



HAL
open science

Gender/sex discrepancies in pronominal references to animals: a statistical analysis

Laure Gardelle

► **To cite this version:**

Laure Gardelle. Gender/sex discrepancies in pronominal references to animals: a statistical analysis. English Language and Linguistics, 2013, 17 (1), pp.181-194. 10.1017/S1360674312000366 . ensl-00778467

HAL Id: ensl-00778467

<https://ens-lyon.hal.science/ensl-00778467v1>

Submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Gender / sex discrepancies in pronominal references to animals:
a statistical analysis**

Laure Gardelle

*Ecole Normale Supérieure de Lyon (University of Lyon, France)/
UMR ICAR (CNRS section 34)*

ABSTRACT

Although the English gender system is a semantic system largely based on sex, it is well known that in references to animals there is widespread discrepancy between pronominal gender and sex, and that gender selection is dependent on speaker's point of view (degree of interest in the animal, projection of personality and so on). What is yet to be established, however, is whether point of view still prevails in references to animals when the antecedent noun specifies the sex of the referent (e.g. *stallion*, *ewe*). In that case the neuter is known to occur but there is no quantitative assessment of the phenomenon, although it is crucial to understanding the influence of sex on gender selection. This paper therefore proposes a statistical analysis of gender use in personal pronouns focusing exclusively on cases in which the antecedent noun specifies the sex of the animal. The analysis is carried out at the scale of a multi-million word *corpus of Contemporary American English* (COCA), using Pearson's chi square test complemented by the odds ratio estimate. Three questions are considered: how common is the neuter? Is its relative frequency the same with female animals as with males? Finally, do the proportions vary according to the position of the anaphor relative to its antecedent?

English pronominal gender¹ is a semantic, as opposed to formal, system in which the sex of the referent is a major determining factor (Corbett 1991: 12). As Huddleston & Pullum (2002: 488) state, '[i]n the most straightforward cases, *he* is used for males, *she* for females, and *it* for entities that are neither male nor female'. For instance *my father*, *one woman* and *my diary* respectively take *he*, *she*, and *it* as their co-referring anaphors. This strong relationship between linguistic gender and biological sex, however, is by no means a one-to-one correlation: it is only with human beings (and even then not babies) that such a correlation is found (ibid.). In references to animals it has been well established since Sweet (1898: 42) that the neuter was not uncommon. For instance Corbett (1991: 12) notes 'a high degree of variability for animals' while Swan (1997: 219) does not include them in his introductory account of gender selection: 'Usually people are *he* or *she* and things are *it*'. This lack of correlation has led linguists to formulate a specific gender choice criterion for references to animals. It varies from one work to another but always pertains to speaker's point of view, with such notions as projection of personality (Sweet 1898: 42, Curme 1931: 551, Quirk et al. 1985: 341, Leech & Svartvik 1994: 56), familiarity (Zandvoort 1965: 132, Biber 1999: 317), animal thought of in its individual aspect (Kruisinga & Erades 1960: 445, Morris 1991: 158), greater degree of interest or empathy (Jespersen 1942: 209, Joly 1987: 234, Biber

¹ More specifically, the gender system we are concerned with applies to the 3rd person singular pronouns *he*, *she* and *it*, including all their forms: thus *he*, *him*, *his*, *himself*, and so on.

1999: 317, Huddleston & Pullum 2002: 489) and relevance or importance of sex (Roggero 1988: 202, Sinclair 1990: 29).

The aim of the present study is to establish whether point of view still prevails in references to animals when the antecedent noun specifies the sex of the referent (e.g. *stallion*, *ewe*). Although it is well established that *it* does not mean ‘neither male nor female’, Huddleston & Pullum (2002: 489) note that ‘[i]n many cases the neuter pronoun is used because the speaker doesn’t know what the sex is, though we may also use *he* or *she* in such contexts’. The question examined here is how common the neuter is when the speaker knows the sex of the animal and specifies it in the antecedent noun – in other words, when the information is part of the categorisation of the referent at the point when the antecedent is uttered. All that is established about this specific case is that a neuter pronoun is possible (Huddleston & Pullum 2002: 488, Gardelle 2006: 109, Siemund 2008: 1). It is attested both in contemporary English, as illustrated by (1), and in earlier modern English, in which the animate genders were more common in references to animals (2):

- (1) (The cow is either admitted for milking, or **it** may be turned away. Siemund 2008: 1) (ICE-GB)
- (2) Sir Jo Finch told us of an exquisite poyson of the D: of Florences that kill’d with a drop : That drawing a threit and needle dipt in it thro a hens thigh **it** perish’d immediatly. (Gardelle 2006: 109) (Evelyn 1661: 190)

However there is no quantitative assessment of the phenomenon. The present article therefore proposes a statistical analysis of gender use in personal pronouns in the specific cases in which the antecedent noun gives the sex of the animal. The analysis is carried out at the scale of a multi-million word *Corpus of Contemporary American English* (COCA). After detailing the data collection procedure (section 1) the study examines three questions in turn (sections 2 to 4). First of all, how common is the neuter in those cases? Secondly, is the relative frequency of the neuter pronoun the same with female animals as with males? Finally, does the position of the anaphor relative to its antecedent affect gender use? In particular, is the neuter disfavoured when the anaphor is part of the same clause or of the same sentence as its sex-denoting antecedent? All the analyses were made using *R* version 2.14.1². The tests carried out are Pearson’s chi square test complemented by the odds ratio estimate. P³-values lower than 0.05 (P < 0.05) were considered to be statistically significant.

1. DATA COLLECTION PROCEDURE

The corpus used for the present study is the *Corpus of Contemporary American English*, which when the data was collected (March to May 2011) contained 176,389 texts (425 million words) spanning the period 1990–2011 (20 million words per year). The corpus

² *R* is an environment for statistical computing initially developed by Robert Gentleman and Ross Ihaka and now expanded by a core group of contributors and other researchers from all over the world. The r-project server (< <http://www.r-project.org/> >) is hosted by the Institute for Statistics and Mathematics of the WU Wien (Vienna University of Economics and Business).

³ The P-value indicates the probability that the distribution observed is not due to chance. A P-value of 0.05 means that there is a probability of 5% that the observed distribution has occurred by chance. It is conventionally considered that P-values have to be lower than 0.05 for a result to be statistically significant.

is designed to be balanced and representative of English language use in the United States. It is equally divided among five genres: spoken (90 million words, although the occurrences are taken from TV and radio programmes, rather than from free conversation), fiction (85 million words, from short stories to movie scripts), popular magazines (90 million words from nearly 100 different publications in a variety of domains, such as news, health or hunting), newspapers (87 million words, from more than 10 different titles) and academic texts (86 million words from about 100 peer-reviewed journals).

In order to draw a list of all the sex-specific nouns denoting animals a systematic search for the words *male* and *female* was carried out in the definition field of the Oxford English Dictionary (2011 online edition).⁴ From the 3,040 occurrences (1,115 for *male*, 1,925 for *female*) the relevant headwords – sex-specific nouns for animals – were extracted. As most headwords specifying sex proved to be for human beings, the process only yielded 63 nouns. That figure does not include compounds with *he*, *she*, *lady* or *man* as their first element, which were not retained. In the case of *lady* and *man* (e.g. *lady elephant*), the possible personification involved in the use of the noun could have biased the results; as for *he* and *she* (e.g. *he-goat*), the OED does not provide a list of entries, so that exhaustiveness could not be achieved. A distinct study of the influence of those modifiers on pronominal gender would have to be carried out. All the potentially sex-neutral nouns (e.g. *duck*, which can denote the species as well as the female of the species) were also discarded and are therefore not included in the count.

The 63 nouns obtained were then searched for in the COCA. 30 of them (such as *cock-lobster*, *spawner* or *vixen*) returned no hits, either for the noun itself or for occurrences of the noun with a co-referential pronoun. As a result the number of nouns considered here is 33 : 16 for males and 17 for females, among which 4 nouns denote sex for more than one species (*buck/doe* and *male/female*)⁵. For each noun, all the hits were read in their context of use in order to determine whether there was a co-referential pronoun. The COCA interface enables an automatic search for collocates, but within nine words to the left or right only, which could have restricted the study. The procedure was therefore carried out manually, and although the mean distance between the anaphoric pronoun and its antecedent was eventually found to be 6.16 words, there were indeed 165 relevant occurrences of pronouns further away than 9 words. For reasons of feasibility, when there were well over 1,000 occurrences (it was the case for 19 nouns, for instance 8,563 hits for *buck*), the search was limited to a random 1,000, as permitted by the COCA interface. A further restriction was imposed on *male* and *female*. Their use as heads of noun phrases being extremely uncommon compared to modifier uses, two samples of 1,000 utterances failed to show any occurrences with co-referential pronouns. An automatic search for the two nouns with collocating pronouns was therefore carried out, but none were found to be co-referential in references to animals. It was consequently decided to restrict the search by imposing a determiner. A sample with *a male/a female* only showed results in compounds or for human beings, so that

⁴ This procedure was followed because there is no such list available today. Ordan and Wintner (2005) propose a tentative set of sex-specific nouns for their project of multilingual lexical databases for natural gender, but it did not prove exhaustive.

⁵ To these must be added 5 nouns which can denote the male or the female of several species, but which are recorded by the OED as being primarily used for one of them: *cock / hen*, *bull / cow* and *sow* (which the OED defines as the female pig, but which was also found in the COCA for the female boar and the female bear).

the search was eventually restricted to *the male/the female*. The figures obtained for these antecedents were included, but only once a comparison of the corpus data with and without these two nouns showed that the determiner restriction did not bias the results. Finally occurrences of pronouns for dead animals, which occurred in the context of hunting or cooking, were discarded, as the [-animate] feature could have had an additional influence on gender selection.

What can be concluded from the data collection is that in a majority of utterances the nouns under study do not present co-referential pronouns in the COCA, either because the referent is not mentioned later on or because subsequent reference does not involve a personal pronoun. Out of the 23,519 utterances examined, only 734 pronominal references were found with sex-specific antecedent nouns for living animals: 413 for males (including 29 for *the male*) and 321 for females (including 18 for *the female*). The figure, however, is high enough for reliable statistical analyses.

Another initial finding is that most of the nouns for which there are hits share a morphological characteristic: they are not morphologically related to a noun that denotes the other sex (e.g. *bull, ewe*). Only two nouns are derivatives (*tigress, leopardess*) and six are compounds sharing one element with the noun for the other sex (e.g. *cock pheasant/hen pheasant*). Because the number of occurrences for derivatives and compounds is too low for reliable statistics (7 and 10 occurrences respectively), it has not been possible to establish whether the morphological pattern of the sex-specific noun had an influence on gender selection. A specific study of this variable would need to be carried out based on a different corpus.

2. ANALYSIS OF THE DISCREPANCIES BETWEEN PRONOMINAL GENDER AND SEX

The data shows that the neuter is far from marginal: it is found in 17.7 % of pronouns (table 1).

	Number of occurrences	Proportion
<i>he/she</i>	604	82.3 %
<i>it</i>	130	17.7 %

Table 1 – Gender distribution in the corpus

The first conclusion to be drawn is that the sole criterion of sex is clearly insufficient to account for gender use even when the antecedent noun specifies the sex of the animal. This suggests that in those cases, as with other nouns, point of view has a major influence on gender selection. For instance one reads:

(3) The sow lifts its snout towards him and exhales sharply. (*Southern Review* 29: 2, 1993)

(4) It goes for the largest male in the troop first. The male rears back on its hefty tail and gives some nasty blows with its powerful hind... (*Analog Science Fiction & Fact* 131: 1/2, 2011)

For 3 of the nouns under study the animate genders and the neuter are even found in similar proportions: *cock* (11 *he*, 10 *it*), *ram* (5 *he*, 6 *it*) and *stag* (12 *he*, 12 *it*).⁶

⁶ Conversely 11 out of the 33 nouns show no occurrences of *it*. It cannot be concluded, however, that these nouns exclude the neuter in their co-referential anaphors. 9 of them only exhibit 1 to 2 occurrences of pronouns (*billy-goat, nanny-goat, peahen, vixen, drone, cock-pheasant, hen pheasant, drake* and

It could be argued that for some nouns such as *stallion*, a small minority of speakers might not know that the nouns denote sex and that as a consequence, they might use them as though they were sex-neutral. This, however, cannot account for the extent of the discrepancy between pronominal gender and sex in the corpus. For instance, speakers using *cock* (11 *he*, 10 *it*) can be expected to be aware that the referent is a male.

One question raised by table 1 is whether the ratio of animate genders is higher with sex-specific antecedent nouns than with sex-neutral nouns. A major practical problem is that for the latter it is impossible to isolate the references in which the sex of the animal is known to the speaker: the context rarely provides clues to establish that. Still, a comparison of gender distribution between the two types of nouns puts the findings given in table 1 in perspective. Rather than consider all the sex-neutral nouns of English, which would be far beyond the scope of the present study, I decided to consider only those with sex-specific counterparts in the corpus, so as to compare data for the same species. In other words, pronominal anaphors for *cock-pheasant*, *ewe* and so on were compared with those of *pheasant*, *sheep* and the like.⁷ For the sex-neutral nouns the figures for each gender were collected in the COCA following the same procedure as that detailed in section 1. The results are as follows:

	Sex-specific nouns	Sex-neutral nouns
<i>he/she</i>	102	35
<i>it</i>	34	45
<i>TOTAL</i>	<i>136</i>	<i>80</i>
<i>Proportion of animate genders</i>	<i>75%</i>	<i>43.8%</i>

Table 2 – *Gender distribution with sex-specific antecedent nouns and their sex-neutral counterparts*

Table 2 shows that for animals of the same species the animate genders are used in 75% of pronouns when the antecedent noun is sex-specific, against 43.8% when the noun is sex-neutral. This difference in proportion is statistically significant ($P < 0.001$). The odds ratio estimate shows more precisely that the animate genders are 3.857 times more likely when the antecedent noun is sex-specific (95% CI = 2.142 to 6.944, $P < 0.001$ ⁸). It should not necessarily be concluded from these figures that knowing the sex entails a higher chance of an animate gender: use of a sex-neutral noun does not mean that the sex of the referent is not known, as illustrated by (5):

bullock), 1 only 6 occurrences (*tigress*). The last one, *female* (19 *she*) exhibits a higher number of associated pronouns, but a Google search among pages in the United States using the phrase '*female and its*' shows that the string is to be found in American English. Extension of the Google search to the other 10 nouns leads to the same conclusion.

⁷ Cases in which the counterpart of the sex-specific noun can be either sex-neutral or sex-specific (e.g. *leopard*, which can denote any member of the species but also just the male) were not considered.

⁸ CI, or 'confidence interval', establishes whether the odds ratio is reliable. '95% CI' indicates that if the same study were repeated 100 times with 100 different corpora of the same size, the odds ratio would be projected to be within the range given afterwards – here between 2.142 and 6.944 – 95% of the time. The odds ratio estimate is considered reliable if the 95% CI is well above or below 1: if the data were randomly distributed, the odds ratio would be expected to have a value of 1.

(5) (*Outdoor Life* 189: 3, 1992) I had just carefully closed my truck door when a turkey gobbled⁹ from its roost tree close by.

What can be concluded, however, is that use of a sex-specific noun when a strictly sex-neutral one exists increases by 3.857 the likelihood of an animate gender in the anaphor.

Coming back to the set of sex-specific nouns, another question to be answered is whether the proportion of neuter decreases when the noun that denotes sex is not species-specific. Indeed in that case the sex information could be expected to be more salient because there is less other information. The nouns concerned in the corpus are *buck* and *doe*, which do not specify one species but are restricted to a subset of animals, and *male* and *female*, which only denote sex.

	<i>Buck</i>	<i>Doe</i>	<i>Male</i>	<i>Female</i>
<i>he/she</i>	26	44	26	18
<i>it</i>	6	11	3	0 ¹⁰
TOTAL for 'it'	17/87 = 19.5%		3/47 = 6.4%	

Table 3 – Gender distribution with sex-denoting antecedent nouns that are not species-specific

It is not found to be less common for *buck* and *doe* than it is in the overall corpus. The proportion is even slightly higher (19.5% against 17.7 %), although the difference is not statistically significant ($P = 0.785$). With *male* and *female* the proportion of neuter seems much lower (6.4% vs. 17.7 %), but given that the data for these nouns is restricted to *the male* and *the female*, the figure must be assessed in comparison with *the* + [other nouns] only. The data is given in Table 4.

	<i>The male/the female</i>	All other instances of <i>the</i> + <i>N</i>
<i>he/she</i>	44 (26 males, 18 females)	338 (172 males, 166 females)
<i>it</i>	3 (3 males)	66 (46 males, 20 females)
TOTAL	47	404
Proportion of neuter	6.4%	16.3%

Table 4 – Gender distribution when the antecedent is a definite description

⁹ The verb *gobble* signals that the referent is a male. It is defined as follows by the OED: 'Of a turkey-cock: To make its characteristic noise in the throat'. The speaker in this extract can be reasonably assumed to know this: he is a turkey hunter, and one common ploy used by hunters is to imitate the clucking of the hen in order to attract the male (very often referred to as the 'gobbler'). It can be noted that in the sentence following this extract the speaker uses *he*: 'The bird's timing was such that I think he heard the faint click as the door shut.'

¹⁰ Although *the female* was not found in combination with *it* in the corpus, Gardelle (2006: 175) and a Google search for the string *the female and its* among pages in the United States show that the neuter is attested with this antecedent.

The neuter pronoun occurs in 6.4% of cases with *the male* or *the female* as antecedent NP against 16.3 % with other definite descriptions. This difference, however, is not statistically significant ($P = 0.073$). If one isolates references to males the differences in proportions (10.3% against 21.1%) are even more clearly found not to be significant (Table 5 – $P = 0.172$):¹¹

	<i>the male</i>	All other instances of <i>the + N</i> for males
<i>he/she</i>	26	172
<i>it</i>	3	46
TOTAL	29	218
<i>Proportion of neuter</i>	10.3%	21.1%

Table 5 – Gender distribution for males when the antecedent is a definite description

It must therefore be concluded that at least when the antecedent is a definite description, the fact that the noun only gives information about sex, rather than sex and the species, does not have a statistically significant influence on gender selection in the anaphor.

3. REFERENCES TO MALES COMPARED WITH REFERENCES TO FEMALES

As stated in section 2, the only three animals for which the neuter and an animate gender were found in similar proportions were all males (antecedent nouns *cock*, *ram* and *stag*) and there are 3 occurrences of *it* with *the male* against none for *the female*. Thus what needs to be examined now is whether gender distribution differs for males and females. The question is all the more relevant as when the sex of an animal is not specifically known to the speaker the animate gender typically used is the masculine (Gardelle [2006: 542])¹². This is true whatever the size of the animal – a mosquito, a rat, a turtle or a giraffe for instance –, and exceptions are restricted to a very small number of referents – mainly cats (for some speakers only) and animals viewed specifically as prey. This would tend to suggest that in the perception of animals the female sex is viewed as more marked than the male; the neuter might therefore be disfavoured for females.

Gender distribution according to sex is as follows:

	Males	Females
<i>he/she</i>	319	285
<i>it</i>	94	36
TOTAL	413	321
<i>Proportion of neuter</i>	22.8%	11.2%

Table 6 – Gender distribution according to sex

The neuter is indeed more common for males. It occurs in 22.8 % of references to males against 11.2% for females, and the difference in proportion is statistically significant (P

¹¹ The analysis cannot be carried out specifically for females because there are no occurrences of *it* with *the female* in the corpus.

¹² The finding does not include children's stories, where the fact that most or all of the characters are animals leads to more variety in the sex chosen.

< 0.001). The neuter is 2.33 times more likely to be used for a male than for a female (OR=2.333; 95% CI: 1.534 to 3.536 ; P < 0.001).

Because the data with the antecedent nouns *male* and *female* is restricted to occurrences with *the*, the analysis was carried out again setting aside the occurrences for these nouns in order to see whether the finding was confirmed. The results are given in table 7:

	Males	Females
<i>he/she</i>	293	267
<i>it</i>	91	36
TOTAL	384	303
<i>Proportion of neuter</i>	<i>23.7%</i>	<i>11.9%</i>

Table 7 – Gender distribution – exclusion of the male/the female

The neuter occurs in 23.7 % of references to males against 11.9% for females. This difference in proportion is again found to be statistically significant (P < 0.001), and the neuter is again predicted to be around 2.3 times more likely for males (OR= 2.303 ; 95% CI= 1.513 to 3.506 ; P < 0.001). It must therefore be concluded that there is an asymmetry between males and females.

In keeping with this, if one considers the nouns of the corpus that are part of a pair of strictly sex-specific nouns (e.g. *buck/does*), there is never more neuter for the females than for the males. Leaving aside the cases with very few occurrences (< 3 for both genders), the asymmetry appears especially with the following pairs:

	Neuter	Animate gender
<i>buck</i>	5 (15.6% of occurrences of <i>buck</i>)	27
<i>doe</i>	4 (8.7%)	42
<i>colt</i>	9 (25%)	27
<i>filly</i>	1 (3.4%)	28
<i>ram</i>	6 (54.4%)	5
<i>ewe</i>	3 (20%)	12

Table 8 – Gender distribution among pairs of strictly sex-specific nouns – 1/2

It is less obvious in the other three cases concerned, but the neuter is again less common for females:

	Neuter	Animate gender
<i>bull</i>	6 (23.1%)	20
<i>cow</i>	7 (22.6%)	24
<i>-cock</i>	10 (47.6%)	11
<i>hen</i>	5 (41.7%)	7
<i>-(the) male</i>	3 (10.3%)	26
<i>(the) female</i>	0 (0%)	18

Table 9 - Gender distribution among pairs of strictly sex-specific nouns – 2/2

No explanation other than the markedness of the female sex could be found to account for this asymmetry. If one considers *buck* and *doe*, for instance, most of the occurrences were found in articles on hunting. In other words with speakers who are apt to feel the same level of closeness to both sexes. Moreover the two nouns in each pair belong to the same level of vertical categorisation. Finally, none of the nouns denoting females appear more specialised or uncommon than those denoting males. Although the figures for each pair are low, they appear to confirm that sex has more influence on gender selection for females than for males.

4. POSITION OF THE ANAPHOR RELATIVE TO ITS ANTECEDENT

This last section examines whether the animate genders are more highly favoured when the anaphor is close to the antecedent. The question is relevant for two reasons. First of all, in same-clause contexts the anaphor bears a stronger grammatical relationship to its antecedent than in other contexts. It is part of the same constituent (the clause) and it is very often bound by its antecedent (in the sense given by Binding Theory: c-commanded by the antecedent and co-referential with it). The hypothesis to be tested is whether this favours an animate gender in the anaphor. The second reason is a cognitive one. The proximity of the antecedent might have a priming effect on the sex of the referent, which might disfavour the neuter. Furthermore, Relevance Theory (Sperber and Wilson 1986) has shown that mentioning an item of information in context implies that it is relevant to the speaker. Choice of lexical information in discourse is guided by the Optimal Relevance principle, which is the result of a balance between processing cost and contextual effect (viz. the triggering of implications). As a result, any mention of information implies that the information is deemed worth the processing cost at the point when it is uttered. If this is applied to gender, the fact that a speaker chooses a sex-specific noun in the antecedent, at least in all the cases in which a sex-neutral word is also available, implies that sex is deemed relevant and important at that stage. When the anaphor is in the same clause as its antecedent, the point of view on the referent has to be the same as it was when the antecedent was uttered: the anaphor and its antecedent belong to the same event description. In other words sex still has to be relevant and important to the speaker when the pronominal gender is selected. Consequently the neuter might be expected to be disfavoured in this context.

This section first considers the overall distribution of genders according to position and then moves on to the specific distinction between same-clause and different-clause contexts.

	Same clause	Different clause in same sentence	Next sentence
<i>he/she</i>	137	296	168
<i>it</i>	28	76	26
<i>TOTAL</i>	<i>165</i>	<i>372</i>	<i>194</i>
<i>Proportion of neuter</i>	<i>17%</i>	<i>20.4%</i>	<i>13.4%</i>

Table 10 – *Gender distribution according to the position of the anaphor relative to its antecedent*¹³

The neuter occurs in 17 % of anaphors in same-clause contexts, 20.4 % in different clause (same sentence) contexts and 13.4 % of anaphors whose antecedent is in the previous sentence. These differences in proportions are not found to be statistically significant ($P = 0.111$). One notes a slightly lower proportion of *it* for next-sentence contexts compared to the other two. An additional statistical analysis along the sole criterion of sentence boundaries was therefore carried out to establish whether a sentence boundary increased the proportion of neuter pronouns:

	Same sentence	Different sentence
<i>he/she</i>	433	171
<i>it</i>	104	26
<i>TOTAL</i>	537	197
<i>Proportion of neuter</i>	19.4%	13.2%

Table 11 – *Gender distribution according to the position of the anaphor relative to its antecedent – same/different sentence*

The neuter is found in 19.4 % of pronouns in same-sentence contexts against 13.2 % when in different sentences. P is slightly above 0.05 (P [uncorrected] = 0.059, P [corrected¹⁴] = 0.064), so that influence of this variable cannot be established.

As for the specific hypothesis of an influence of same-clause contexts formulated at the beginning of this section, the results are as follows:

	Same clause	Different clause
<i>he/she</i>	137	467
<i>it</i>	28	102
<i>TOTAL</i>	165	569
<i>Proportion of neuter</i>	17%	17.9%

Table 12 – *Gender distribution according to the position of the anaphor relative to its antecedent – same/different clause*

The neuter is found in 17 % of pronouns in same clause contexts against 17.9 % in other contexts. This difference in proportion is not statistically significant ($P = 0.816$).

Finally, in order to assess the possible influence of Optimal Relevance, which would only apply when the speaker could have used a sex-neutral term denoting the same species, an extraction was made of pronouns whose antecedent nouns had a specific sex-neutral counterpart (e.g. *cock-pheasant*, *ewe*), for the sole contexts in which the pronoun and the antecedent were in the same clause. Although the number of occurrences is fairly low (26 occurrences) and might therefore not guarantee reliability, it seems to show that the neuter in same-clause contexts is not any less favoured there than in the corpus as a whole: it is found in 26.9% of occurrences (7 out of 26) against

¹³ To these must be added 2 occurrences of *she* 2 sentences away from the antecedent and 1 of *he* 4 sentences away.

¹⁴ This value is Pearson's chi square corrected from continuity using the Monte Carlo procedure with 10,000 replications.

17 % in the corpus as a whole. Thus there is no evidence that the position of the anaphor relative to its antecedent (same clause / ...) has any influence on gender selection.

5. CONCLUSION

The statistical analysis presented in this paper establishes a number of facts on the relationship between pronominal gender and biological sex in the specific case in which the animal's sex is given by the antecedent noun. Some of the results are expectable in that they confirm the modern theoretical descriptions of gender in English, while others allow for a better understanding of the system.

First of all, even when the sex is specified and part of the categorisation of the referent in context, the neuter is widely used: it accounts for over 1/6 of occurrences in the corpus (17.7 %). The fact that gender and sex do not correlate in references to animals is part of most grammatical descriptions, but what was less expectable perhaps is the extent of this lack of correlation – 17.7% of cases. This figure should be emphasised because as was stated in the introduction, only three works among those cited mention the possibility of the neuter with sex-specific antecedent nouns, and because the notion of a core correspondence between gender and sex when the sex is known to the speaker is still found in some works that mention English gender to compare it with other, formal systems. For instance, to Mülhäusler & Harré (1990: 29), 'English is strongly characterized by natural gender (...) and we can be sure of selecting the correct anaphoric third-person pronoun if we reflect on the biological sex of "man", "cow", "stallion" and so on'. The present study clearly establishes that such a statement is a misrepresentation of the gender system (and not just an oversimplification), even if one considers only sex-specific nouns. By focusing solely on the distinction between sexed beings (animates, that is, humans and animals) and sexless entities (inanimates), the statement neglects the influence of another categorisation principle: the fact that humans are typically ranked higher than animals. This is encapsulated in the Animacy Hierarchy (Corbett 2000: 56):

speaker > addressee > 3rd person > kin > human > animate > inanimate

Use of *he* or *she* for an animal signals that it is promoted by the speaker, while *it* indicates that it retains its basic ranking. In other words, even though *it* is more common with inanimates in English than it is with males and females, and although *he* and *she* are found mostly with males and females, these facts should not be misread as 'for any male you typically use *he* and for any female, *she*': this holds only for human beings.

This is confirmed by two other facts established in this paper. Firstly, there is no statistically significant evidence in the COCA that *male* and *female*, which make the sex information particularly salient (it is their only lexical information), favour the animate genders more than other sex-specific nouns, at least in definite descriptions. Secondly, when the sex is given explicitly by a noun and when the speaker is bound to have the same point of view on the referent upon selection of the anaphor as upon selection of the antecedent (same-clause contexts), there is no statistically significant evidence that the neuter is disfavoured.

The second major finding of this study is that in the corpus, the neuter is 2.3 times more likely for males than it is for females. This fact, which does not seem to be mentioned in existing studies, could be evidence that females are regarded as the marked gender (in the sociological sense of the term, e.g. Romaine 2000) even among

animals. This finding also provides evidence that while sex does not always trigger promotion of the referent (and thus use of *he* or *she*), it can play a part – a further study could help to determine in what contexts. In keeping with this result, the proportion of neuter was found to be lower with sex-specific nouns than with the sex-neutral nouns denoting the same species (when such nouns exist). Although in the latter case it is impossible to distinguish between the utterances in which the sex of the animal is not known to the speaker and those in which it is, the difference in proportion could point to an influence of sex in the promotion of the referent (and thus in the use of *he* or *she*).

A larger sample would allow for finer-grain statistical studies. In particular, these could identify gender behaviour depending on species or speakers' profiles. Given the very low proportion of sex-specific animal nouns with co-referring anaphoric personal pronouns in discourse, however (734 occurrences in a 425-million-word corpus), this was impossible to achieve. Further studies on corpora in other varieties of English would also enable us to determine whether the findings reported in this paper are restricted to American English or can be considered valid for the description of pronominal gender in English as a whole.

REFERENCES

- Biber, Douglas et al. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Corbett, Greville G. 1991. *Gender*. Cambridge : Cambridge UP.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge UP.
- Curme, George O. 1931. *A Grammar of the English Language in Three Volumes*, vol.3 : *Syntax*. Boston : D.C. Heath & Co.
- Davies M. 2008-. The Corpus of Contemporary American English: 425 million words, 1990-present. <<http://corpus.byu.edu/coca/>>.
- Evelyn, John. 1983 [1620–1706]. *The Diary*, ed. John Bowle. Oxford: Oxford UP.
- Gardelle, Laure. 2006. *Le genre en anglais moderne (16e siècle à nos jours): le système des pronoms*. Ph.D. dissertation. Université Paris IV-Sorbonne.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge UP.
- Joly, André. 1987. Toward a Theory of Gender in Modern English. In *Essais de systématique énonciative*. Lille: Lille UP, 197–244.
- Jespersen, Otto. 1965 [1942]. *A Modern English Grammar on Historical Principles*, vol.7 : *Syntax*. London: Allen & Unwin.
- Mühlhäusler, Peter & Ron Harré. 1990. *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: Blackwell.
- Kruisinga, Etsko & P.A. Erades. 1960. *An English Grammar*, vol.1 : *Accidence and Syntax*. 8th edn. Groningen: Noordhoff.
- Larrivé, Pierre (ed.). 1997. *La Structuration conceptuelle du langage*. Leuven: Peeters.
- Leech, Geoffrey & Jan Svartvik. 1994. *A Communicative Grammar of English*. 2nd edn. London: Longman.
- Morris, Lori. 1997. The Grammatical Role of English Pronominal Gender. In Larrivé (ed.), 149–70.

- Ordan, Noam & Shuly Wintner. 2005. Representing Natural Gender in Multilingual Lexical Databases. *International Journal of Lexicography* 18(3), 357–70.
- Oxford English Dictionary* online edition, 2nd edition, 1989; online version December 2011. <<http://www.oed.com.sidproxy.ens-lyon.fr/>> (accessed 20 July 2011).
- Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Roggero, Jacques. 1988. *Grammaire anglaise*. Paris: Nathan université.
- Romaine, Suzanne. 2000. *Language in Society: An Introduction to Sociolinguistics*, 2nd edn. Oxford: Oxford UP.
- Siemund, Peter. 2008. *Pronominal Gender in English: A Study of English Varieties from a Cross-Linguistic Perspective*. New York: Routledge.
- Sinclair, John (ed.). 1990. *Collins Cobuild English Grammar*. London: HarperCollins.
- Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Cambridge, MA: Harvard UP.
- Swan, Michael. 1997. *Practical English Usage*, 2nd edn. Oxford: Oxford UP.
- Sweet, Henry. 1898. *A New English Grammar Logical and Historical*. Oxford: Clarendon.
- Zandvoort, R.H. 1965. *A Handbook of English Grammar*, 3rd edn. London: Longman.