2012_62)

The vague plural pronoun is indeed very convenient and suitable in this context, as it preserves the ambiguity of the (hypothetic) referent. In contrast, no singular pronouns and only 1 instance of *he or she* appeared in a similar context.

Similarly to findings in previous literature (e.g. Abbott 1984, 48) there were also cases where *they* was used even when it could be argued that only persons of one particular sex were discussed:

> (57) We are not going to pick *someone* because *they* are the only Pacific Island girl who turns up. (sunday_news_6_6_2010_105)

Here any possible referent would clearly be female, and yet *they* is used rather than *she*. Following Baranowski's (2002, 390) argumentation, in cases like this the scope of the singular antecedent can be seen to be extended to all possible girls who fit this description, representing a class of individuals, and thus the use of *they* seems justified. Another possibility is that usage like this displays awareness of gender-neutral or politically correct language, and perhaps of a need to preserve the ambiguity even where a sex-specific pronoun would be appropriate. In such context using *they* may reduce the chance of the recipients interpreting the reference as specific.

6.3 *He*

There were relatively few instances of the "old-fashioned" generic *he* compared to the plural pronoun, although with 54 tokens (1.5% of all results) it was the most frequent singular pronoun in the data and, rather surprisingly, it was also more frequent than the coordinated pronoun *he or she*

(36 tokens, 1.0%). However, instances of generic *he* were not entirely randomly distributed but some of them appeared in rather restricted contexts. As was noted above, half (51%) of the instances were from quoted texts, and these instances typically appeared to be quotes from novels or other literature, or quotations from celebrities, for example. Some examples also seemed idiomatic in style.

> (58) Christmas is a time when *everybody* wants *his* past forgotten and *his* present remembered. (dominion_post_24_12_2011_38)

> (59) There is *no-one* who became rich because *he* worked on a holiday, and *no-one* who became fat because *he* broke a fast. (manawatu_standard_7_5_2011_51)

The first example (58) turned out to be a quote from American comedian Phyllis Diller, while in example 59 the context reveals that this phrase is in fact an Ethiopian proverb. Quotations in general, and especially well-known and established ones such as these examples, are expected to be accurate and precise, and are therefore unlikely to be subject to editorial interventions. If such examples are disregarded, "truly" generic instances of *he* were in fact quite rare in the data.

Some general features can be distinguished in the use of generic *he*. In many instances the hypothetical referent could be seen as stereotypically male or it could be argued that all possible referents are male. Some "typical male activities" detected in the corpus samples include, for example, sports and cars:

> (60) *Nobody* who holds the record for major championships in *his* chosen sport does it without a supersized ego. (manawatu_standard_1_9_2011_51)

> (61) When driving at the speed limit, it's not unusual to be overtaken at high speed by *someone* who feels the need to demonstrate *his* new Mercedes, or driving skills, but who is otherwise probably a law-abiding citizen. (nelson_mail_21_10_2010_103)

Technology and, interestingly, spying were also described in masculine terms:

> (62) The video spy watch and a pen audio bug are supposed to be for the budding secret agent, but *anyone* with half a brain would not allow *his* child anywhere near the stuff. (daily_news_26_5_2010_88)

> (63) In spy talk, a 'sleeper' is *somebody* who lives *his* life in the target country, keeping *his* nose clean and climbing up the ranks of the local hierarchy, until *he* reaches a

position in which *he* can be of great service to *his* true employers abroad. (timaru_herald_19_8_2011_30)

Example 62 is an excerpt from a review of children's spy gear. As the writer is male, this use may reflect the author's gender and/or he perhaps considers men, or fathers, the primary target audience for this review. As such, the text seems to support the stereotypical view of technology and gadgets as boys' toys. Likewise, in the second example the repeated use of the pronoun makes it clear that spies are predominantly assumed to be male. Thus in these texts spying as a profession and the "tools of trade" are clearly seen as something that primarily men would be interested in. Some examples also portrayed another traditionally very androcentric area, the business world:

(64) It may be that *someone* was a line manager in the past but now *he* doesn't want that responsibility. (the_press_13_10_2012_39)

(65) I knocked and he said "What is it, Tall?" I said, "There's *someone* here to see you, sir". He replied: "Show *him* in". (southland_times_10_11_2012_68)

Here the second example is from an interview with a newspaper photographer, where the interviewee reminisces the "good old days". The full text reveals that *someone* is non-specific, i.e. it does not refer to any known visitor, but in the past business partners were of course automatically assumed to be male and generic *he* was used uncontroversially. In the wider context this example was quite unique, however. It was also interesting to note that although the sports news are "male dominated" as was discussed in 2.5.2, contrary to expectations only a few examples of generic *he* appeared in the sports section and *they* was far more popular in this section as well. Thus the effect of the sports category on the overall results can be considered to be quite marginal.

Generic *he* was not very frequently found in the reflexive form, but the proportion of the reflexives in relation to other forms was similar to previous studies (e.g. Holmes 1998; Laitinen 2007). Even in the reflexive form the masculine pronoun was somewhat restricted to special contexts, as one of the instances (equalling 25%) appeared in sporting context (51), while another example (66) appeared to be from a legal document.

(66) *Everyone* is justified in using, in defence of *himself* or another, such force as, in the

circumstances as he believes them to be, it is reasonable to use. (nelson_mail_5_10_2010_28)

There were also a couple of interesting cases that displayed mixed usage. In one example (67) the writer used both singular *he* and plural *they*, which may or may not be intentional, however, as the rest of the text did not display any pronoun alternation. In another example (68) the coordinated pronoun was alternated with generic *he*, making it seem like the feminine pronoun was added as a kind of an afterthought. Both texts were letters to the editor written by amateur writers, and do not necessarily display typical newspaper journalism, but they may be indication of a (perhaps subconscious) need for writers to consider gender-neutral language.

(67) Surely then, when *someone* is making every effort to turn *his* life around, it is a very mean response to pillory *them* as front-page news. (daily_news_20_9_2010_47)

(68) I do not want my well-informed vote to be negated by the unenlightened tick from *someone* who has been told that it is *his* duty to vote. It is not at all *his (or her)* duty to vote. (dominion_post_4_11_2011_84)

6.4 *He or she*

The coordinated pronoun accounted for only 1% of all instances and, rather surprisingly, in all but the universal paradigm this variant was less frequent than generic *he*. As the coordinated pronoun is often regarded as the most "correct" generic pronoun by many grammarians (see chapter 2.2.3), one would perhaps expect to find more instances of this use. Wales' (1996, 121) suggestion that this alternative is more likely to occur in subjective form rather than objective form seems to hold in the present study since the pronoun only appeared once in the objective form (less than 3% of all), although this grammatical form was rare with other variants as well.

The result seems to confirm Biber et al.'s (1999, 317) notion that the pronoun is "dispreferred in news". The complexity of the pronoun is perhaps the feature that has most bearing on its popularity, and its clumsy feel likely makes it less preferred in this register. Besides its clumsiness, it also takes up more space than *they*, and considering that the amount of space in

printed newspapers is usually quite limited, it would seem logical if shorter pronouns were preferred in this context. The coordinate pronoun has its merits as well, however, for example Wales (1996, 124) finds that often the easiest option for replacing pseudo-generic *he* in a publication is to simply insert *or she* (or *his/her,* etc.) in order "to save on rewriting time and printing costs", although she does admit that such forms may often read awkwardly in extensive runs.

Another reason for dispreferring the coordinated pronoun could lie in its even clumsier reflexive form, *himself or herself.* This is supported by the fact that only one instance of the reflexive pronoun was found in the data. It may be that this construction is simply too cumbersome and long for the newspaper genre (cf. Wales 1996, 121). Indeed, the one instance of the reflexive form appeared in what seems like an excerpt from law text or other legal text (see 5.5 above). The coordinated pronoun is perhaps more suited for this kind of formal context that requires precision, and where any possible ambiguity needs to be avoided.

Since this form can be especially cumbersome if used repeatedly, some writers have avoided this in an interesting manner by alternating between the coordinated pronoun and the plural pronoun, as in these examples:

> (69) When my credit card was lost in Wellington recently, *someone* pounced. Within hours, *he or she* had dined and drunk *their* way around several establishments in the capital to the tune of $1000 before I had realised and stopped the card. (daily_news_22_10_2010_87)

> (70) My idea of elegance is that *someone* is elegant when *he or she* shows a good knowledge of what fits *them*, where you can find naturalness and self-esteem. (the_press_29_2_2012_32)

This strategy of mixing the pronouns is quite innovative and seems to improve the flow of the text. There might be another reason for this alternation, though. By first introducing *he or she*, the writers may want to emphasize the singularity of the referent, after which it can be regarded "safe" to switch to the plural pronoun. This alternation seems like a clever way to avoid repeating and wasting space while at the same time preserving the singularity of the antecedent, and language users seem to have no problem with using these pronouns side by side.

One example displayed a unique pattern of pronoun alternation:

> (71) I am what is known in the business as a sleeper. No, not the sort of person who falls asleep at his desk after lunch, although I do that too. I am *someone* who worms *his or her* way into a foreign community with the aim of some day having access to valuable information and sources which would be useful to another power. (waikato_times_7_7_2010_30)

Unlike in the examples discussed in chapter 6.3 above where masculine pronouns were used consistently throughout the text, here the author first uses *his* in reference to *person* but then switches to the dual pronoun as if in a sudden need to emphasize the gender-neutralness of the antecedent. It is interesting how the writer mixes in the first person pronoun *I* as well, speaking from a personal perspective but keeping the reference generic at the same time. This type of pronoun mixing, as displayed in the examples above, perhaps exemplifies the development of new patterns of variation (cf. Cameron 2006, 739).

Considering that the length of the pronoun may be a crucial factor in newspaper language, it was interesting to note that contracted forms such as *s/he,* which are considered especially a phenomenon of written language, also appeared in the data.

> (72) Empathy is putting yourself in *someone* else's place, looking at things from *his/her* perspective. (dominion_post_2_2_2011_142)

> (73) The goal is not to find someone who is perfect, but rather to find *someone* whose character you respect – and who can either recognise *his/her* faults or accept feedback so *s/he* can be better in the future. (dominion_post_16_3_2011_122)

Although their advantage is that they are convenient and brief, the viability of these forms is somewhat limited, and some of these variants, such as *s/he,* have no objective or possessive forms. The fact that they are difficult to read has likely contributed to their rareness in the data as well. Furthermore, repeated use of these forms will likewise appear cumbersome, and possibly "smack of the bureaucratic form" as Wales (1996, 120) points out. Although these forms occasionally appeared in the data, they might seem too gimmicky to become standard usage.

6.5 *She*

Instances of generic *she* were quite rare, as only two tokens were recorded in the data. Based on previous findings, and especially due to the conflicting attitudes towards this variant and its strong ideological connotation, this result is perhaps not surprising. Gerner (2000), for instance, reported that only two non-referential examples of the feminine pronoun were found in the BNC (unlike in Gerner, referential instances have not been included in this study). The two non-referential and non-specific instances that occurred in the CNZNE, like many of the examples of generic *he*, appeared in somewhat restricted context:

> (74) If you spilled your drink on *someone* else's silk blouse, however, apologise immediately, do what you can to help *her* clean up, and offer to have the item cleaned or replaced. (the_press_5_12_2012_34)

> (75) *No-one* wants a lover *she* feels *she* has to mother, or who feels he has to dominate *her* with incessant rooster crowing and wing flapping. (timaru_herald_8_12_2011_47)

In example 74, the objective pronoun refers back to the hypothetical *someone* and the context reveals the text is an article offering advice for surviving the office Christmas party. This example is quite odd since elsewhere in the text *he or she* is normally used. The author is a woman, and the assumption here that a person wearing a silk blouse would be female perhaps reflects the writer's gender. In example 75, then, the word *lover* is neutral, but the pronoun *no-one* is assigned female reference. The context reveals that also here the author is female, and the text is a column discussing how short men have trouble in romance. The author seems to argue on the basis of her own experience, so it can be assumed that the gender of the writer is reflected in this use of generic *she* as well. It seems safe to say on the basis of these findings that at least in newspaper language generic *she* can not yet be considered standard usage.

6.6 Comparison of *-one* and *-body* forms

It is useful to also to look at the frequencies and the distribution of the anaphoric pronouns between *-one* and *-body* indefinite compounds and investigate how they may influence the choice of epicene

pronoun. The proportions of all pronoun combinations are presented in Table 17 below (only subjective forms are displayed for convenience' sake).

Table 17: Distribution of -one and -body compounds, raw figures

	someone	*somebody*	*everyone*	*everybody*	*anyone*	*anybody*	*no one*	*nobody*
they	1132 (90%)	124 (10%)	1056 (86%)	177 (14%)	736 (92%)	62 (8%)	216 (69%)	97 (31%)
he	27 (81%)	6 (19%)	2 (50%)	2 (50%)	10 (100%)	0	7 (100%)	0
she	1 (100%)	0	0	0	0	0	1 (100%)	0
he or she	17 (100%)	0	8 (89%)	1 (11%)	6 (80%)	1 (20%)	1 (33%)	2 (67%)
TOTAL	**1177 (90%)**	**130 (10%)**	**1066 (86%)**	**180 (14%)**	**752 (92%)**	**63 (8%)**	**225 (69%)**	**99 (31%)**

Overall, *-one* compounds appeared to be much more frequent than *-body* compounds, with 3220 (87%) and 472 (13%) hits in total, respectively. This reflects the observations made by Biber et al. (1998, 352-353) that *-one* pronouns are usually preferred in written registers, as well as the notion that in the negative paradigm the difference is less distinctive. The plural pronoun *they* is the most frequently used epicene in all paradigms and it consistently appeared more often with *-one* compounds, while, accordingly, in the negative paradigm the distribution was slightly more distinctive. Singular pronouns also anaphorised considerably more often in *-one* than *-body* compounds, amounting to 80 (87%) and 12 (13%) tokens in total. The masculine generic appeared more often with *-one* compounds in all except the universal paradigm, while generic *she* only appeared with *-one* forms. This perhaps indicates that the singulars are more readily used with these more individualised compound pronouns. The findings here seem to reflect those in Laitinen (2007) in that both generic *he* and *they* appeared more often with *-one* compounds, and *-body* forms were less frequent overall. The distribution of *he or she* also seems similar to that of *they*, except for the negative paradigm where *-body* forms were slightly more frequent with this variant. Perhaps the

most surprising finding here is the complete absence of the combination *somebody* followed by *he or she*.

6.7 Comparison with other varieties

The predominance of *they* in New Zealand English was somewhat expected considering that the same trend has been observed in American and British English and also in Australian English (see chapter 2.6), which have been influential varieties for NZE. Furthermore, the present findings for *they* also seem to correlate with those in spoken NZE (Holmes 1998, 33-34) where the plural pronoun accounted for 90% of all third person pronouns, although the results did include other non-specific antecedents such as *person* as well. With *someone/-body,* and thus also overall, the proportion of *they* seemed to be higher in the present study (96%) compared to, for example, 86% in the spoken part of the BNC (Gerner 2000), although it should be noted that those results included referential instances as well and the number of "truly" generic singulars may be smaller.

As the present investigation concerns newspaper language, it is especially worthwhile to see how the results compare to other contemporary newspapers, and I decided to use the fairly recent findings by Baranowski (2002) for comparison. Table 18 shows comparison of the frequencies of epicene pronouns in relation to quantifier antecedents in the CNZNE and in the two newspaper corpora in Baranowski (2002), with normalisations towards 1 million words by me.

Table 18: Frequencies of epicene pronouns with quantifier antecedents in the present study (bolded) and in Baranowski (2002), N/1,000,000

	Independent	*San Francisco Chronicles*	**CNZNE**
they	61.63	24	**61.64**
he	3.49	6	**0.92**
he or she	3.49	10	**0.62**
she	1.16	0	**0.03**

Independent represents BrE and *San Francisco Chronicles* represents AmE. Based on this

comparison, pronoun usage in NZE appears to be more consistent with BrE rather than AmE usage, especially in preferring the plural pronoun. Although quantifiers in Baranowski (2002) included also determiners (e.g. *some people)* and the findings may somewhat differ from those in the present study, the overall proportions of the pronouns appear to be quite similar in the CNZNE and BrE corpus. It is immediately obvious from the table, however, that in the NZE corpus there is less variation in pronoun use overall. In AmE especially the coordinated pronoun appears to be more popular. Even though the frequencies for this pronoun were lower in other studies (e.g. Pauwels 2001; Neubauer 2008) and in other AmE newspapers the proportion of *he or she* was only 4% (Balhorn 2009), the comparison shows that this variant was nevertheless infrequent (1% of all hits) in the New Zealand newspapers. Generic *he* is similarly less frequent in the present data than in the other two newspapers. Considering how the rates of *they* are similar in the BrE and NZE corpora, possibly New Zealand writers favour other strategies, such as pluralisation, or prefer to use more precise expressions such as NPs instead of singular generic pronouns. Verifying this hypothesis is beyond the scope of this study, however.

6.8 Evaluation of the findings

Overall, the material from the CNZNE seemed to produce adequate and reliable results and provide a comprehensive view of current generic pronoun usage in New Zealand English. As was pointed out in the methods section, there were some initial problems with the search method and the manual pruning of the results proved more challenging and time-consuming than anticipated, but going through all the results manually provided greater precision and as a result improved the reliability of the findings. The amount of data found was quite large, despite the fact that other types of strategies, such as pronoun avoidance, may be more preferred in news (cf. Biber et al. 1999, 239). In studies focusing on this topic it seems common that the total number of tokens is very small, or the material is quite old for such a time-sensitive issue, which also presented difficulties for finding

data that could reliably be used for comparison. The comparison way not be ideal, but it was important that findings from same register (newspapers) and type of construction (anaphoric reference to indefinite pronouns) were compared. The slight differences observed may not be statistically significant due to the small numbers, which may well be a result of chance. This could be confirmed by statistical analysis, but unfortunately acquainting myself with the theory and running and analysing statistical tests had to be abandoned due to time constraints, and therefore the findings may not be statistically significant all around. On a wider scale the results and their distribution can, however, be regarded as reliable and representable of the NZE variety. The predominance of *they* in NZE newspapers is irrefutable, and the results reflect the trends observed in other varieties.

7 Conclusion

This study set out to investigate the frequencies and distribution of generic third-person pronouns with indefinite antecedents in New Zealand English, which linguistic factors may influence their use, and whether these uses differ from those of American and British English. On the basis of the corpus findings it is safe to say that this variety appears to follow the same trend of favouring the plural pronoun *they* as has been observed in previous studies. The plural pronoun accounted for 97.5% of all pronouns and other variants were quite marginal. As expected, *they* was most common in the universal paradigm, which more easily lends itself to plural interpretation, although even in the assertive paradigm, where it was least popular, *they* still accounted for 96% of the tokens. Only traces of the prescription of generic *he* could now be observed in New Zealand newspaper English, and although it was the second most frequent epicene in the data (1.5 %), most of the instances appeared in quoted text or in special contexts. "Truly" generic instances of *he* were thus quite rare. A more surprising finding was that the coordinated pronoun *he or she* was quite infrequent (1.0 %), considering that it is regarded in the literature as the "most correct pronoun of all" (e.g. Baranowski 2002, 394). The clumsiness of the variant may be the most important reason for its dispreference in the newspapers. Generic *she,* on the other hand, was extremely rare in the data with only 2 instances in total, which indicates that this variant is not in standard use in NZE. Reflexive forms were also investigated and the results were similar to those of other forms, and *themselves* was clearly the preferred reflexive pronoun. Here the most interesting finding was that there was only one *themself* recorded in the data, which confirms that this "singular" reflexive is not yet in standard usage in New Zealand English. The distribution of *-body* and *-one* nouns was also investigated, and *-one* pronouns were the more frequent type as could be expected. Similarly to previous findings (e.g. Laitinen 2007) singular pronouns appeared much more often with *-one* pronouns, indicating that the singulars are perhaps more readily used with these more individualised compound pronouns.

The singular *they* was consistently the preferred generic pronoun in all of the newspapers

investigated, and there was no great variation between them, apart from the tabloid magazine *Sunday News* preferring *they* slightly more than the broadsheet magazines. Extralinguistic features, i.e. publishing region or political standing did not appear to influence the choice of pronoun either. Although sociolinguistic features could not be investigated in more detail, it should be pointed out that in a couple examples the choice of a generic pronoun seemed to reflect the author's gender, but this is only marginal and most writers opt for *they* in generic contexts. Interestingly, a few examples of innovative pronoun mixing and possible new patterns of variation were also detected in the data.

In a more detailed comparison with Baranowski's (2002) findings in British and American newspapers, the findings in NZE seemed to correspond more to BrE than AmE usage and the frequencies of *they* were almost identical in BrE and NZE. There was less variation overall in the present data, however, and the other generic pronouns were much less frequent in NZE than in the other two varieties. There were also differences between the varieties in the distribution of pronouns in quoted and non-quoted texts: in the present data *they* was equally common in both registers, while *he or she* seemed to be almost exclusively a feature of written register. The singular pronouns were more evenly distributed and in this case the results seemed to deviate from Balhorn's (2009) finding that in AmE singular pronouns are more common in non-quoted texts.

The popularity of *they* is also likely to have been affected to some extent by the increasing awareness of politically correct language. Compared to the other variants, *they* has unbiased connotation and is also simple and brief, and considering that *they* is now increasingly used in reference to non-binary gendered persons as well, it is no surprise that *they* is the most popular generic pronoun variant. Use of singular *they* is, however, not unproblematic and some still regard it as grammatically incorrect. Yet in the language of newspapers the construction can today be considered standard usage with indefinite pronouns, and the results suggest that *they* might well be on its way to becoming the accepted epicene singular pronoun. To further investigate the present-day status of singular *they* and pronoun usage in New Zealand English it would be useful to extend

the focus to dual pronouns and other antecedents. If possible, it might also be fruitful to examine

and compare other strategies, such as using NPs, to form a more complete picture of gender-neutral

usage.

References

Primary source

CNZNE – The Corpus of New Zealand Newspaper English

Secondary sources

Abbott, Gerry. 1984. "Unisex 'they'." *ELT journal* 38, 1: 45-48.

Adami, Elisabetta. 2009. "'To each reader *his*, *their* or *her* pronoun': Prescribed, proscribed and disregarded uses of generic pronouns in English" in *Studies in Practical Linguistics Volume 69: Corpus Linguistics: Refinements and Reassessments*, eds. Antoinette Renouf and Andrew Kehoe, 281-308. Amsterdam: Editions Rodopi.

Baker, Paul, Andrew Hardie and Tony McEnery. 2006. *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Balhorn, Mark. 2004. "The Rise of Epicene They." *Journal of English Linguistics* 32, 2: 79-104.

Balhorn, Mark. 2009. "The Epicene Pronoun in Contemporary Newspaper Prose." *American Speech* 84, 4: 391-413.

Ball, Catherine. 1994. "Automated Text Analysis: Cautionary Tales." *Literary and Linguistic Computing* 9, 4: 295-302.

Baranowski, Maciej. 2002. "Current usage of the epicene pronoun in written English." *Journal of Sociolinguistics* 6, 3: 378-397.

Baron, Dennis. 1981. "The Epicene Pronoun: The Word That Failed." *American Speech* 56, 2: 83-97.

Baron, Dennis. 1986. *Grammar and Gender.* New Haven: Yale University Press.

Bauer, Laurie and Paul Warren. 2008. "New Zealand English: Phonology." In *Varieties of English: The Pacific and Australasia. Handbook of Varieties of English, Volume III* , eds. Kate Burridge and Berndt Kortmann, 39-63. Berlin: Mouton de Gruyter.

Bell, Allan and Koenraad Kuiper. 2000. *New Zealand English.* Wellington: Victoria University Press.

Biber, Douglas. 2009. "A corpus-driven approach to formulaic language in English." *International Journal of Corpus Linguistics* 14, 3: 275–311.

Biber, Douglas. 2012. "Register as a Predictor of Linguistic Variation." *Corpus Linguistics and Linguistic Theory* 8, 1: 9-37.

Biber, Douglas, Ulla Connor and Thomas Upton. 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan, eds. 1999. *Longman Grammar of Spoken and Written English.* 3rd ed. London: Longman.

Bodine, Ann. 1975. "Androcentrism in prescriptive grammar: Singular 'they', sex-indefinite 'he', and 'he or she'." *Language in Society* 4, 2: 129-146.

Burchfield, Robert, ed. 1996. *The New Fowler's Modern English Usage.* Oxford: Clarendon Press.

Burridge, Kate and Jean Mulder. 1998. *English in Australia and New Zealand.* Oxford: Oxford University Press.

Cameron, Deborah. 1985. *Feminism and Linguistic Theory.* London: The Macmillan Press Ltd.

Cameron, Deborah. 2006. "Gender and the English Language." in *The Handbook of English Linguistics*, eds. Bas Aarts and April McMahon, 724-742. Oxford: Blackwell Publishing.

Carter, Ronald and Michael McCarthy. 2006. *Cambridge Grammar of English.* Cambridge: Cambridge University Press.

Comrie, Margie and Kate McMillan. 2012. "New Zealand National Report 2010. Global Media Monitoring Project." [Internet] Available from http://www.whomakesthenews.org/index.php/gmmp-2010/reports [Accessed 30 January 2015]

Comrie, Margie and Kate McMillan. 2013. "Running on the spot: NZ's record in news media gender equity." *Pacific Journalism Review* 19, 2: 196-214.

Deverson, Tony and Graeme Kennedy, eds. 2005. *New Zealand Oxford Dictionary.* Oxford: Oxford University Press.

Frank, Francine and Paula Treichler, eds. 1989. *Language, Gender and Professional Writing.* New York: Modern Language Association of America.

Fowler, Henry. 1926. *A Dictionary of Modern English Usage.* Oxford: Clarendon Press.

Garrett, Peter and Allan Bell. 1998. "Media and Discourse: A Critical Overview." In *Approaches to Media Discourse*, eds. Allan Bell and Peter Garrett, 1-20. Oxford: Blackwell.

Gerner, Jürgen. 2000. "Singular and plural anaphors of indefinite personal pronouns in spoken British English" in *Corpora Galore: Analyses and Techniques in Describing English,* ed. John Kirk, 93-114. Amsterdam/Atlanta: Rodopi.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, and Peter Trudgill. 2004. *New Zealand English: its origins and evolution.* Cambridge: Cambridge University Press.

Grimshaw, Patricia. 2001. "Reading the Silences: Suffrage Activists and Race in Nineteenth-Century Settler Societies." in *Women's Rights and Human Rights: International Historical Perspectives,* eds. Patricia Grimshaw, Katie Holmes and Marilyn Lake, 31-48. New York: Palgrave Macmillan.

Hay, Jennifer, Margaret Maclagan and Elizabeth Gordon. 2008. *New Zealand English*. Edinburgh: Edinburgh University Press.

Henley, Nancy. 1987. "This New Species That Seeks a New Language: On Sexism in Language and Language Change." in *Women and Language in Transition,* ed. Joyce Penfield, 3-27. New York: State University of New York Press.

Holmes, Janet. 1998. "Generic pronouns in the Wellington Corpus of Spoken New Zealand English." *Kotare* 1: 32-40.

Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Huddleston, Rodney and Geoffrey Pullum. 2005. *A Student's Introduction to English Grammar.* Cambridge: Cambridge University Press.

Hundt, Marianne, Jennifer Hay and Elizabeth Gordon. 2008. "New Zealand English: Morphosyntax." in *Varieties of English: The Pacific and Australasia, Handbook of Varieties of English, Volume III*, eds. Kate Burridge and Berndt Kortmann, 305-340. Berlin: Mouton de Gruyter.

Kachru, Braj. 1985. "Standards, codification and sociolinguistic realism: the English language in the outer circle." in *English in the world: teaching and learning the language and literatures*, eds. Randolph Quirk and Henry Widdowson, 11-30. Cambridge: Cambridge University Press.

Laitinen, Mikko. 2007. *Agreement Patterns in English: diachronic corpus studies on common-number pronouns*. Helsinki: Société néophilologique.

Lindquist, Hans. 2009. *English Language: Corpus Linguistics and the Description of English.* Edinburgh: Edinburgh University Press.

Lüdeling, Anke and Merja Kytö. 2008. *Corpus Linguistics: An International Handbook*. Berlin: Walter de Gruyter.

Martyna, Wendy. 1980. "The Psychology of the Generic Masculine." in *Women and language in literature and society*, eds. Sally McConnell-Ginet, Ruth Borker and Nelly Furman, 69-78. New York: Praeger.

Mayor, Michael, ed. 2009. *Longman Dictionary of Contemporary English* 5th ed. Harlow: Pearson Longman.

McConnell-Ginet, Sally. 1989. "The Sexual (Re)Production of Meaning: A Discourse-Based Theory." in *Language, Gender and Professional Writing,* eds. Francine Frank and Paula Treichler, 35-50. New York: Modern Language Association of America.

McEnery, Tony and Andrew Wilson. 2001. *Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Meyer, Charles. 2002. *English Corpus Linguistics: An introduction.* Cambridge: Cambridge University Press.

Meyers, Miriam. 1990. "Current Generic Pronoun Usage: An Empirical Study." *American Speech* 65, 3: 228-237.

Miller, Casey and Kate Swift. 1977. *Words and Women*. New York: Anchor Press/Doubleday.

Motschenbacher, Heiko. 2011. "Taking Queer Linguistics further: sociolinguistics and critical heteronormativity research." *International Journal of the Sociology of Language* 212: 149-179.

Mucchi-Faina, Angelica. 2005. "Visible or influential? Language reforms and gender (in)equality." *Social Science Information* 44, 1: 189-215.

Neubauer, Marion. 2008. "Coreference between singular epicenes and the plural pronoun." *Annual Review of Cognitive Linguistics* 6: 145-167.

News Works NZ. 2014. NZ Newspaper Titles [Internet] Available from http://www.newsworksnz.co.nz/titles [Accessed 10 January 2015]

Pauwels, Anne. 1998. *Women changing language.* New York: Addison Wesley Longman Ltd.

Pauwels, Anne. 2001. "Non-sexist language reform and generic pronouns in Australian English." *English World-Wide* 22, 1: 105-119.

Pauwels, Anne. 2003. "Linguistic sexism and feminist linguistic activism." in *The handbook of language and gender*, eds. Janet Holmes and Miriam Meyerhoff, 550-570. Berlin: Blackwell Publishing Ltd.

Pauwels, Anne and Joanne Winter. 2005. "Feminist linguistic activism in the 21st century: a view across the English-speaking world." in *Contexts - Historical, Social, Linguistic: Studies in Celebration of Toril Swan*, eds. Kevin McCafferty, Tove Bull and Kristin Killie, 107-128. Bern: Peter Lang.

Penfield, Joyce, ed. 1987. *Women and Language in Transition.* New York: State University of New York Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* London: Longman.

Reppen, Randi, Susan Fitzmaurice and Douglas Biber. 2002. *Using Corpora to Explore*

Linguistic Variation. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Rickman, Paul (forthcoming). "A Diachronic Corpus of New Zealand English, and a Study in Verb Complementation" in *Changing English: Global and Local Perspectives*, eds. Anna Mauranen, Juhani Klemola, Svetlana Vetchinnikova & Markku Filppula. Berlin: Mouton de Gruyter.

Simpson, John, ed. 2013. *Oxford English Dictionary Online*. 3rd ed. [Internet] Available from http://oed.com [Accessed 30 March 2015]

Sinclair, John et al., eds. 2006. *Collins Cobuild Advanced Learner's Dictionary* 5th ed. Glasgow: Harper Collins.

Soanes, Catherine and Angus Stevenson, eds. 2003. *Oxford Dictionary of English* 2nd ed. Oxford: Oxford University Press.

Stringer, Jeffrey and Robert Hopper. 1998. "Generic *he* in conversation?" *Quarterly journal of speech* 84: 201-221.

Wales, Katie. 1996. *Personal Pronouns in Present-Day English.* Cambridge: Cambridge University Press.

World Trade Press. 2010. *New Zealand Media, Internet and Telecommunications Complete Profile.* Petaluma: World Trade Press.