

WHAT CAN PRONOUNS TELL US?  
A CASE STUDY OF ENGLISH EPICENES

MICHAEL NEWMAN  
*Queens College, City University of New York*

ABSTRACT

The article reports an oral corpus-based study of epicene pronominal constructions (i.e., pronouns coreferent with singular antecedents and referring to referents of indeterminate sex) in English. *They* is used in 60% of the tokens, *he* in 25%, and other forms were used minimally. That variation corresponds to three semantic factors: perceived sex stereotypes associated with the referent, notional number, and, surprisingly, degree of individuation. These findings support accounts of the importance of agreement as a discourse-level phenomenon (e.g. Barlow's Discourse-Linking Theory), and of pronouns as elements whose informational content goes beyond mere denotation.

1. Introduction

1.1 *Pronoun theory*

Intuitively, a pronoun coreferent with an antecedent can be described as being in a relation expressed as a triangle such as that shown in Figure 1. The pronoun is in the right top corner, the antecedent in the left top corner, and their common referent at the bottom. The top side of the triangle represents a direct anaphoric relation, a semantic or syntactic link between the pronoun and antecedent. The right side indicates a referential relation linking the pronoun to the referent, and the left side denotes a similar link connecting the antecedent to the same referent.

As an intuitive description, this diagram makes few claims about the particular relations and entities shown respectively as lines and vertices.

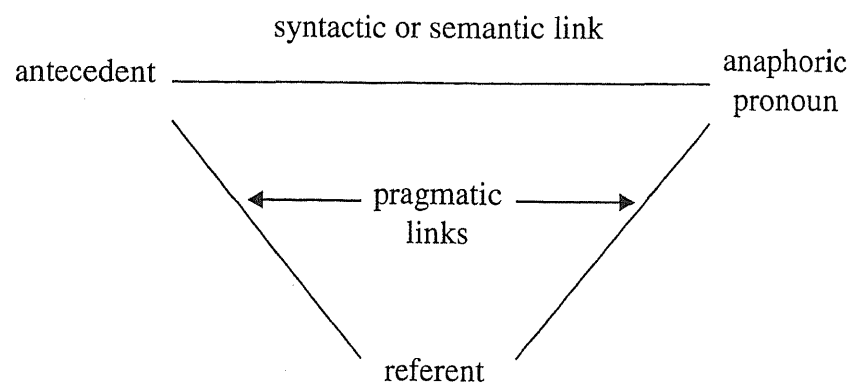


Figure 1. *The potential relations between pronoun-antecedent and referent*

Nevertheless, even a brief examination highlights an important distinction between two types of relations: The top line marks a link between two linguistic elements whereas the laterals indicate a mapping or encoding of some form of meaning of the referent. Therefore, any full account of anaphoric pronouns will necessarily be complex because it requires the integration of at least two different treatments. The approach needed to elucidate the anaphoric relation will be crucially unlike that needed to understand the anchoring — to use Barwise and Perry's (1983) and Pollard and Sag's (1994) term — of the pronoun and antecedent to the referent.

Given these considerations it can only be considered unfortunate that the bulk of recent research has concentrated exclusively on the anaphoric relation represented by the top line. For example, the following five issues, which have dominated pronoun theorists' research agendas, are all primarily or exclusively efforts to explore the relation between antecedent and pronoun: [i] the circumstances under which the anaphoric link can be described as a one of semantic dependence of pronoun upon antecedent (e.g. Lasnik 1976; Bolinger 1979; Evans 1980; Bosch 1983; Reinhart 1983; Rigau 1986; Van Hoek 1992; Fiengo & May 1994); [ii] the nature of semantic dependence, and how can it be modeled formally (e.g. Lasnik 1976; Evans 1979; Bosch 1983; Fiengo & May 1994); [iii] the nature of purely discursive anaphoric relations (e.g. Donnellan 1978; Bosch 1983; Cornish 1987; Fox 1987; Roberts 1989, 1995), [iv] how the anaphoric relation is resolved by linguistic agents, human or machine (e.g. Webber 1980; Sidner 1983;

Dahlbäck 1991; Grodzinsky, Wexler, Chien, Marakovitz 1993), and [v] the acquisition of anaphoric relations between pronouns and antecedents by children (e.g. Chien & Wexler 1990, 1991; Deutch, Koster, & Koster 1986; Foster-Cohen 1994; McKee 1992; Mazuka & Lust 1994).

By contrast, the right line connecting the pronoun to the referent has not been given much consideration. As a result of this imbalance in research we are left with an incomplete understanding of how pronouns function as referring expressions. Instead, they are typically assumed to be simple designators; their presence is presumed to obey only principles of economy, and their form is considered to be entirely dependent on factors of antecedent morphology and syntactic structure.

Such an approach is problematic because it implies pronouns are static elements, a view soundly criticized by Bollinger (1979), Bosch (1983), Cornish (1986, 1987), and Wiese (1983), and Fiengo & May (1994), among others. For example, it leaves no way to account for the pronominal variation evident, just in the case of English, in reference to collectives, higher animals, and epicenes. Similarly, pronominal variation arising from contrasts between natural and formal gender in many languages is left without explanation, as is the indexing of sociolinguistic factors as in T/V and honorific systems. Finally, there are indications (Utakis 1995, 1997) that referential characteristics of pronouns, such as definiteness, play a significant role in the acquisition process.

To be fair, it should be pointed out that this neglect of pronoun meaning has not been complete; there has long been some speculation in the literature about what the relation between pronoun and referent might look like. Lasnik (1976:9), for example, tentatively offers what might be called a *lexicalist* account of pronoun reference, that is one that assimilates the right to the left side of triangle:

One interpretation [...] although certainly not the only possible one, is that *he*, for example, means 'male human being,' and consequently, that that pronoun can be used to refer to any member of that class.

Later, another generative linguist, Rigau (1986:146), proposed an alternative, though related, picture in which full pronouns lie between more typically grammatical elements, such as clitics, and lexical items in their semantic properties.

Pronouns are not descriptions. Deictic pronouns designate but do not describe. In spite of that, the overt pronouns are closer to definite descrip-

tions — terms according to Hornstein — than are clitic or empty pronouns. Because of their lexical characteristics and the logical feature of uniqueness — i.e., because of their inherent semantic content — strong pronouns may be dealt with as terms.

Whatever the difference in the respective theorists' overall linguistic frameworks, these accounts resemble Bolinger's (1979) suggestion that pronouns are similar to lexical items, only semantically leaner.

Moreover, a few linguists have proposed more developed models of pronominal meaning. Bosch (1983, 1987) and Cornish (1986, 1987), for example, offer related though somewhat differing psycholinguistic theories to account for facts of pronoun variation in a number of languages. In both models pronoun form is determined by the presence of psycholinguistic entities in mental discourse models. Cornish goes so far as to see these "controllers" as the true antecedents of the pronoun.<sup>1</sup>

Rather than conceiving of pronoun form as a linguistic reflex of the properties or features of discourse entities, an intriguingly different view follows from the applications of Situation Theory (Barwise & Perry 1983, Devlin 1991). In a situated model the flow of information is the opposite of that assumed by Bosch and Cornish; pronouns are seen as sources not targets, supplying information regarding the referents they are anchored to. Speakers use that information in the construction and decoding of entities in discourse models.

The first elaborated situated view of pronoun form can be found in Barlow's (1992) Discourse-Linking Theory (DLT) of agreement. This theory provides a specific model of how pronoun selection may influence referential perspective, or the accumulation of properties ascribed to a referent beyond that of simple identity. Specifically, according to DLT, all agreement exponents — be they nouns, inflections, or pronouns — contain partial information that is added to representations of referents in discourse.

The data to be examined here provide an opportunity to confirm or deny certain predictions made by DLT. Specifically, the study examines the use of epicene and related pronouns in a corpus of extemporaneous English.<sup>2</sup> This domain of pronoun use provides a good case study for this model because pronoun variation in epicene contexts is common, making it possible to explore the different information potentially contributed by the various forms.

The layout of the paper is as follows: Section 2 consists of a description of the problems to be examined and the corpus used. Section 3 gives the

results of the study. These include the relative proportions of pronouns used and their correlations with various features of antecedents and properties of referents. They also include qualitative analyses of individual tokens, which provide data not discernible in the numbers. Section 4 discusses the theoretical implications. Specifically, pronoun choice was found to express specific limited sets of semantic properties. These can be combined with the properties of other coreferent elements (nominal or pronominal) giving rise to certain referential perspectives, precisely some of those found by Barlow (1992) in other languages. In Section 5, the conclusion, the notion that pronouns contribute information is further elaborated and suggestions are made for future research.

## 2. The case study of epicene pronouns

### 2.1 *Epicenes*

English epicenes are most frequently discussed as a prescriptive issue that arose from the problematic proscription of singular *they* and, relatedly, through complaints by feminist language critics about the sexism of the traditionally prescribed use of *he*. The term *epicene* is taken from Baron (1980), although as Corbett (p.c.) points out, this usage is shifted from its original Greek sense of common gender. Nevertheless, I follow Baron because the more common term used for this phenomenon, *generic*, has a different basic meaning of reference to a type or class, and no other term is readily available. Note that I use *generic* in its basic type-class sense in this study.

These terminological rough edges are perhaps symptoms of the fact that concern with this phenomenon has been more focused on critiquing and influencing usage than studying the linguistic facts behind it. If so, a more serious consequence is likely to be the frequent lack of any coherent explicit definition of the phenomenon, whatever it is called. Yet it seems that when writers discuss what I am calling 'epicenes,' they consistently refer to a pronoun used in one specific set of circumstances in which no canonical form is consistently found. Specifically, *an epicene pronoun is one that is coreferent with a singular antecedent and that refers to a referent of unknown or indeterminate sex.*

Despite the implicit consensus in the literature on these characteristics,

the mix of pragmatic and formal criteria gives rise to the suspicion that epicenes do not share any consistent constellation of purely linguistic characteristics. For example, pronouns placed into the empty slots in (1), will be epicenes by this definition assuming appropriate pragmatics. Nevertheless, it is hard to see what else they have in common, semantically, syntactically, pragmatically, or even with respect to constraints on pronoun choice.

- (1) a. When a person looks at \_\_\_\_\_ in the mirror, what might \_\_\_\_\_ see?  
 b. Under those conditions, anybody might end up contradicting \_\_\_\_\_ own position.  
 c. Every student continued in the program, and the results showed that \_\_\_\_\_ profited from the experience.  
 d. — Have you seen my stupid intern anywhere?  
 — I didn't know you had an intern. When did \_\_\_\_\_ start?  
 e. You never told me you had a pen pal in Spain. Do you have a photo of \_\_\_\_\_.  
 f. The typical American high school student believes that \_\_\_\_\_ life will be more difficult than \_\_\_\_\_ parents' was.

It is possible to use *he*, *she*, *he or she*, or *they* (1a) and (1b). However, only *they* is possible in (1c). Contrarily, a singular pronoun is highly favored if not required for most speakers in (1d) and (1e). Likewise, while intuitions vary, a good number of speakers also favor a singular pronoun in (1f) (see McConnell-Ginet 1979). Clearly, certain factors are limiting or constraining pronoun use in some contexts while permitting wide variation in use in others. Although there has been some speculation about what a few of those factors might be (Whitley 1978, McConnell-Ginet 1979, Weidmann, 1984), there is no comprehensive account in the literature.

## 2.2 Methodology

The corpus used in this study consists of the spontaneous discourse found in the transcripts of three episodes of each of eight (for a total of 24) nationally televised interview programs. The programs are:

<i>Crossfire</i>	<i>Donahue</i>
<i>Geraldo</i>	<i>Larry King Live</i>
<i>The MacNeil-Lehrer News Hour</i>	<i>Nightline</i>
<i>The Oprah Winfrey Show</i>	<i>Sally Jessy Raphaël</i> .

All the programs are an hour in length except *Nightline* and *Crossfire*, which were 30 minutes each. The shows were broadcast on three separate days: July 20, 1990; June 5, 1991; and November 20, 1991.<sup>3</sup> Not all the discourse which appears in the transcripts of these programs is included in the corpus; scripted sections — monologues, commentaries, and news reports — are excluded. No commercials are included in the transcripts, so the corpus consists of approximately 15 hours of predominantly extemporaneous discourse. The transcripts are commercially produced.

This corpus was chosen for several reasons. First the transcripts were readily available and easily converted into electronic form through scanning. This advantage was made possible by the unimportance of phonological characteristics and turn-taking pragmatics, which permitted the use of a commercially produced transcription using standard orthographic English spelling. Aside from convenience, the subject matter of the talk shows tended to be abstract, and so there was a high proportion of hypothetical and other irrealist referents. The presence of a good number of such referents was crucial for the success of the study because cases of real existing human referents whose sex is unknown to the speaker are rare.

One concern with the use of commercial transcripts was accuracy. So one the transcript of program, the *Sally Jessy* of June 5, 1991, was compared with a videotape for discrepancies between the language used and that recorded in the transcript. A total of 53 divergences between transcript and actual speech were found in the 7980-word document. The majority of these errata were the elimination of minor discourse features such as false starts, hesitations — such as word repetitions — back channel cues, and floor-holding signals. There were two cases in which the affirmation [mh-m] was transcribed as *yes*. Only ten discrepancies had any grammatical or lexical significance, and most of these cases seem to be lapses in the transcription process. However, two errors were of a type that is of concern considering the objective of this study: standard forms were substituted for nonstandard ones. In (2a) *was* — spoken by one of the guests — was replaced by *were* by the transcriber as in (2b), and in (3a) an African-American's zero copula was replaced by the reduced form 's as in (3b).

- (2) a. There **was** a lot of guilt feelings  
 b. There **were** a lot of guilt feelings
- (3) a. So you both have to keep praying and praying and praying because he always out there.  
 b. So you both have to keep praying and praying and praying because he's always out there.

These changes, however significant grammatically, were phonetically slight, and so they are possibly inadvertent — perhaps an unconscious translation into the transcriber's dialect. Evidence for this interpretation can be gleaned from the faithful transcription of other — sometimes identical — nonstandard features in this and other transcripts. Evidently, it is possible that similar mistranscriptions and unavoidable information losses (such as the homophony of the fully reduced forms of *him* and *them* and the lack of marking of stress) could cause some distortion in the actual count of pronouns. Nevertheless, in the absence of any systematic distortions on the part of the transcribers towards prescriptive norms, that disturbance is unlikely to be significant for the type of analysis aimed at in this study. No large-scale distortions were found as the vast majority of nonstandard features were left intact.

The analysis was begun by correlating antecedents with the various pronouns used. Then, tokens — consisting of antecedents and associated pronouns together — were correlated with a number of syntactic, and semantic, and pragmatic characteristics. Finally, certain tokens were studied individually for particular effects. For the sake of brevity, the relevant issues are elaborated on together with the results.

### 3. Results and discussion

#### 3.1 Overall epicene pronoun use

The results include two components: (i) quantitative data consisting of instances of the different pronouns used in tokens and correlations of those pronoun-based distributions with three descriptive properties, and (ii) qualitative data extracted from selected tokens which help illustrate properties not observable in the numbers.

The corpus provided 1,356 tokens containing coreferent (defined as those cases in which the antecedent and pronoun referred to the same

noncollective referent) relations, of which 106 or 7.8% of the total were epicene. Table 1 shows the distribution of these epicene tokens by pronoun used.

Table 1: Pronouns used with epicene tokens (% is in parentheses)

	<i>they</i>	<i>he</i>	<i>she</i>	<i>it</i>	inconsistent	disjunctive	total
Epicene tokens	65 (61.3)	26 (24.5)	1 (0.9)	5 (4.7)	5 (4.7)	4 (3.8)	106 (100)

The fact that *they* is used in more than 60% of these tokens supports the intuitions of a number of writers (e.g. Bodine, 1975; Miller & Swift 1980/1988; Newmeyer, 1978; Stanley, 1978; and Valian, 1977) that it is the most common pronoun used in epicene contexts. Considering this consensus, the greatest surprise might be the frequency of epicene *he*, used in almost one-quarter of epicene tokens. In addition to *he* and *they*, there were five tokens with *it* (4.7%) and four tokens with disjunctives, that is *she or he* or *he or she* (3.8%). Finally, there were five inconsistent tokens (4.7%), in which pronoun use varied. Epicene pronoun use in this corpus was, thus, characterized not by a particular "unmarked" pronoun but by variation.<sup>4</sup> In the following subsections, three meaning-laden factors that potentially motivate the choice of pronoun are discussed: sex bias, notional number, and individuation.

#### 3.2 Sex stereotyping in epicene tokens

The findings tend to support the arguments of feminist critics that in spite of a logically sex-indeterminate status of a referent, the use of *he* promotes masculine associations. Evidence to this effect can be found in previous research on sexism in language. However these studies have always made use of experimental instruments of various sorts (e.g. Kidd 1973; Martyna 1978, 1980, 1983; Cochran 1988; Khosroshahi 1989; Gastil 1990). This study provides the first confirmation of the allegations of male bias of *he* in supposedly sex-indefinite contexts in extemporaneous discourse.

The major way feminist critics have attempted to demonstrate bias in pronoun use is through correlation of a pronoun with classes of referents that are stereotyped as being male. Table 2 contains the type classes of the referents represented by epicene *he* and *they*. The bias of these classes was determined by 20 informants who were asked whether they felt that the class

described contained a masculine stereotype in American society. All informants were graduate students in linguistics or applied linguistics in the New York area. Twelve were females and eight were males. They were given written surveys with the classes of referents and three possible choices, "stereotypically male," "stereotypically female," and "no sex stereotype involved." To reduce any influence of worries about self presentation, it was emphasized that their responses were to reflect perceptions of social stereotypes in American society rather than their own personal impressions. The investigator remained with the respondents to answer any questions as they worked through the survey. The column marked "bias %" indicates the proportion of informants reporting male bias for that class.

Table 2: Types of referents in epicene **he** and **they** tokens

Stereotypically masculine referents	bias %	tokens with <i>they</i>	tokens with <i>he</i>	Stereotypically neutral or feminine referents	bias %	tokens with <i>they</i>	tokens with <i>he</i>
baseball scout	100		1	child	0	2	4
building contractor	100	1	1	customer or consumer	0	2	1
congressional rep.	70		1	divorcing spouse	0	4	
doctor	80	2	4	hypochondriac	15	3	
drug addict	85		1	malpractice victim	10	1	
infantalist	87	2	1	medical patient	0	1	
judge	95	1	1	talk show panelist	5	1	
lawyer	65	1		person	0	32	2
mass murderer	95		1	realtor	15	3	
Supreme Court nominee	90	1		sixteen-year old	5		1
professional athlete	95	2	1	store sales assistant	10	1	
Soviet general	100		1	witness to a crime	5	2	
timber worker	100		4				
TV executive	95		1				
Total		9	18	Total		52	8

*He* shows a strong correlation with male-biased referents as it was used in 18 of the 27 — 67% — of the male-biased tokens with either *he* or *they*. On the other hand, *he* was used in only 8 of the 60 neutral or feminine tokens, amounting to only 13%. Looking more closely at referents of these eight sex-neutral *he* cases, both *he* and *they* appeared in reference to *person*, but in a lopsided manner with only two out of thirty-four with *he*. *Child* would seem to be an exception, with four of six cases with *he*, but it could

be explained by the fact that on these shows, boys were far more frequently discussed than girls. *Consumer* and *16-year old* were also referred to by *he* alone. All these cases are discussed below.

There were too few tokens with other epicene pronouns to determine any patterns. Readers are invited to decide for themselves if there is any bias in the lone token containing epicene *she*. It is shown here together with one of the disjunctive tokens, which is the next token in the same anaphoric chain; the speaker is a male divorce lawyer speaking to the audience. Note that *you* has an arbitrary value.

- (4) In other words, if you can catch **your spouse** doing something, if you can catch **your spouse** out with **her** boyfriend, or you can catch **your spouse** doing something **she** shouldn't be doing **or he** shouldn't be doing, then it becomes a negotiating tool, in terms of the money. [S-I #87a and #87b]

The other three disjunctive cases follow. Two type-classes overlap with the categories in Table 2, and the other is explicitly ambisexual. An intriguing feature of these disjunctives concerns the likelihood that they may arise from some degree of conscious monitoring. Although there is no way to be sure, there are two factors supporting this interpretation. First, two cases (4 and 5a) are not so much disjunctive pronominal units as apparent repairs. Second, all the speakers employing disjunctive forms were members of professions involving careful public speaking. Two cases (4) and (5a), were spoken by lawyers. Example (5b) is from a spokesman for a professional organization, and the speaker of (5c) was a male politician. Furthermore, in the last two cases, the speakers were acting in their professional capacity, advocating the position of their association or party.

- (5) a. You should go to a **lawyer who cares about your particular case**, is interested in helping you, and does what **he** can **or she** can to help you. [S-1 #32a]  
 b. That leads **the physician and his or her** office having tremendous costs. [M-2 #25a]  
 c. **The American man and woman who has children and grandchildren** gets real value for **his or her** dollar. [N-3 #11a]

Finally, the degree of sex bias appears to vary with inconsistent cases. Again, readers are invited to judge for themselves:

- (6) a. What happens to a **Congressman** when **they**'re reprimanded? Do **they** stand in the well of the House and have a denunciation or a reprimand read to **them** by the Speaker?... All right. But after **he**'s reprimanded, what — does the Speaker list the case against **him**, and **he** stand and listen to it? [C-1 #18a]
- b. And the nicest thing about laughter is that the moment a **person** laughs, **he or she** is never more **himself**. [K-1 #62a]
- c. — If **the person** [avoiding payment of child support] leaves the state?  
— Well, unfortunately, you have to track **him** down sometimes. You have to go to another state. Get your judgment into another state, we recognize by full faith and credit. And then go after **them** where you find **them**. And it's difficult. [S-I #88a]
- d. Sometimes what I do is I'll tell **the client**, [in a criminal case] "Come on down with me. We'll take a lie-detector test. Lie detectors work." On the way down in the cab, **they** tell me the truth. ...But I won't put **them** on the witness stand, I won't lie to the jury, I won't lie to the public about whether **he**'s innocent or guilty. I'll take the case and I'll defend **him** on the legal merits. [K-II #25a]
- e. And this will sound maybe bad against contractors. However, if **someone**'s doing that, for example, and doing a plumbing job, and all of sudden you say, Can you do this? **They** will bite off more than **they** can chew. But that's the point where you've got to go — say, if you wanted, for example, a better area as a deck — you want to add a deck to your house. You have **him** do — "Aw, **he** can add a deck." You know, don't assume — You can't just — unfortunately, because **he** did do a good job, you thought — but just like — here, we'll take a totally different analogy — [O-II #19a]

Another way that critics have imputed sexism to epicene *he* is to note that the form is used more frequently by men than by women. Women and girls were found to reject purportedly epicene uses of *he* presented to them in greater number than men and boys (Martyyna 1978, 1980, Cochran 1988).

Female college students were also found to make greater use of nonsexist options in their own writing than their male counterparts (Meyers 1990). The extemporaneous data in this study did not show the same results. A  $\chi^2$  test showed no significant differences in the behavior of the two sexes with respect to epicene uses ( $p > .05$ ). Table 3 tabulates the total epicene tokens by the sex of the speaker:

Table 3: *Epicene pronouns by speaker sex*

Speaker sex	<i>they</i>	<i>he</i>	<i>she</i>	<i>it</i>	disjunctive	inconsistent	total
Males	44 (64.7)	13 (19.1)	1 (1.5)	1 (1.5)	4 (5.9)	5 (7.4)	68 (100)
Females	19 (55.9)	11 (32.4)	0	4 (11.8)	0	0	34 (100)
Unknowns	2 (50.0)	2 (50.0)	0	0	0	0	4 (100)
Total	65 (61.3)	26 (24.5)	1 (0.9)	5 (4.7)	4 (3.8)	5 (4.7)	106 (100)

In sum, there is a fairly strong tendency for *he* to be used with male-biased words and referents, but there are tokens that cannot be accounted for in this manner. Speakers of both sexes seem to avoid it though not consistently in neutral cases. *They* has mostly neutral referents, but there are some with strong potential for masculine bias. Female bias was rare, and, judging by the use of pronouns overall — 265 tokens with *she* versus 588 tokens with *he* — reference to females of any kind occurred less than half as frequently as did reference to males on these shows. Finally, in contrast to previous experimental research and analyses of planned writing, no conclusions could be drawn about the potential role of speaker sex.

### 3.3 Notional Number

The masculinity associated with the use of *he* can only be considered surprising if one starts from a theoretical stance that denies pronouns a contribution to meaning beyond mere designation. Once the door to a further contribution is open, masculine bias follows easily from the male reference inherent in *he*. If such is the case, it might also be expected that a sense of

singularity would be associated with *he*, but this time together with the other singular pronouns. Plurality would then be expected to be associated with *they*. However, upon initiating the analysis, number proved to be a more complicated category to map from form to meaning than originally supposed. The fundamental difficulty is that notional and syntactic number do not necessarily correspond any more than formal and notional categories do elsewhere. As Barlow (1992) notes, semantic and syntactic classifications are often confused, but they are two fundamentally different dimensions. One way they differ is that whereas syntactic number is essentially digital in nature, notional number is surprisingly analogic. The prototypical conceptions of oneness and multiplicity are two poles that only characterize those referents that are clearly composed of one unit or an aggregate of units. Many hypothetical and quantified referents do not present either property discretely. For example, the distributive sense characteristic of many quantifiers combines singular and plural aspects. The same can be said of formally singular generics because they refer to whole classes by a single example. For these reasons a third category called *neutrality* was added to the formal singular and plural. Any referent which is not unambiguously a single entity — and so singular — or clearly multiple — and therefore plural — was classified as neutral. The following examples are typical cases of number neutrality:

- (7) a. So, what does **a mother** do when **she** hears that? [O-III #11a]  
 b. Already **anyone** can take initiative of **their** own volition and get tested. [K-III #38a]

Even given this more flexible view of number, however, classification of tokens was not always straightforward. Although cases of distributional forms of quantification were considered neutral in general, certain usages maintain an emphasis on the singular side of the distributionality, and these cases were classified as singular. For example in (8), the distributional sense that reference could be to any attorney is almost hidden behind the highlighted singularity of *one*.

- (8) — But, you know, I thought, okay, **one attorney** would be able to do —  
 — But how do you know **they're** not going to be the one to help you? [S-I #50a]

Conversely, the plural side of distribution is most evident with *every* and *each*. Evidence for this plurality can be found in the strong formal constraints found upon the interpretability of singular pronouns coreferent to these elements.<sup>5</sup> The criteria for assignment follow:

- (i) A token is classified as singular if there is only one entity composing the referent.  
 (ii) A token is classified as plural if there is clearly more than one entity composing the referent  
 (iii) A token is classified as number neutral when it is not possible to discern whether there is one or more than one entity composing the referent. Usually these tokens contain a formally singular quantifier apart from *every* or *each* in the antecedent or are formally singular generics.

### 3.3.1 *The distribution of notional number*

The role of notional number of the token was assessed by comparing the distribution of pronouns across the three classes of referents as described above. Table 4 shows the distribution of notional number among epicenes.

Table 4: Epicene tokens by pronoun and notional number

Pronoun	singular	plural	neutral	total
<i>they</i>	4	15	46	65
<i>he</i>	4	0	22	26
<i>she</i>	0	0	1	1
<i>it</i>	2	0	3	5
disjunctives	0	0	4	4
inconsistent	0	0	5	5
All epicenes	11	15	81	106

While there is little evidence of absolute rules, there was a tendency for *they* to skew the distribution towards the plural end of the spectrum while *he* tended to do the opposite; these tendencies were too slight to be worth performing statistical analyses. On the other hand, there were a few truly singular, singular *theys*, and no plural epicene *hes*. Interestingly, the typical school grammar pattern (also popular in much of the literature on syntactic



constraints on anaphoric pronouns) of *every ... he* did not show up. That absence, of course, should not be interpreted as implying that it was in some way linguistically banned, only that it is probably rare in extemporaneous discourse.

### 3.3.2 Exceptions to general trends

There were six notionally singular referents in the corpus anaphorized with *they*, including four epicenes (9) and two with unambiguously male reference (10).

- (9) a. Somebody said to him, "Prince, what would you think of this place?" He said, "It's fantastic." **They** said, "How would you like to have a place like this in Monaco?" [G-1 #48a]
- b. He said that he was going and I assumed he was going hunting, because he often went hunting for squirrels, to shoot bottles, different things like that. Anyway, some time elapsed and my mother and I and my little sister went outside and I heard — I heard — saw **someone** run by and **they** were screaming, 'cause we lived right up the street from where McDonald's was. [G-III #30a]
- c. I'm sure that **whoever is chosen**, probably **their** viewpoints will be at least reasonably known on those issues, [N-1 #3b]
- d. — But, you know, I thought, okay, **one attorney** would be able to do —  
— But how do you know **they're** not going to be the one to help you? [S-I #50a]
- (10) a. Just before the program, we looked up **one of them**, and **they're** not in there. [reference is to a male doctor mentioned on the program] [K-1 #38a]
- b. If your child does something — wait a minute. Don't jump in yet. And they reprimand — you reprimand a child when they do something wrong, even though **this** is an adult, and perhaps **they** are saying that **they** are — [O-II #9b] [reference is to a misbehaving male contractor]

In (9a, b) and (10) the reference is singular because it is to a real individual, whereas in (9c, d) the referents are hypothetical. These two were

singular either because of explicit specification of number as in (9d) or, as in (9c), because of pragmatic factors — the referent is the person to be selected to fill the one vacancy on the Supreme Court. The conclusion to be drawn from these cases is that it appears that at least in this corpus, the singular pronouns were more singular than the plural one was plural.

### 3.4 Individuation

The first two potential factors associated with the use of pronouns in the corpus are intuitively attractive, but do not offer, by any means, a clean account of the data. Nor did they show any uniquely pronominal contribution of information regarding a referent, since that information was typically recoverable elsewhere. The final factor, individuation, though the least intuitive, turned out to be the most robust on both these accounts.

The finding that the use of singular *they* cannot be accounted for only through appeals to plural notional number and sex-indefiniteness is not original to this study. The point was made earlier by McConnell-Ginet (1979) and Weidmann (1984). Weidmann, for example, points to a number of cases, like those in (10), where singular *they* was used although the reference was sex-definite and singular:

- (11) If there is a Barbara Wassman on board, could *they* make *themselves* known to the cabin?  
Any person called Barbara may be assumed to be female, so the choice of *they* for anaphora cannot be prompted by the wish to remain non-committal about the sex of the person in question. What *they* does is to reiterate the meaning of the indefinite article before the name: it expresses uncertainty about the presence of any person called Barbara Wassman. (Weidmann 1984: 65)

Weidmann called this use of *they* "nonassertive" because it did not force a conclusion that the person was present, though she was named. McConnell-Ginet (1979) studied different, almost inverse, cases. They consisted of usages where singular *they* is barred in the view of many, though not all, speakers although it might be expected on the basis of sex indefiniteness and number neutrality:

- (12) \***The child** produces many utterances that **they** could not have heard. (from McConnell-Ginet (1979:76))

Importantly, McConnell-Ginet does not attribute the awkwardness these speakers feel solely to the definiteness of the antecedent but to a special semantic effect that she called "prototypicality." A prototype consists of the embodiment of the class that the generic represents in a sort of archtypical figure, similar to a fictional character. It is, essentially, a personification of the genus.

Although our shared prototype of a speaker is probably not very specific and indeed may not (in any very interesting sense) exist at all, to use a singular definite generic involves us in the pretense of a person-like prototypical genus representative to be named by that phrase. 'The speaker' invites us to hang flesh on the metaphorical bones of one who speaks, bringing to life a prototype to whom we can refer — and whom we can almost hear 'arguing with his wife.' (McConnell-Ginet, 1979 p.77)

These two concepts are evidently quite specific, and, moreover, in the context of an account of the use of singular *they*, they appear *ad hoc*. They account for real effects, nevertheless, and the application of the notion of individuation is an attempt to unify them as manifestations of a consistent and otherwise well-motivated semantic phenomenon.

This amalgamation is supported by the fact that individuation effects have been noted for pronoun usage and agreement in other languages. Monville-Burston (1983), for example, describes the influence of individuation on French pronominal clitic combinations. Following Burston (1983) she defines individuation as "a feature resulting from the presence in the referent of the clitic of properties such as humanness, animacy, definiteness, capacity to act or react" (Monville-Burston, 1983:246). As she describes it, individuation is a scalar property, and she concludes that in French two clitics can be combined only if at least one carries a reduced degree of it.

Roblee (1993) similarly asserts that the degree of individuation can constrain agreement patterns in Russian. Of particular interest for this study, formal mismatches in agreement can be licensed, as she sees it, by low degrees of individuation. Her definition of individuation is more comprehensive than Monville-Burston's. Specifically, she accepts Timberlake's (1975) understanding of it as referring to "the degree to which the participant is characterized as a distinct entity or individual in the narrated event." (Timberlake, cited in Roblee, 1993:424). In fact, Monville-Burston's more

concrete definition can be seen as falling out from this general construct, at least when the referents are human or human-like.

Furthermore, although they did not make use of the word "individuation," it is likely that the same concept can be found in some more general work on referentiality and specificity. For example, Givón's (1982) notion of "pragmatic referentiality" appears indistinguishable from that of individuation (as described by Roblee and Timberlake) in that he describes this construct as a dimension that varies between,

whether a particular individual argument (NP) is going to be *important* enough in the subsequent discourse, i.e. whether its *specific identity* is important, or only its generic *type membership*. (Givón 1982:84)

If this amalgamation of "pragmatic referentiality" and individuation is accepted, then Givón can be seen as usefully opposing individuation and genericness as two poles on a single scale. Such a position is further strengthened by purely semantic studies that appear to treat the same opposition. Donnellan (1966/1971, 1978), for example, famously describes definite terms such as *Smith's murderer* as varying between 'attributive' and 'referential' senses. The attributive sense corresponds to a generic reading because it refers to whoever must have murdered Smith in their capacity as *Smith murderer* without pointing to a specific identity. The referential sense corresponds to an individuated reading because in it *Smith's murderer* is equivalent to name, say, *Jones*.<sup>6</sup> Perhaps Jones was acquitted on grounds of insanity, and the speaker, not remembering his name, was giving an account of that fact. In any event, the working hypothesis in this study was that in epicene contexts, *they* would indicate low levels of individuation of a particular referent (a more generic sense) while singular pronouns would express high levels.<sup>7</sup>

While the concept of individuation seems cross-linguistically and theoretically well supported, to measure it does present a challenge in a language, such as English, in which there are few precedents for it.<sup>8</sup> To avoid circularity in deciding whether particular tokens presented individuated or generic referents, I established an individuation scale from one to five. All coreferent tokens, both epicene and nonepicene, were then placed on the scale on the basis of two criteria. Any referent referred to by a proper name or physically present in the context of situation was considered fully individuated and received the maximum individuation score of 5. Four raters examined the remaining tokens, giving them a score between 5 and a

minimum individuation score of 1. A 5 represents those referents treated as individuals with their identity relevant while a 1 corresponds to referents considered fully generic. A token that was ambiguous or of intermediate individuation would receive a score of 3. The raters were three advanced doctoral students in linguistics and one recent Ph.D. All had at least one course in semantics, and each had the concept of individuation explained, and as a test proved capable of repeating the salient points. The raters were given a survey containing all the remaining tokens. In most cases, the entire token was used, but in 18 cases the token extended over a considerable length of discourse, and only the first portion was used.

In the results, The raters were found to have a more than acceptable level of reliability (86.2%), with disagreements on opposing sides of the midpoint score of 3 occurring in a total of only 75 out of 546 tokens.<sup>9</sup>

#### 3.4.1 Individuation and epicene pronouns

The data support the hypothesis concerning the role of individuation in pronoun choice. Epicene *he* has a mean individuation score of 1.89 and all singular epicene pronouns together have a similar mean of 1.80. By contrast epicene *they* has a mean of 1.39. An ANOVA showed that these variances were significant ( $F=5.984$  (1, 89),  $p=0.016$ ) for *he* versus *they* and ( $F=6.591$  (1, 101),  $p=0.012$ ) for all singulars versus *they*. The variances between other pronouns were not significant. Interestingly, an examination of individual cases reveals a complex picture of the role of pronouns in fixing the individuation of the referents. Pronoun choice appears in these tokens as one ingredient in an elaborate calculus involving multiple sources of information. For example, in (13) there are various references to a hypothetical spouse in a divorce case. Each one shows a somewhat different referential perspective which appears to result from an interaction of the context with linguistic information provided by the pronoun and antecedent. The individuation scores attributed by the judges follow each example in parentheses.

- (13) a. — What does **the person** do when **they** don't have enough money? For one moment, off the pedestal and into the people. What do **they** do?  
 — Sally, let me tell you, **they're** in trouble. **They're** in trouble. And let me tell you something. Let me — Just a minute. Wait a minute. Because there are lawyers at different levels that charge different monies. (1) [S-I #64a]

- b. M: I've never seen a marriage yet work that had a pre-nuptial, always ends up in a divorce.  
 S: Marvin, in later marriages, I think there is a place for pre-nuptial agreements, people who are older, who have children they want to protect from their first marriage.  
 M: If they're fair and if they have supervision.  
 S: It's a different situation. ...  
 M: But usually they're designed to give **someone** less than what **they'd** get under the law. (1) [S-I #71a]
- c. It's a negotiating tool. In other words, if you can catch your spouse doing something, if you can catch **your spouse** out with **her** boyfriend, or you can catch **your spouse** doing something **she** shouldn't be doing **or he** shouldn't be doing, then it becomes a negotiating tool, in terms of the money. (1.5, both cases) [S-I #87a, #87b]

Examples (13a) and (13b) seem to be standard generic referents. They are situated in relatively unspecified hypothetical circumstances and are referred to not only by *they* but also by nominals with general reference. The last case, consisting of two tokens shown earlier, appears to invoke the kind of personified generic referent described by McConnell-Ginet as a prototype. This effect is achieved not only through the use of a generic with the definite article — as in McConnell-Ginet's account — but also by the use of an incipient hypothetical story, a more semantically specific antecedent, and the pronoun.

Example (14a) shows that, in fact, a definite NP is not even a necessary element in the creation of a prototype. This example is spoken by a different lawyer defending his modus operandi in divorce cases:

- (14) a. You want to win, the same way that a **brain surgeon** goes into an operation and **he** wants to cut out the entire tumor, you want to go all the way in a particular case. [S-I #44a]

Notice how the use of *he* in this case relates not only to the sex-role stereotype that brain surgeons are expected to be male but gives us the image of a brain surgeon to which the lawyer is comparing himself. The effect might be fairer with *they* but only at a loss of vividness that follows the lowering of individuation and the consequent lessening of the prototype effect:

- (14) b. You want to win, the same way that a **brain surgeon** goes into an operation and **they** want to cut out the entire tumor, you want to go all the way in a particular case.<sup>10</sup>

Example (15) again shows the power of narrative elements in augmenting the individuation. Yet this case is more complex because the lexical and pronominal reference together appear to contrast with listeners' knowledge of the real situation expressed in the narrative.

- (15) a. R: Right. Then, she did it and she wore this dress that was full of roses and her hair was soaking wet so she put a turban on because she was late for the appointment. They found that they really did have something to talk about. They both loved animals. They both loved the sea. They both didn't like prying photographers even though they were taking pictures of them. As a matter of fact, Rainier once told me that he felt like one of his monkeys in the zoo when photographers took pictures of him. He had a wonderful sense of —

G: But then, after that meeting, they didn't see each other for months.

R: Well, not so many months. It was from then till Christmas.

J: Eight months, nine months. It was between Cannes and Christmas. One of the things that Rainier told me that they had in common was that they both had lonely childhoods. They were both Catholic, so that they had something to talk about.

G: But how do you fall in love with **somebody** at a photo opportunity, not see **them** for eight months and the next time you see **them** you're engaged? [G-1 #40a] (3.25)

The score in (15) is near the midpoint because the raters split in their judgments. Surprisingly, this indeterminacy may be an appropriate response. Note that the situation was one where a claim has been made that two real people, Grace and Rainier, had an unusual courtship. Geraldo, the host, questions the account by referring to the improbability of such an event taking place as described. To make his case, he uses a generic, as indicated

by the use of *you*, *somebody* and — I believe — *them*.<sup>11</sup> At the same time, however, the reference is mediated by the concrete, and so individuated, case at hand that is available to the interlocutors and the audience. Thus, the raters' split (one gave it a 1, and the others a 4) reflects a subtle and possibly communicatively useful semantic ambiguity.

In all these cases, it would be possible to alter the pronouns without the sentence necessarily being barred grammatically or becoming incoherent semantically. That is not the case of (16a) which does not admit the change of pronoun exhibited in (16b):

- (16) a. — I have a **16-year-old** that thinks it's the greatest thing in the world.  
— Does **he**? [D-III #20a] (4.75))  
b. — I have a **16-year-old** that thinks it's the greatest thing in the world.  
— #Do **they**?

In this way, (16) provides perhaps the most clear-cut demonstration of role individuation in pronoun selection. The accuracy of the high score is indisputable. The teenager in question is identified as a person who believes certain things, but, atypically, without sex being specified. Similarly, *they* is also unlikely in the invented examples in (17) which are different types of individuated epicenes:

- (17) a. — I'm visiting **my cousin in Connecticut**  
— Where #do **they**/does **he**/does **she** live?  
b. Do you know a student named **Chris Davis**? #**They** are/**he/she**/**he** or **she** is on my roster but never came to class.  
c. — What do you think of **Pat** on "Saturday Night Live"?  
— I think #**they** are/?**he**?/**she** is very funny.  
d. I don't remember whether **Jaye Davidson** was nominated for best supporting actor or actress for **his/her/#their** role in *The Crying Game*.  
e. I think **RuPaul** is a real role model for today's youth. #**They/He/She** overcame tremendous odds to achieve success.

Note that it might be convenient for speakers to evade commitment to sex-specificity in these cases. For instance, if a speaker wrongly uses *he* or *she*, they are liable to be corrected in cases like (16a), (17a), and (17b). The fact that no correction occurred in (16a) thus implies that the interlocutor

guessed correctly. An evasive form of reference would be even more convenient in (17c–e) since no correct guessing is possible. Pat is a fictional TV character who is defined precisely by indeterminate sex, and viewers generally use *it*, the awkwardness of which has become part of the joke. Similarly, referring to transsexuals and transvestites such as Jaye Davidson or RuPaul can leave speakers acutely uncomfortable with pronoun choice. Again the option of using *they* as a way out is not available; the high individuation of these referents renders it impossible.

Returning to the corpus, example (18), mentioned previously in reference to notional number, presents precisely the contrary situation. Here low individuation appears to be supported by the use of *they* in spite of the singular notional number and the fact that the sex is presumably known to the speaker, who actually saw the person referred to:

- (18) He said that he was going and I assumed he was going hunting, because he often went hunting for squirrels, to shoot bottles, different things like that. Anyway, some time elapsed and my mother and I and my little sister went outside and I heard — I heard — saw **someone** run by and **they** were screaming, ‘cause we lived right up the street from where McDonald’s was. [G-III #30a] (3.00)

The individuation judgments were divided here as well with one judge giving the example a score of 5, whereas two gave it a 2, and one gave it a 3. I would have favored a low degree of individuation for this token for the following reasons: Not only is no clue available for the identity of the individual, but that identity is not actually relevant to the role the referent is playing in the discourse. It was a specific person, but it might as well have been anyone.

A final demonstration of individuation comes from those puzzling cases, discussed earlier, in which *they* is used in spite of the sex definiteness and singularity of the reference. In these tokens (shown as (10)) the use of singular *they* can now be explained on the basis of low individuation:

- (19) a. Just before the program, we looked up **one of them**, and **they’re** not in there. [reference is to a doctor mentioned on the program] [K-1 #38a]  
 b. If your child does something — wait a minute. Don’t jump in yet. And they reprimand — you reprimand a child **when**

they do something wrong, even though **this** is an adult, and perhaps **they** are saying that **they** are — [O-II #9b] [reference is to a misbehaving contractor]

In these cases, the identities of the referent are being downplayed in relation to their generic type class roles. Such cases are not anomalous; something similar appears in these cases from J. D. Salinger’s, *A Catcher in the Rye*:

- (20) a. You take **somebody’s mother**, all **they** want to hear about is what a hot-shot **their** son is.  
 b. You take **a guy like Morrow** that’s always snapping **their** towel at people’s asses — really trying to *hurt* somebody with it — **they** don’t just stay a rat while **they’re** a kid. **They** stay a rat **their** whole life. (J.D. Salinger *A Catcher in the Rye*, NY: Little Brown, pp. 56–7)

It can also be found in these other sex-definite singular *theys* culled from various sources:

- (20) c. I mean I don’t know **a single guy** who talks to **their** father about anything other than sports maybe, or who’s going to take the garbage out or mow the lawn or when *HE* was a kid or other very weighty things like that. (John Fox *The Boys on the Rock*, novel, 1984 NY: St. Martin’s, p. 59)  
 d. Circle No if you would not like this attribute in **a match**. A No acts as a vote against **the person**, it does not automatically disqualify **them**. (“Buddies and Pals” gay male dating service ad found in *Outweek* #42)  
 e. When you have a friendship with **a female**, you don’t want to ruin it by telling **them** that you like **them**. (Douglas Martin, an ex homeless man, quoted in “About New York” *NY Times*, 5–1–91)

The conclusion to be drawn from these examples is that pronouns, determiners, and nouns all contribute information that must be actively interpreted in terms of the context of situation. When this information is incompatible with that from other sources (i.e., the antecedent or context), a coherent referential perspective cannot be computed, and the utterance is infelicitous. Note that this understanding implies an active role for the

linguistic agent in the process of making reference, presumably in terms of schemata of various sorts.

This active role may be helpful in clearing up the otherwise seemingly intractable problem of the variation between speakers noted by several linguists in determining the acceptability of various definite singular generics with *they*. McConnell-Ginet (1979) and Bodine (1975), for example, disagreed over how acceptable such combinations were. Similarly, Whitely (1978) was puzzled about the fact that speakers would often reject definite singular *theys* when asked to judge their acceptability, in spite of the fact that they were found in widespread use.

These contradictions are easier to accept if pronouns are seen as one of a number of factors that contribute information concerning referential perspective, an effect which is ultimately created or interpreted actively by linguistic agents. It is certainly plausible that if these agents are asked to reflect on sentences containing contrasting information, they may come to differing conclusions as to the felicity of the resulting referents. Similarly, they are more likely to be intolerant of contradictions under such conditions than if they simply receive or even create them in extemporaneous discourse. If this perspective is accepted, then the contradictions would have to be particularly blatant, such as in (16) and (17), to result in a consensus of infelicity.

#### 3.4.2 Individuation and notional number

Nevertheless, the impossibility of *they* with indisputably high individuation does not appear absolute. Example (21) presents such a case.

- (21) And you all — **everybody** has gotten **their** biological babies except Jodie and that's what you want? [O-III #53a]

Although this token is a counterexample to any proposed absolute incompatibility of singular *they* and individuated cases, the hypothesis of the importance of individuation does not have to be abandoned. It simply must be restricted to referents that are not notionally plural. There can be little doubt, of course, that one of the main purposes of *they* is to refer to aggregates of sometime highly individuated people.

## 4. Interpretation

In sum, the descriptive findings of this study were as follows:

1. Epicene pronoun use is characterized by variation, and so there is no basis for a postulation of a formal class of epicene or indeterminate gender.
2. *They* is the most common pronoun, followed by *he*. Other forms are rarely used.
3. At least three factors are related to the variation:
  - (i) Probable reference to males is associated with *he*.
  - (ii) Plural notional number is associated with *they* and singular notional number is associated with *he*.
  - (iii) Low individuation is associated with *they*; high individuation is associated with *he* and singular pronouns generally.

### 4.1 *The semantics of pronouns*

This study provides strong evidence for theories that see anaphoric pronouns as dynamic elements in addition to elucidating a clearer understanding of the local though enigmatic issue of epicenes. The picture provided is one in which pronouns are seen as one of a number of sources of information about a referent. In that way, the findings support Barlow's (1992) Discourse-Linking Theory of agreement. The basic premise of DLT is summed up in the following description:

- Agreement is neither redundant (necessarily) nor a morphosyntactic relation. [...] It is argued that what are normally referred to as the source and target are not directly linked by an agreement relation, rather they instigate Discourse Referents, which consist mainly of properties that are determined by the morphology of the source and target. Interpreting agreement as a consistency condition linking Discourse Referents leads to new insights on the nature of agreement and its role in language. (Barlow 1992:3)

A Discourse Referent (DR) in DLT is best seen as a bundle of information that arises from and develops during discourse. Each referential term or related agreement morpheme contributes information to the DR. The agreeing information takes the form of properties, in Barwise and Perry's (1983:50-51) sense of simple 1-ary relations "of the kind recognized by human beings."

Properties are related to, but by no means identical with, morpho-syntactic features. In the simplest cases the relations are obvious. For example, Barlow (1992:157) gives the following illustration: In the case of a nominal such as *the boys* "the plural morphology may be associated with the property:  $\langle$ composed of individuals(x) $\rangle$ ." In this sense, the plural marker contributes that property to the discourse referent in a relation captured by the following function. On the left side is the morphosyntactic information, while on the right we find the corresponding semantic property:

$$f(\text{AGR}[\text{NUM:PL}]) \rightarrow \langle \text{COMPOSED-OF-INDIVIDUALS}(x) \rangle$$

In other cases, as in gender languages, the relation is not necessarily so straightforward. The information contained in the DR may not map directly into properties of the easily recognizable types that Barwise and Perry refer to. The fit between, say, masculine gender, and male sex is of course not particularly good. Instead, gender refers alternatively to word classes. So the information contributed by gender markers refer to the membership of related items in the relevant class; in other words, the semantic property of gender is, in many cases, ultimately linguistic. The use of a gendered pronoun (or other agreement morpheme) limits the class of potential referents to those that are likely to be referred to by nouns within that class.

Note that in Barlow's theory an antecedent is not seen as referring to a DR, but as initiating one; in this DLT crucially differs from the psycholinguistic models assumed by Bosch and Cornish. Any subsequent nominal elements referring to the same referent, or verbal or adjectival agreement morphemes associated with it morphosyntactically, similarly provide information about that object by establishing new DRs. These DRs merge as the discourse proceeds, adding information. In this sense, we see agreement operating as a dynamic system of information manifestation, transfer, and synthesis.

Most of the time the information supplied by subsequent DRs is redundant with the earlier information established by the initiating nominal as in the following case.

(22) The boy often argues with his teacher.

*His* matches the maleness property of *boy*, and both *his* and *-s* match *boy* in the feature of singularity. The predominance of this pattern of strict consistency of properties and features has led to the view that agreement is a redundancy relation. Similarly it may have led to the assumption that

semantic properties are isomorphic with the corresponding grammatical features. However, as we have seen, this consistency, however typical, is by no means universal, nor is there any reason to accept it as the essential quality of agreement. At times, there are contrasts, and in these cases the felicity of the combination is determined by whether these contrasts are resolvable in terms of properties of the DRs, not features of the linguistic elements. As Barlow (1992:159) puts it:

It is crucial to DLT that compatibility between antecedent and anaphor is based on consistency of properties associated with discourse referents and not on matching of syntactic features. The different predictions made by a property-based discourse account and a feature-based account are quite clear. If linking of discourse referents is based on compatible properties, then the antecedent and anaphor can differ in syntactic features as long as their properties are consistent.

Barlow thus argues that cases of feature mismatches such as singular *they* are in no way aberrant, and in fact what appear to be contradictory features can give rise synergistically to properties that are not themselves encoded in any single DR. For example, note in these tokens how the inconsistent pronominalization of the collective gives rise to two different referential perspectives. The singular forms point to the institutional aspect of the collectivity and the plural ones to the multiplicity of the component members:

- (23) a. I think **the administration** has **itself** a political problem, probably of **its** own making. At a time when we don't have money for education, nutrition, infrastructure and all here in the United States, we're told that **they** want \$500 million for Latin America and now \$4 billion for the Soviet Union. [N-III #7a]
- b. I think **baseball** has a responsibility to show minorities in this country that since Jackie Robinson broke into baseball that **they** are concerned about their presence, not only as a baseball player but being involved from a total standpoint, not only from coaching and management, but also from being vendors, doctors, lawyers, whatever **it** has to offer. [N-II #4a]

As Barlow (1992:95) points out:

Noun-agreement mismatches are connected with the encoding of classificatory information. These include: collectives, collections, definiteness, animacy, gender, and respect. These are not isolated examples, they illustrate patterns that show up repeatedly in languages.

#### 4.2 DLT and epicenes

The question remains, however: How can DLT be applied to the phenomena looked at here? Figure 2 schematizes an example using a simplified version of the diagrams developed for describing the interaction of DRs in Barlow (1992) for the arguments in the example sentences below.<sup>12</sup>

- (24) a. A **teenager** often thinks that **he** is immortal.  
 b. A **teenager** often thinks that **they** are immortal.

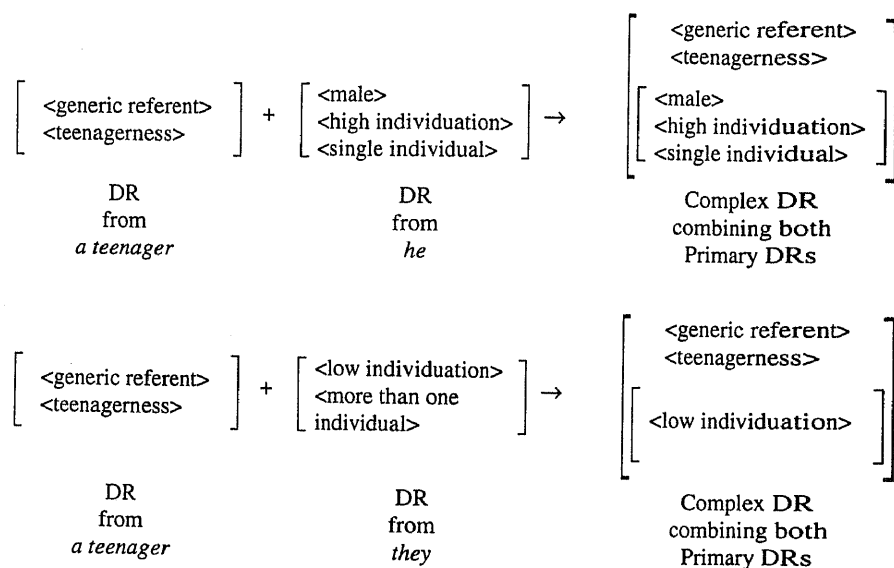


Figure 2. Discourse referents for a teenager ... he and a teenager ... they.

I have assumed that neutral notional number is simply an absence of specification of that dimension. *He* adds the qualities of singularity, high

individuation and maleness. These qualities are combined as shown in the complex DR. The apparent contrast of information regarding genericness and high individuation is resolved by interpreting the referent as a prototype.<sup>13</sup> In contrast to *he*, the only information contributed by *they* is low individuation, and the resulting complex DR thus contains no information regarding sex. As for number, the contrast between singularity and plurality is unproblematic because the plural feature is not necessarily isomorphic to a property of 'more than one' though it may very well favor it. In any case, the generic is number neutral.

In this way, DLT not only supplies a linguistic mechanism to account for singular *they*, but it also provides an intuitively attractive way of dealing with the phenomenon of prototypes, discussed by McConnell-Ginet (1979), and the "nonassertiveness" of Weidmann (1984). They are nothing more than specific schematic resolutions of combinations of semantic properties. Recall that the former consisted of treating a generic as if it were individuated, while the latter involved taking a real referent and downplaying the individuation. In both these cases, the pronoun signals an individuation that is not in accord with the logical status of the referent, yet the result is not incoherent but a subtle and evocative product of the mix of two contrasting qualities. DLT by treating referential information on a more abstract level thus provides a far more fine-grained model of the semantics of referential perspective than we are accustomed to in linguistic research.

## 5. Conclusions

Epicenes have largely been considered an applied or sociolinguistic issue, mainly of concern to those interested in issues of sex-bias in language and/or prescription. This study shows that they are of considerable interest for linguistic theory as well. Particularly, the relations involving morphosyntactic mismatches revealed factors that are hidden when features match — as they do most of the time — due to the redundancy of semantics and syntax in these cases. In any case, the relation between anaphors and their referents is largely virgin territory in pronoun research. This case study shows, I believe, the potential richness of the phenomena that can be discovered. Pronouns were shown to contribute significant information concerning referential perspective, and they thus involve choice about reference rather than simply responding to morphosyntactic features.



In order to further pursue that nature of the choices and the range of informational contributions, further investigation is warranted. Work of this nature should be crosslinguistic, and involve other classes of referents (e.g. collectives) where agreement mismatches between pronoun and antecedent are frequent.

*Author's address*

Michael Newman  
 Department of Linguistics and Communications Disorders  
 Queens College, CUNY  
 65-30 Kissena Blvd.  
 Flushing, New York 11367-1597  
 newman@cuny.campus.mci.net

NOTES

1. The lexical item normally considered the antecedent is considered by him as merely the "antecedent trigger" because it instigates the controller in a speaker's mental discourse model.
2. The work presented here extends a larger study of epicene relations, Newman (1997).
3. The only exception was that the Sally Jessy Raphaël program of November 19, 1991 was used in place of that of November 20, which was temporarily unavailable for reasons which are unknown to me.
4. This variation is consistent with historical data, dating as it does back at least to the Middle English period (see Newman 1992, 1997). Some authors, both linguists (e.g. Mühlhäusler & Harré 1993) and feminist critics (e.g. Miller and Swift 1980/1988), probably misconstruing Bodine (1975), have claimed that epicene *he* was merely an invention of prescriptive grammarians. This early evidence of its use should lay such ideas to rest.
5. There is ample cross-theoretical literature on these issues (e.g. Bosch 1983; Reinhart 1983, Fiengo & May 1994; Roberts 1995), which coincides in the existence of some kind of limit on singular agreement between antecedents with *every* and pronouns. Beyond these

limits, when presumably only notional factors are determining agreement, only a plural pronoun is possible.

6. However, see Abbott (1993) for an account of these phenomena in purely logicosemantic terms as "rigid designators."
7. I should note that this account provides a substantial refining of my previous conclusions on the semantics of singular *they* (in Newman 1992), in which I supposed that the crucial factor was an ad hoc category, which I termed 'solidity.'
8. In English, the term "individuation" was used by Mufwene (1984, 1986) as the basis for a reanalysis of the count/mass distinction. Undoubtedly, the conception is related to the one described here; whether it is ultimately identical or not goes beyond the scope of this article.
9. Of course on a subtle semantic property like individuation, disagreement was to be expected. The disagreements included maximally opposing judgments (i.e., 1 and 5) 3.7% of the time, an indication of the robustness of the concept.
10. Although the prototype effect is quite clear in this contrast, it was not picked up by the raters who gave (16a) an individuation score of 1.
11. Franklin Horowitz (p.c.) points out the possibility that *them* in some way is a continuation of the earlier anaphoric chain composed of both Rainier and Grace, as when Geraldo says "they didn't see each other for months." In this sense this *they* would be notionally plural. Evidently, that level of analysis is too subtle to determine with the data at hand.
12. Much of the theoretical apparatus has been left out because it would take us into concepts far beyond the scope of this work. Also, Barlow uses properties such as ⟨unspecified number⟩ and ⟨unspecified gender⟩ which I have left out.
13. The model is no doubt an oversimplification. As we have seen, the qualities of sex reference, notional number and individuation are not unrelated semantic properties, but interact with each other. The context, situation, and other elements in the sentence, such as the verbal aspect and the adverb, are also implicated in the determination that the teenager in question is generic, not one coming down the street.

REFERENCES

- Abbott, Barbara. 1993. "Referentiality, specificity, strength, and individual concepts". In Duncan, Erin; Farkas, Donka; & Spaelti, Philip (eds), *Proceedings of the 12th West Coast Conference on Formal Linguistics*.
- Barlow, Michael. 1992. *A situated theory of agreement*. New York: Garland.
- Baron, Dennis. 1981. "The epicene pronoun: The word that failed." in *American Speech* 56: 83-97
- Bodine, Anne. 1975. "Androcentrism in prescriptive grammar: Singular 'they,' sex-indefinite 'he' and 'he or she'." *Language in Society* 4: 129-146.

- Bolinger, Dwight. 1979. "Pronouns in Discourse". In T. Givón (ed.), *Syntax and Semantics V*. 289–310. New York: Academic Press.
- Bolinger, Dwight. 1980. *Language — the Loaded Weapon*. London: Longman.
- Bosch, Peter. 1983. *Agreement and anaphora: a study of the role of pronouns in syntax and discourse*. London, NY: Academic Press.
- Burston, Jack. 1983. "Clitic Object sequence and cooccurrence restrictions in French". *Linguistic Analysis* 11: 237–265.
- Chien, Yu Chin & Wexler, Kenneth. 1990. "Children's knowledge of locality conditions as evidence for the modularity of syntax and pragmatics". *Language Acquisition* 1: 225–295.
- Chien, Yu Chin & Wexler, Kenneth. 1991. "Children's knowledge of pronouns as bound variables in a long-distance context". *Papers and reports on Child Language Development* 30: 25–38.
- Cochran, Effie. 1988. *Generic masculine pronominal usage and sex-linked occupational stereotypes among High School students*. Doctoral Dissertation, Teachers College, Columbia U.
- Cooper, Robert. 1985. "Selling language reform". in D. Tannen & J. Alatis (eds) *Georgetown Roundtable on Language and Linguistics*.
- Cornish, Francis. 1986. *Anaphoric Relations in English and French*. London: Croom Helm.
- Cornish, Francis. 1987. "Anaphoric pronoun: Under linguistic control or signaling particular discourse representations?" *Journal of Semantics* 5: 233–260.
- Dahlbäck, Nils. 1991. *Representations of discourse: Cognitive and computational aspects*. Linköping, Sweden [*Linköping Studies in Arts and Science* 71/*Science and Technology* #264].
- Deutsch, Werner; Koster, Charlotte & Koster, Jan. 1986. "What can we learn from children's errors in understanding anaphora". *Linguistics* 24: 203–225.
- Donnellan, Keith. 1966/1971. "Reference and definite descriptions". In Danny D. Steinberg & Leon A. Jakobovits. (eds), *Semantics; an interdisciplinary reader in philosophy, linguistics, and psychology*, 100–114. Cambridge: Cambridge University Press
- Donnellan, Keith. 1978. "Speaker References, Descriptions and Anaphora". In P. Cole, (ed.), *Syntax and Semantics [Pragmatics 9]*.
- Evans, Gareth. 1980. "Pronouns". *Linguistic Inquiry* 11: 337–362.

- Foster-Cohen S. 1994. "Exploring the boundary between syntax and pragmatics: relevance and the binding of pronouns". *Journal of Child Language* 21: 237–255.
- Fiengo, Robert & May, Robert. 1994. *Indices and Identity*. Cambridge, MA: MIT Press.
- Fox, Barbara. 1987. *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press
- Gastil, John. 1990. "Generic pronouns and sexist language: The oxymoronic character of masculine generics." *Sex Roles* 23: 629–643.
- Givón, Talmy. 1982. "Logic vs. Pragmatics, with human language as the referee: Toward an empirically viable epistemology". *Journal of Semantics* 6: 81–133.
- Gregersen, Edgar A. 1979. "Sexual linguistics". In J. Orasanu, M. Slater, L. Adler. (eds), *Language, sex and gender: Does la difference make a difference*. New York: Academy of Science
- Grodzinsky, Yosef; Wexler, Kenneth; Chien, Yu Chin; Marakovitz, Susan. 1993. "The breakdown of binding relations". *Brain-and-Language* 5: 396–422
- Khosroshahi, Fatemeh. 1989. "Penguins don't care, but women do: A social identity analysis of a Whorfian problem". *Language in Society* 18: 505–525.
- Kidd, Virginia. 1971. "A study of the images produced through the use of male pronouns as the generic". *Moments in Contemporary Rhetoric and Communications*. 1: 25–30.
- Lasnik, Howard. 1975. "Remarks on Coreference". *Linguistic Analysis* 2: 1–22.
- Martyna, Wendy. 1978. "What does 'he' mean? Use of the generic masculine". *Journal of Communication* 28(1): 131–138.
- Martyna, Wendy. 1980. "The psychology of the generic masculine". R. Borker, & N. Ferman. (eds), *Women and language in literature and society*, 69–78. New York: Praeger.
- Martyna, Wendy. 1983. "Beyond the He/Man approach: The case for nonsexist language". In B. Thorne, C. Kramarae, & N. Henley. (eds), *Language, Gender and Society*, 25–37. Rowley, MA: Newbury House.
- Mazuka, Reiko & Lust, Barbara. 1994. "When is an anaphor not an anaphor?" In Barbara Lust, Gabriella Hermon, Jaklin Kornfilt (eds), *Syntactic*

- theory and first language acquisition: Cross-linguistic perspectives, Vol. 2: Binding, dependencies, and learnability*, 145–175. Hillsdale, NJ: Erlbaum
- McConnell-Ginet, Sally. 1979. "Prototypes, Pronouns and Persons". In M. Mathiot (ed.), *Ethnolinguistics: Boas, Sapir and Whorf revisited*, 63–83. The Hague: Mouton.
- McKee, Cecile. 1992. "A comparison of pronouns and anaphors in Italian and English acquisition". *Language Acquisition* 2: 21–54.
- Meyers, Miriam. 1990. "Current generic pronoun usage". *American Speech* 65: 228–237.
- Miller, Casey & Swift, Kate. 1980/1988. *The Handbook of nonsexist writing*. NY: Harper 2nd ed.
- Monville-Burston. 1985. "Individuation: Its role in clitic sequences in French". In Larry D. Kind & Catherine A. Maley (eds), *Selected papers from the XIII Linguistic Symposium on Romance Languages*.
- Moulton, Janice; Robinson, George M & Cheris, Elias. (1978) "Psychology in action: Sex bias in language use: Neutral pronouns that aren't". *American Psychologist*. 33: 1032–1036
- Mühlhäusler, Peter & Harré, Rom. 1990. *Pronouns and People: The linguistic construction of social and personal identity*. Oxford: Blackwell.
- Mufwene, Salikoko. 1984. "The count/mass distinction and the English lexicon". In David Testen, Veena Mishra, and Joseph Drogo (eds), *Lexical Semantics*. Chicago: Chicago Linguistic Society.
- Mufwene, Salikoko. 1986. "Number delimitation in Gullah". *American Speech*. 61.1: 33–60.
- Newman, Michael. 1992. "Pronominal disagreements". *Language in Society* 21: 447–475.
- Newman, Michael. 1996. *Epicene pronouns: The linguistics of a prescriptive problem*. NY: Garland.
- Newmeyer, Frederick. 1978. "Prescriptive grammar in reappraisal". In W. McCormack, & S. Wurm (eds), *Approaches to Language*, 581–593. The Hague: Mouton.
- Pinker, Steven. 1994. *The language instinct*. NY: Morrow.
- Nilsen, Alice 1984. "Winning the great HE/SHE battle". *College English* 46: 151–157.
- Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. Chicago: University of Chicago Press.

- Rigau, Gemma. 1986 "Some remarks on the nature of strong pronouns in null-subject languages". In Inwonne Bordelais, Helas Contreras, Karen Zagana (eds), *Generative studies in Spanish syntax*, 143–163. Dordrecht: Foris.
- Roberts, Craige. 1989. "Modal subordination and Pronominal anaphora in discourse." *Linguistics and Philosophy* 12: 683–721.
- Roberts, Craige. 1995. "Anaphora in intensional contexts". In Shalom Lappin (ed.) *Handbook of Semantics*. London: Blackwell.
- Robblee, Karen E. 1993. "Individuation and Russian Agreement". *Slavic and East European Journal*,
- Sidner, Candace. 1981. "Focusing for the Interpretations of Pronouns". *Computational Linguistics* 7: 217–231.
- Stanley, Julia. 1978. "Sexist Grammar". *College English* 39: 800–811.
- Timberlake, Alan. 1975. "Hierarchies in the genitive of negation". *Slavic and East European Journal* 19: 123–138.
- Utakis, Sharon. 1995. *Quantification and definiteness in Child Grammar*. Doctoral Dissertation NY: Graduate Center, CUNY.
- Utakis, Sharon. 1997. Pronouns and definiteness in child language. In A. Sorace, C. Heycock, & R. Shillcock (eds), *Proceedings of GALA '97 — Conference on Language Acquisition*, 178–183. Edinburgh, Scotland: Human Communications Research Centre, University of Edinburgh.
- Valian, Virginia. 1977. "Linguistics and feminism". In M. Vetterling-Bruggin, F. Elliston, & J. English (eds), *Feminism and Philosophy*, 154–166. Totowa NJ: Littlefield, Adams.
- Van Hoek, Karen A. 1992. *Paths through conceptual structure: Constraints on pronominal anaphora*. Doctoral dissertation. University of California at Santa Barbara.
- Webber, Bonnie. 1980. "Syntax beyond the sentence: Anaphora". In R. Spiro, B. Bruce, W. Brewer (eds), *Theoretical issues in reading comprehension*, 141–164. Hillsdale, NJ: Erlbaum.
- Weidmann, Urs. 1984. "Anaphoric *they* for singular expressions". In R. Watts & U. Weidmann (eds), *Modes of interpretation: essays presented to Ernst Leisi on the occasion of his 65th birthday*, 59–68. Tübingen: Narr.
- Wiese, Berndt. 1983. "Anaphora by pronouns". *Linguistics* 21: 373–417.
- Whitley, M.S. 1978. Person and number in the use of *we*, *you*, and *they*. *American Speech* 53: 18–39.