Poznań Studies in Contemporary Linguistics 35, 1999 pp. 133-152
© School of English, Adam Mickiewicz University, Poznań, Poland

LANGUAGE CONTACT AND THE FUNCTION OF LINGUISTIC GENDER

PETER TRUDGILL
University of Fribourg

1. Introduction

Most functional approaches to linguistics have in common "the belief that language must be studied in relation to its role in human communication" (Foley and Van Valin 1984: 7).[1] Foley and Van Valin's approach is in particular directed towards explaining grammatical phenomena "in terms of recurrent discourse patterns in human language" (Foley and Van Valin 1984: 13). Givón (1979: 49) similarly argues that "the formal or 'structural' properties of syntax will ... be shown, to quite a degree, to emanate from the properties of human discourse"; and elsewhere he argues that the goals of a functionally oriented linguistics include the study of how grammatical devices "are used in coding and communicating knowledge" (Givón 1984: 10). The goal of this paper is not to question such approaches but to suggest that there may be rather more grammatical devices than is sometimes thought in human languages which, although not dysfunctional (see Lass 1997: 16), may not have any particular function at all in coding and communicating knowledge. There may be, that is, a number of grammatical phenomena that can perhaps be legitimately regarded as being afunctional. We examine this suggestion with particular reference to grammatical gender and to the natural gender of self-reference.

[1] Much of this paper is derived from conversations on this topic over the years with Jean Hannah. I am very grateful for her help and her insight. I am also very grateful to many other people who have supplied corrections, encouragement, references and information, shared ideas with me, and helped with data from and insights into languages other than English, including Sasha Aikhenvald, Lyle Campbell, Andrew Carstairs-McCarthy, Grev Corbett, Bill Croft, Alice Davison, R. M.W. Dixon, Małgorzata Fabiszak, Paul Fletcher, George Grace, Raymond Hickey, Adam Jaworski, Jussi Klemola, Miklós Kontra, Didier Maillat, F. R. Palmer, Klára Sándor, and Theo Vennemann.

2. Grammatical categories

Although there is not total agreement about what the major "grammatical categories" are, or about exact terminology, by relatively common consent (see, for example, Jespersen 1924, Bloomfield 1933, Lyons 1977, Bybee 1985) they would appear to include at least: number, case, tense, aspect, voice, mood, person, and gender. Crucially for this paper, gender can be further subdivided into natural gender and grammatical gender. Natural gender is related to the sex of humans and, often, animals, while grammatical gender is not. Many languages, such as English, have natural gender but not grammatical gender. Note, however, that the distinction between the two types of gender can be a fuzzy one, and that grammatical gender can often be seen as the result of a (perhaps metaphorical) diachronic extension of natural gender down the animacy hierarchy (see Croft 1994). (This view is disputed by Fodor (1959), who also gives an interesting overview of attempts in the 19th century and the first half of the 20th to account for the origins of linguistic gender.)

Some of these categories are more frequent in the world's languages than others. For example, all the languages of the world have the category *person*, but of the other categories some are more and others less common. It is the contention of this paper, moreover, that these categories may also be of different statuses or degrees of importance, and that their functions may be less or more clear. Some categories, that is, would appear to be more central in some way than other categories. In this paper it is argued that the category of gender is a strong candidate for the description of "less central".

Languages differ, of course, in the extent to which certain of these categories are optional or obligatory. And there are also important differences in how the categories may be expressed, the most obvious difference being between morphological (or synthetic) expression, on the one hand, and lexical and syntactic (or analytic) expression, on the other. As is also well known, the same language at different points in its history can change with regard to which categories it possesses, as well as the way in which they are expressed. Old English, for example, expressed case almost entirely morphologically, whereas modern English expresses it to a considerable extent syntactically.

Language contact is often cited as an important factor leading to a change from the synthetic to the analytic expression of these categories. In Trudgill (1978), for instance, I showed that, in a situation involving language contact, language shift and language death, the Albanian dialects of central Greece (Arvanitika) were undergoing the loss of the morphologically expressed optative mood which was now being expressed instead lexically by the use of the originally Greek word *makari* 'would that'. The point would appear to be that imperfect learning by adults of a language in a contact situation can lead to the loss of features which constitute difficulties for learning, and that analytic structures are in some way easier to learn and/or process than synthetic structures. If we are interested in the centrality or importance of the major grammatical categories in the world's languages, it is therefore revealing to

see what happens to them, and their expression, in situations of intense language contact, the most dramatic of which involve pidginisation.

3. Pidginisation and creolisation

Adults and adolescents who are beyond the critical threshold for language learning more or less necessarily subject new languages that they are learning to *pidginisation*. Pidginisation is a process which consists of three major sub-processes (see Trudgill 1996). The first is *reduction*, or, as it is sometimes called, *impoverishment*. The immediate cause of reduction is restriction in function. Since non-native speakers typically use language for a narrower range of purposes than native speakers, there are large areas of a target language which are simply not present in the usage of non-native speakers. That is to say, a language which has been subject to reduction has, as it were, large areas of itself missing: vocabulary will be smaller, grammatical devices fewer, and stylistic alternatives less elaborate. Imperfect learning will also play a role here, however: reduction occurs because learners have not (yet) learnt all there is to learn.

The second sub-component of pidginisation can be labelled *admixture*. As a result of the fact that adults are less than perfect foreign-language learners, a target language will be subject to interference from the learner's native language.

The third component of pidginisation can be labelled *simplification*. Paradoxically, this is a somewhat complex notion, referring for the most part to regularisation and loss of redundancy. It also refers, however, to an increase in *transparency*, by which is meant an increase in forms such as *eye-doctor* as opposed to *optician*, and *did go* as opposed to *went*. Imperfect learning, that is, leads to the removal of irregular and non-transparent forms which naturally cause problems of memory load for adult learners, and to loss of redundant features. This can in turn lead to an often dramatic increase in analytic over synthetic structures.

In some particular sociolinguistic circumstances, pidginisation may lead to the development of a *pidgin language*. What is required for the development of a pidgin language is, firstly, for the degree of reduction, admixture and simplification to be rather extensive. Typically, mutual intelligibility with the target language will be lost. Secondly, the considerably pidginised target language will acquire a relatively stabilised or *focused* form which will eventually be susceptible to learning by future speakers and to description by linguists.

Thus, a pidgin language, relative to the original target language, is reduced, mixed and simplified and has undergone a process of focusing, whereby it has acquired a fixed, stabilised form. As is well known, a pidgin language will not be adequate for all the needs of a native speaker because of the process of reduction which it has undergone, but, as is also well known, this is of no consequence, since pidgin languages, by definition, do not have native speakers.

Subsequently, in certain rather unusual sociolinguistic circumstances, a further development may take place, i.e. a pidgin language may be subjected to *creolisation*. Creolisation is a process which, as it were, repairs the reduction which a pidgin language undergoes during the course of pidginisation. This repair process can be la-

belled *expansion*. Expansion may occur, less or more rapidly, when the pidgin language, originally employed simply as a non-native *lingua franca*, becomes more and more important as a means of communication within a particular community. This is what happened, for example, at least according to one scenario favoured by some creolists, in the case of West African Pidgin English as a result of the transatlantic slave trade. Slaves of West African origin transplanted to the Americas found that West African Pidgin English was the only language that they had in common, and children born into slave communities thus subjected the original pidgin language to expansion in order to render it viable as a means of dealing with all the functions that a native speaker requires of a language. That is to say, the vocabulary was expanded, there was an increase in the grammatical devices available, and stylistic differentiation was increased. A *creole language*, as a pidgin language which has undergone expansion, is thus a perfectly normal language from the point of view of its native speakers, but it demonstrates admixture and simplification with reference to its historical source language. (There are also a number of languages in the world which merit the description *creoloid*. These are languages such as Afrikaans which, as a result of language contact, demonstrate considerable amounts of admixture and simplification as compared to their source language (Dutch in the case of Afrikaans), but which have maintained a continuous native speaker tradition and have never been pidgins – see Trudgill 1996.)

4. Pidgins, creoles and grammatical categories

As a result of simplification and reduction, pidgin languages usually lack all morphology: there are usually no cases, numbers, tenses, aspects, moods, voices, persons or genders that are morphologically marked. Person can be signalled by pronouns or nouns; case can be pragmatically deduced or, perhaps, marked by word order; location of events in or through time can be indicated if necessary through adverbs, as can modality. Interestingly, though, there is typically no role at all for grammatical gender – no pidgin language in the world has grammatical gender – and only a very small role, except lexically, for natural gender. Even lexically, the labelling of natural gender is confined to a few basic terms such as *man* and *woman*, and marking is then achieved, if required, on other nouns in a semantically transparent way by compounding e.g. Tok Pisin *hosman* 'stallion', *hosmeri* 'mare', where *hos* means 'horse', *man* means 'man', and *meri* means 'woman'. This, then, is a preliminary indication suggesting that in some way gender may be a grammatical category which is less central than most or all others.

Note further, then, what happens during creolisation. The expansion process inherent in creolisation – the "repair" of the reduction of pidginisation – involves the reintroduction of many of the grammatical categories that have been lost during pidginisation. We have seen that a pidgin is not adequate for all the needs of a native speaker. We can therefore assume that native speakers "need" some of these grammatical categories and hence reintroduce them. Creole languages thus typically have (optional) aspect and tense markers such as the Sranan particles *ben* 'past' and *de* 'habitual', which are preposed to the main verb. They typically have (optional) plural markers, such as Jamaican Creole *dem* which is postposed to nouns. They continue the pidgin practice of signalling person by pronouns. Case is marked by word order, even for pronouns: Jamaican Creole has *mi* 'I, me'; *im* 'he, him, she, her', *wi* 'we, us', *dem* 'they, them'. Mood can be signalled lexically, as in Arvanitika (see above).[2] And, as far as the active-passive voice distinction is concerned, "creoles usually express passive voice with constructions that correspond semantically but not syntactically to the passive construction found in varieties of English" (Holm 1997: 71). Thus many Caribbean English-based creoles, for example, permit constructions such as *De chicken eat* which are ambiguous as to whether the chicken is eating or being eaten and where the meaning has to be deduced pragmatically.

However, and vitally for our discussion, there is not a creole language in the world which has reintroduced, during the expansion process, the category of grammatical gender on nouns or verbs.[3] Typical creoles, moreover, do not even demonstrate natural gender. For example, in Sranan, the third-person singular pronoun *en* 'he, him, his, she, her, hers, it, its'. A very strong inference is that gender is a category that languages and their speakers can more readily do without than many or most other categories.

5. Gender marking systems

The way in which both natural and grammatical gender are marked in the world's languages varies considerably. One obvious way is through pronouns. Greenberg showed (1966) that there are hierarchies and implicational universals in the expression of natural gender in pronoun systems. Some languages, like Hungarian and Finnish, have no gender-marking on pronouns at all. Others, like English, have gender only in the third-person singular. Others, such as French, have it also in the third person plural – but there are no languages which express gender in the third-person plural but not in the singular. A smaller number of languages also have gender in the second person, where there may also be complications involving T and V pronouns: Polish, for example, has gender marking for the second-person V quasi-pronouns *pan/pani* (which are actually third-person forms in origin) but not for T pronouns; Spanish has gender-marking only in the second-person plural T pronoun but not for the V pronoun, and not at all in the singular. Yet other languages may have gender marking in the first person. Some have this feature only in the first-person plural, such as Spanish *nosotros* versus *nosotras*. Others have it also in the first person singular (see below). Here again there are implicational universals: if

[2] Most of the well-known creole languages derive from languages in which mood plays only a small role, and we can therefore deduce little from the fact that, for example, French creoles have no subjunctive.

[3] One apparent exception, pointed out to me by Sasha Aikhenvald, is Kituba, which is derived from KiKongo, and which does have noun classes (see Stucky, 1978). However, Kituba appears to have arisen out of contact between closely related languages which all have numerous noun classes, and is probably in any case more properly to be described as a creoloid (see Trudgill, 1996). Even here, moreover, we can note that the number of noun classes has been reduced.

gender is marked in the first person it will also be marked in the second or third, but not necessarily vice versa (Greenberg 1966: 96).

Gender marking can also be effected through articles and adjectival agreement in the noun phrase, as in French. And it can be effected through finite verb forms, as for example in past tense and conditional verb forms in some Slavic languages: Russian, Ukrainian and Belorussian have gender marking in these verb forms in the singular; Polish has it in both the singular and the plural; and modern forms of Kashubian (Stone 1993) also have it in the singular and the plural.

In languages which have both grammatical and natural gender, there may be interesting complications in gender-marking systems concerning the relationship between the two. We can even find interesting differences between two forms of the same language. Norwegian Nynorsk, for example, uses the pronouns *han, hon* and *det* for all nouns which are respectively grammatically masculine, feminine and neuter, in the manner of German. Norwegian Bokmål, on the other hand, uses the equivalent forms *han* and *hun* only for natural gender – for animates, especially humans – while *den* is used for grammatical gender to pronominalise nouns which are grammatically masculine or feminine but not animate. Thus Nynorsk *han* can be translated as 'he' and 'it', while Bokmål *han* can be translated only as 'he'. There is similarly the interesting fact about Polish (see below) that in both nouns and verbs male humans form a distinct gender in the plural only, while female humans are grouped together with non-humans and inanimates. And there are other well-known problems of gender resolution, such as the clash between grammatical and natural gender in cases such as German *das Mädchen* 'the girl', which is grammatically neuter but semantically feminine. For a full survey of these issues, see Corbett (1991).

6. The function of natural gender

It is the thesis of this paper that the function of linguistic gender is to a considerable extent obscure. However, it must be conceded that the function of natural gender is a good deal less puzzling than that of grammatical gender (see below). It is much less surprising that human languages have gender distinctions for human beings than that they have grammatical gender, since the distinction between male and female is the most fundamental one there is between human beings, and it is therefore presumably often important to know if a man or a woman is being talked about. Maybe we can, too, make the same point about other, non-human animate beings.

Three functions for natural gender have been discussed in the literature:

1 –The primary function of natural gender is presumably that it can be helpful in making clear the sex of a third person where this has not been previously established. This may be particularly useful (or on some occasions problematical!) in cases where nouns such as *friend, person, teacher* are not marked for gender. Compare the following examples:

(1)	*Meine*	*Freundin*	*ist*	*gestern*	*angekommen.*
	my-f	friend-f	is	yesterday	arrived.
	Sie	*ist*	*glücklich.*		
	She	is	happy.		
(2)	*My friend arrived yesterday. She is happy.*				
(3)	*Ystäväni*	*saapui*	*eilen.*		
	friend-my	arrived	yesterday.		
	Hän	*on*	*onnellinen.*		
	S/he	is	happy.		

In the first, German, example, we know from the very beginning that the friend is female. The noun *Freundin* bears the female morphological marker *-in*, and the possessive pronoun *meine* is also feminine in form. The pronoun *sie* 'she' is thus superfluous. In example (2), on the other hand, it is not clear in English what sex the friend is until we get to the pronoun *she*. And in (3), Finnish, it is never clear what sex the friend is. Other languages of course have yet other ways of doing this.

(4)	*I*	*fili*	*mou*	*eftase*	*xtes.*
	the-f	friend-f	my	arrived	yesterday.
	Ine	*eftixismeni.*			
	Is	happy-f.			
(5)	*Moja*	*koleżanka*	*przyjechała*	*wczoraj.*	
	my-f	friend-f	arrived-f	yesterday.	
	Jest	*szczęśliwa.*			
	Is	happy-f.			

In (4), modern Greek, which is a pro-drop language, does not employ a pronoun at all but signals feminine gender from the very first definite article. In (5), Polish, there is the additional information in that gender is also marked on the verb.

2 –The other important function would appear to be the disambiguation and reference tracking function which operates in certain constructions, as in sentences such as *John kissed Mary and then he ran away*, as opposed to *John kissed Mary and then she ran away* (cf. Comrie 1988).

> Human referents are by far the most frequent in text frequency studies. For this reason, humans are further subclassified in order to aid in identification and tracking of multiple human referents in discourse. The primary salient distinction among human beings is sex. This is manifested ... in the very common distinction of masculine and feminine genders in noun class systems (Croft 1994: 162).

Many languages, however, such as Hungarian, Turkish and Finnish and most creoles, have no gender in their pronominal systems and manage perfectly well without. What do languages such as Finnish do in these cases? The answer is blind-

ingly obvious. They do the same as English does in the case of sentences such as *John kissed Bill and then he ran away*, where the tracking function is of no help at all and where, in the absence of a switch-reference system (see Foley and Van Valin 1984), one has to say, if one wants to be clear, *John kissed Bill and then Bill ran away*. In Finnish, for example, these two sentences would run as follows:

(6) *Jussi suuteli Marjaa ja juoksi sitten pois.*
John kissed Mary and ran then away.

(7) *Jussi suuteli Marjaa ja sitten Marja juoksi pois.*
John kissed Mary and then Mary ran away.

Other strategies can also be employed in such languages, such as Hungarian:

(8) *János megcsókolta Marit és elfutott.*
John kissed Mary and ran-away.

(9) *János megcsókolta Marit aki elfutott.*
John kissed Mary, who ran-away.

3 –Corbett also tells us that gender systems can have the function of indicating speaker attitudes. Rothstein (1993: 697) confirms that in Polish "most personal nouns can be 'depersonalised' [e.g. masculines can be made neuter] for emotional effect, usually pejorative" (see also Dressler et al. 1998, Lewandowska-Tomaszczyk 1992). This is presumably, however, a relatively minor and derivative sociolinguistic function.

The widespread though not universal expression of natural gender in the third person in the world's languages is thus unsurprising. If someone is not present, it may often be useful to know if they are male or female.

7. The function of grammatical gender

Much more puzzling is the function of grammatical gender. It is very interesting indeed to observe that, of the 323 pages of text in Corbett's enormously erudite and stimulating book *Gender* (1991), 321 are devoted to the origins, nature and workings of gender systems, and only 2 to the function of grammatical gender. This is not surprising: it is not at all clear what gender is for. As Hickey (forthcoming) says: "Grammatical gender ... is largely semantically redundant". The function of noun classes in general and of grammatical gender in particular in human languages is actually largely obscure.

Corbett asks: "why do languages have gender systems?". From a Chomskyan perspective, Alice Davison (p.c.) suggests that, since gender is spread out over syntactic constituents such as noun phrases, gender agreement may aid in processing strings of words into syntactically coherent phrases, thereby aiding interpretation. Features such as gender may also be crucial in cross-referencing predicates with subjects: "if you can identify what is linked by the agreement relation, you have solved a number of computations about what the structure of the sentence is". (A similar point is made by Fodor (1959: 206), who points to the function of gender (and other forms of) concord in helping with parsing in languages with free word order, particularly in literary style, as in the classical Latin of Ovid: *lurida terribiles miscent aconita novercae*.) On the other hand, gender still clearly constitutes an additional complication at the level of production.

Aikhenvald (1998) has also shown that in languages – and there are many such – like Manambu (East Sepik Province, Papua New Guinea) in which gender is related in nonhumans to size and/or shape, gender can have a minor semantic function. In Manambu, gender assignment to nonhuman animates is based on their size, so that "large animals belong to the masculine gender, and smaller animals belong to the feminine gender", while in inanimates it is based on their size and shape, so that "long and/or large objects are treated as masculine, and small and/or round ones as feminine". However, there are a number of cases in which the same noun can be either masculine or feminine, depending on its size or other characteristics. Thus *val* 'canoe' "is masculine if big, feminine if small". (Lyle Campbell (p.c.) has speculated that this type of phenomenon, like the unusual Polish plural gender differentiation system already mentioned, may ultimately be linked in some way to animacy and agentivity: just as animates make 'better agents' than inanimates, so large objects may have been perceived as making better agents than small, and, in a sexist society, men better agents than women.)

Another minor function is identified by Fodor (1959: 206), who suggests that gender "lends itself to the purposes of animation, sexualisation and personification in literature" and cites a Russian folk song where a rowan tree *r'abina* (feminine) is yearning for an oak tree *dub* (masculine).

The most common answer that has been given, however, as Corbett points out, is that, as with natural gender marking, gender systems help with disambiguation and reference tracking. Foley and Van Valin (1984) give extensive evidence showing the importance of gender in languages in which "gender functions as the dominant system of discourse cohesion" (Foley and Van Valin 1984: 326). Heath (1975) argues that there is an inverse relationship between the number of verbal means – such as switch-reference, passive and anti-passive – for reference tracking in a particular language, and the number of nominal gender classes, the point being that the more you have of the one, the less you need of the other. And Lyons (1977: 288) also writes "it is clearly the pronominal function of gender which is of primary importance in communication". Thus, for example, the German sentence (Zubin and Köpcke: 1981) *Der Krug fiel in die Schale, aber er zerbrach nicht* is not ambiguous (as the corresponding English translation *The jug fell into the bowl but it didn't break* would be) because the two nouns are of different genders. One cannot help wondering, however, whether this function is, as it were, "worth it". After all, as with the solution to the "problem" of the absence of natural gender marking in Finn-

ish, it is not an enormous effort to say, in English, *The jug fell into the bowl but the jug didn't break.* In what sense does the (one has to assume) occasional German sentence such as the above "justify" the wealth of morphological complexity demonstrated by the German gender system, particularly in view of the fact that the disambiguation only works anyway if the two nouns involved just happen to be of different genders? It seems likely that the reference tracking role of gender can only be seriously important if there are many more than the three genders which German has.[4] Foley and Van Valin themselves convincingly demonstrate (Foley and Van Valin 1984: 326) the very important reference tracking role gender plays in the New Guinea language Yimas which, however, has about 16 different noun classes(!), and say that reference tracking of this type only works if "there is only one noun from each class in a discourse" (Foley and Van Valin 1984: 324).

Further problems with this interpretation can also be noted:

1 –Gender marking can in some cases lead to tracking failure and ambiguity: in German, *Katze* 'cat' is feminine and *Hund* 'dog' is masculine, so that in a household with a male cat and a female dog, conflict between natural and grammatical gender can lead to considerable pronominal confusion.
2 –There is also the the important point made by Croft (1994: 162) that "people talk more about people than about anything else" and that, therefore, reference tracking is most important for human referents – which is precisely where we find natural as opposed to grammatical gender.
3 –We also have to consider the perplexing fact, pointed out by Fodor (1959), that languages with gender often do not employ it in an efficient, functional way. He points out, for example, that German distinguishes between male and female horses lexically (*Hengst* vs. *Stute*) rather than by means of grammatical gender; and that the French for a female elephant is not *une éléphant* or *une éléphante* but *un éléphant femelle*, just as it is *nóstényelefánt* (literally 'female elephant' in Hungarian, a language without gender).

A further possible, psycholinguistic role, to do with processing, has been suggested to me by Paul Fletcher (p.c.). Given that most adults know several thousand nouns, and given that the time available for the recognition of a word can be measured in milliseconds, listeners need all the help that they can get in finding the right item in the lexical store. Anything which might cut down the range of possibilites for identification of an upcoming noun might be functional in this sense (see also Grosjean et al. 1994). Once again, however, this seems likely to be of most benefit in languages which have large numbers of genders, or very extensive classifier systems.

As we have seen, case, number, tense, aspect, person, mood and voice are all, like gender, grammatical categories that can be morphologically manifest in languages. Unlike grammatical gender, however, which appears to be of relatively little

[4] I owe this point to Małgorzata Fabiszak.

benefit for purely communicative purposes, especially if a language has only two or three genders, it is much easier to see intuitively what these other categories are "for" in the sense, in Givón's terms, of the knowledge they communicate. We have also established that grammatical gender is perhaps somewhat different from other grammatical categories – and, perhaps we can say, relatively afunctional – by its non-reappearance during creolisation. Indeed, we can even suggest that its relative lack of importance makes its appearance in languages where it does occur somewhat mysterious. Pidgins can manage without most of these grammatical categories. Creoles need all of them – except grammatical gender. There is therefore a case for suggesting that grammatical gender is a rather perplexing category. Our first question then is: why, unless like Yimas they are going to have enough noun classes to do an important job of reference tracking, do languages have grammatical gender?

8. Natural gender again

There is also another important question we have to face up to. We have distinguished between grammatical and natural gender, and suggested that the function of the former is more puzzling than the function of the latter. Even natural gender, however, is not without its enigmas. We have agreed that natural gender marking in the third person may tell us something that we did not already know, as well as, sometimes, help with reference tracking. But what of natural gender marking in the second and first persons? This, like grammatical gender generally, most often tells us very little indeed that we do not already know. It is very unlikely to help us with reference tracking: all of Foley and Van Valin's crucial examples involve the importance of gender as a reference tracking device in the third person. There can by definition be no reference tracking problems in the case of the first-person singular, and such problems are also very unlikely to occur very often in the first-person plural or in the second person.[5]

It is possible that some disambiguation may occur from time to time with second person pronouns. For example, a question such as *How are you* [sg.]*?* addressed to one person in the presence of another might be ambiguous as to the addressee unless one is male and one female and the language in question distinguishes between male and female second person pronouns (or, in pro-drop languages, verb forms). The same may also be true with first person plural pronouns – making it precisely clear, sometimes, who "we" are if there are different groups of people involved that are distinguished by sex.

But what can possibly be the function of gender distinctions in the first person singular? This form of gender-marking is particularly puzzling. It is true that there is the secondary function that written narratives in languages which have such marking reveal the sex of the narrator in a way which is not possible in other languages. But except in the written language – and the masculine/feminine first-person distinctions we see in some of the world's languages cannot be assumed to have arisen as a result

[5] Didier Maillat has pointed out to me that there may be a useful disambiguating function here when it is not clear whether a speaker is using indirect or reported speech, as in "He said I love you".

of the advent of writing (or crackly telephone lines!) – this form of gender marking gives us no "information" as such at all. It communicates no knowledge to us that we do not already have. It is quite normal to be able to tell whether a speaker is male or female – we do not, most usually, need distinct pronouns or other forms of grammatical marking to tell us this. Our second question is therefore: why do languages have natural gender other than for third person pronouns?

9. The gender of self-reference

Now, if gender is a relatively marginal grammatical category, this makes the gender of self-reference an even more remarkable phenomenon. All human languages would appear to allow speakers to refer to their own status as being male or female. Given that, as we have already remarked, the difference between male and female is the most fundamental difference there is between human beings, and undoubtedly therefore a semantic universal, it is hardly surprising that this difference is universally lexicalised and that all languages distinguish between lexical categories of the type *man-woman, boy-girl*. This then permits – indeed often requires – individual speakers to signal not only the sex of others but also their own sex lexically: *I am a happy woman* and *I am a sick man* and their equivalents are thus likely to be unremarkable sentences in all the languages of the world.

Languages, however, do differ considerably in the extent to which sex differences are lexicalised. This can be true of kinship terms, where for example *cousin* is not marked for gender in English but is in many other languages. It can also be true of occupational descriptions where, for example, languages may or may not distinguish between *actor-actress*, *manager-manageress*, etc. This issue has of course been the subject of much controversy recently, witness discussions in the English-speaking world as to whether a woman may be a *chairman* or not, and in the French-speaking world as to whether a female firefighter should be called a *pompière* or not.

This is an interesting and important topic. More interesting for our purposes, however, is the extent to which the sex of speakers is signalled **grammatically.** In the languages of the world there seem to be a number of possibilities for how the expression of self-referential natural gender may occur grammatically:

1 –It may not occur at all – as in English and Hungarian. This appears to be linked to the fact that such languages do not have grammatical gender.

2 –It may occur through the use of adjectival gender marking, as in French *je suis heureuse* versus *je suis heureux* 'I am happy'. In European languages this appears to occur only in languages which also have grammatical gender, although of course it is not inevitable in such languages – witness many languages including, for example, German, where it does not occur except, as Theo Vennemann (p.c) has pointed out to me, in appositional deadjectival nominal predicates such as *Ich unglückliche(r)!* 'poor me!' f(m). In many Slavic and Romance languages verbal past participles also behave like adjectives as far as gender marking is concerned. (Interestingly, this gives rise in Portuguese to gender marking in the word for 'thank you': *obrigado* versus *obrigada.*)

3 –It may occur through the use of distinct gender-marked verb forms in the first-person singular, as in Polish past tense and conditional verb forms. This also appears to be true, at least in Europe, only of languages which also have grammatical gender.

4 –It may occur through the use of distinct gender-marked first-person singular pronouns. Laycock (1965: 133) reports, for example, that the New Guinea language Ngala has the forms /wn/ 'I(m.)' and /ɲən/ 'I(f.)'.

The second question we posed above concerned the function of the gender of self-reference: what is it *for*? This question is not an easy one to answer. What is clear is that we do not receive any additional information as a result of first-person singular gender marking. In Givón's terms, no knowledge is communicated to us:

a) if a Ngala man says /wn/ while a Ngala woman says /ɲən/; nor
b) if an Italian man says *sono stanco* while an Italian woman says *sono stanca* 'I am tired'; nor
c) if a Polish man says *przyjechałem* while a Polish woman says *przyjechałam* 'I arrived'.

So how can we explain the existence of such self-referential gender-marking in the world's languages? Does it have any function? If so, what is it?

10. Possible functions

One interesting issue we should perhaps consider in this connection is whether there are any psycholinguistic consequences or even "benefits" to such gender marking. Let us approach this issue in the following way. Corbett points out that in many languages grammatical gender has to be maintained even when there is no disambiguating function possible at all e.g. when an object is being pointed to or is in some other way the obvious topic. Corbett discusses this interesting phenomenon and points out that it is of considerable semantic and psycholinguistic interest at what level of classification pronominalisation is decided, e.g. do French speakers pointing to, for example, a car and exclaiming the equivalent of *It's dirty!* use a pronoun translating *it* appropriate for pronominalising *thing* or *vehicle* or *car* or *Peugeot*?

It was a source of great psycholinguistic surprise to me, when I was learning Norwegian as a young adult, to be told when somebody passed something to me and I exclaimed *Det er tungt* 'It's heavy' that I had made a mistake because the thing I had been passed, a hammer, was grammatically masculine, while I had used a neuter pronoun and adjectival form – I should have said *Den er tung*. I protested that I was not thinking of it as a hammer but just as, well, something – but to no avail! Can this mean that speakers of languages which have grammatical gender are doing some-

thing which speakers of other languages do not? Do Norwegians constantly think, when they are picking something up, what it is, in a way English speakers do not? If so, can this also apply to natural gender? Does it mean that French speakers are more aware of the sex of the person they are talking to than English speakers? Does this mean that speakers of languages which extend natural gender into the first and second persons are also doing something which speakers of other languages do not do? If so, what is it? Can it possibly be that Russian or Ngala speakers are more aware, in some sense, of their own sex than speakers of English? In what sense could it be? Does it make any difference that French speakers have to learn as infants to say *heureux* rather than *heureuse* – or vice versa? In the absence of any psycholinguistic evidence, this all seems rather improbable.

Another possibility in our search for a function is to remind ourselves of work, particularly on American Indian languages, which has dealt with the issue of separate men and women's "languages". The well-known data from Haas's study (1944) of the American language Koasati, for instance, shows that men and women employed different verb morphology (see also Foley 1997: 300). In Trudgill (1974) I tentatively suggested that it was perhaps not a coincidence that such sex-exclusive differences were maintained (see Crystal 1971 for other instances) in societies with rather rigid and institutionalised gender-role differentiation. We can also remind ourselves of the issue of "appropriateness" (see Trudgill 1982) as an explanation for linguistic sex differences as revealed in sociolinguistic studies of western societies. Western societies typically have sex-preferential tendencies in language use, such as the well-known tendency for men, on average, to use prestige forms less often than women. It is not clear, however, where reminding ourselves of such phenomena will get us. Can there be a connection between societal structure, or social roles, and the grammatical signalling of ones own sex? Again this seems unlikely. As Corbett is forced to concede:

> When we consider work with sociolinguists and sociologists, where the concern is the link between language and society, we find the problems are more challenging than expected...it is not at all straightforward to establish links between grammatical gender and the relative status of those classified by the different genders....In Polish we find a distinction male human versus all other in the plural, which appears to be a particularly sexist division.[6] Russian, which is related to Polish, has no such feature: however, this does not reflect any obvious difference in the relative status of Polish and Russian women and men.[7] (Corbett 1991: 323)

One further point to consider, however, is the possibility suggested by Wierzbicka (1992: 394) that certain linguistic features may be of a reflection of aspects of culture from "the past, possibly the remote past", rather than from the present. (On the supposed sexism of Indo-European society, and its role in the development of the feminine gender in proto-Indo-European, see Miller 1977, as well as other references cited there.)

In any case, it is clear that in Ngala society it is considered appropriate for men and women to use different first-person singular pronouns, perhaps in rather the same way in which it has traditionally been considered appropriate in many western societies for women to swear less than men. But it is not clear *why* this is so, nor where this notion of appropriateness came from. And it is not clear, in particular, why failure to observe this distinction would be not only inappropriate but also ungrammatical.

11. Non-functionality

One further explanation that we should consider seriously is that self-referential gender-marking is not really functional at all, but that it occurs as a natural consequence of gender marking where it *is* useful, namely in the third-person. Note that it is not very surprising that gender is marked on adjectives or verbs in pro-drop languages such as Italian or Polish where third-person pronouns meaning 'he' and 'she' are likely to be absent. The distinction between *e stanco* 'he is tired' and *e stanca* 'she is tired' is a helpful distinction. Similarly the Polish distinction between *przyjechał* 'he arrived' and *przyjechała* 'she arrived' is also very functional.

It could therefore be argued that, grammatical systems being the regular systems that they are, it is only natural to extend adjectival and verbal agreement to other persons. It would be unsystematic to have agreement for third person adjectives but not for the other persons. (This patterning also explains how children eventually acquire correct gender marking, even where, for instance, boys are mostly exposed to the speech of women or girls.[8] There is anecdotal evidence that initially boys in this situation may use, for example, feminine verb forms in Polish, but that this state of affairs does not last long.)

This explanation will clearly not work, however, in the case of non-pro-drop languages such as French, where the adjectival distinction between *il est heureux* versus

---

[6] This point is made less tentatively by Sullivan (1981). However, it does not seem to me immediately obvious that, for example, the Polish distinction between *oni* "they: masculine personal" and *one* "they: masculine impersonal, feminine, neuter" is necessarily structurally sexist. It is just as likely that we are dealing with a case of polysemy. Polish recognises, we could say, five categories of nominals in the plural: 1.masculine personal; 2. masculine impersonal; 3. feminine personal; 4. feminine impersonal; 5. neuter. It just so happens that while, in the singular, to simplify somewhat, forms for 1. and 2., as well as for 3. and 4., are homophonous, in the plural it is 2., 3., 4. and 5. that are all homophonous. That is to say, the collapse of the formal distinction in the plural is not necessarily accompanied by any semantic collapse between these categories that is recognised by or has any cognitive effect on native speakers.

[7] Wierzbicka (1992: 323) does however make the interesting observation, concerning unsuccesful attempts by the Polish communist government to encourage the use of the second-person plural T pronoun *wy*, that 'Polish courtesy stresses respect for every individual and is highly sex conscious. The collectivist and genderless ring of the form *wy* was jarring in that tradition'. See also the comments above about animacy and agents; Jaworski (1986, 1989); and Herbert and Nykiel-Herbert (1986), who argue rather convincingly that Polish is in some respects structurally sexist.

[8] I owe this point to Grev Corbett.

elle est heureuse tells us nothing additional. And it does not work either for the existence of gender-marked first-person singular pronouns.

Let us consider further, however, the notion of non-functionality. One conclusion we may be able to draw from the evidence cited above is the following. Natural gender marking in the third person does indeed have a number of functions. So does grammatical gender marking in those languages which have large numbers of gender classes. However, grammatical gender marking in languages such as European languages which have only two or three genders seems to be almost totally non-functional. And, as Hickey (forthcoming) says, grammatical gender is a category which is "not guided by semantic needs"; if it were, he asks, why would we find languages such as modern Swedish and Danish which do not distinguish between masculine and feminine grammatical gender at all but simply between neuter and "common" gender (historical masculine and feminine combined)? Similarly, natural gender marking in the second and first persons – particularly the first person singular – has little or no function at all. We are used to the idea that human languages contain and indeed need redundancy to aid with processing. But do not these particular forms of gender marking represent redundancy on a somewhat nonfunctional scale?

The only way we can explain these phenomena satisfactorily would appear to be historically. We know that languages drag along with them a certain amount of, as it were, unnecessary historical baggage. This is most obvious in the case of grammatical irregularities which all languages appear to be able to tolerate up to a point. If the plural of *foot* in English is *feet* rather than **foots*, native learners can cope with this, and linguists can explain why it is so on historical grounds. But it may well be that in languages, or at least in some languages, there is much more of this afunctional historical baggage than has sometimes been thought. For example, the presence of different declensions for nominal forms and different conjugations for verbal forms in inflecting languages would appear to provide good evidence that languages can demonstrate large amounts of complex and non-functional differentiation which provide afunctionally large amounts of redundancy and whose presence in such languages can again, presumably, only be explained satisfactorily in historical terms.

Corbett tells us that "we are still some way from understanding how gender systems arise" (Corbett 1991: 310). Nevertheless, he argues that a likely origin for noun classes in general is to be found in nouns themselves and in particular in "nouns with classificatory possibilities such as 'woman', 'man', 'animal'". We then have to suppose diachronic processes involving the grammaticalisation of such nouns as classifiers (see also Lee 1988), which are well known to occur in languages such as Chinese. Classifiers can then in turn either come to be used anaphorically and turn into demonstratives – and subsequently pronouns and other gender markers – or they can be repeated within the noun phrase and give rise to gender agreement in that way (see also Harris and Campbell 1995: 341-2).

Other, probably secondary, forms of gender marking can also be explained historically. For instance, Slavic gender-marked verb forms derived originally from compound tenses which consisted of the verb *be* plus a past participle which, like adjectives (see above), agreed in gender. Then developments such as that which

occured in Russian took over: "the present tense of the verb 'be' in Modern Russian [became] the null form, which has left the original participle as the only verb element present" (Corbett 1991: 126).

When we say "explained", of course, it should be clear what manner of explanation we are talking about. These gender phenomena are "linguistic male nipples", in the sense of Lass (1997: 13). They came into being for a reason, but with no purpose. The reason was a series of grammaticalisation processes, as suggested by Corbett, which would appear to be "invisible hand" phenomena, in the sense of Keller (1994), in that they occur for reasons which have nothing to do with the ultimate outcome. (Lyle Campbell (p.c.) points out that the gendered Polish forms *pan/pani* and the gendered Spanish forms *vosotros/vosotras, nosotros/nosotras* are also the result of – rather different – grammaticalisation processes.) They are phenomena which, as biologists would say, have an explanation but no function. Whether or not it is clear why such grammaticalisation processes take place, it is clear that their motivation is not originally to divide nouns into agreement classes, or to aid with reference-tracking or disambiguation. The possibility of reference-tracking and disambiguation using gender differentiation in the third person of the type described by Foley and Van Valin is, as it were, a bonus (i.e. an example of exaptation, in the sense of Lass 1990). And it is a bonus which is scarcely operative in languages with few genders, or in gender marking in the second person and the first person plural, and which is not operative at all in gender marking in the first person singular.

Gender marking occurs with a very high degree of frequency indeed in those languages which have it, and is thus a feature with a very high degree of *entrenchment* in the sense of Langacker (1987: 59).[9] It is thus very readily maintained in the speech of individuals;[10] and because of the amazing language learning abilities of the human infant, languages readily maintain this type of complex historical baggage from one generation to another even though it represents a complication and/or an excess of redundancy, and even though it may have no particular or very important function. Indeed, Weist et al. (1991) have demonstrated, in connection with child language acquisition of tense and aspect marking in Polish, that a well-differentiated and regular set of paradigms may facilitate early learning, although it is not at all clear that we can generalise from this to gender.

Gender marking of the afunctional type disappears only when adults start playing an influential role in language learning in contact situations such as those which give rise to the development of pidgins, creoles and creoloids (see above). This disappearance also occurs in the development of koinés, by which is meant varieties which arise in dialect contact situations and which result from dialect mixture, levelling and simplification (see Trudgill 1986): for instance, the standard forms of Swedish and Danish have only two genders, while many non-standard dialects of

[9] I owe this point to Bill Croft

[10] This does not mean to say that languages cannot demonstrate considerable dialectal or idiolectal variation in gender assignment – see Fisiak (1975); Kryk-Kastovsky (1998).

these languages still have three. Given that language contact is becoming an increasingly important fact in the modern world, this opens up an intriguing possibility. In Trudgill (1992), I argued that it was interesting to consider the title of Labov's influential (1975) paper "On the use of the present to explain the past". While totally supporting the Labovian enterprise, I suggested that, increasingly, the present is going to be unlike the past in demographic and social network terms, and that this might well lead to differences in the direction of linguistic change and in the distribution of structures over the world's languages. I also suggested that increasing language and dialect contact means that creoles, creoloids and koinés are on the increase, and that languages spoken in small, isolated, communities with tightly-knit social networks – which I hypothesised were the types of language most likely to produce complexity and redundancy and to transmit them to descending generations (for fuller argumentation, see also Trudgill 1998) – were becoming less and less common. It is therefore not unlikely that languages with large numbers of afunctional grammatical devices will become less numerous, and indeed it is not entirely impossible that linguistic gender, except perhaps for natural gender in the third person, will one day disappear from the languages of the world, never to return. If this is so, we should, like Foley and Van Valin, do as much as we can, as quickly as we can, to investigate languages with complex gender systems before it is too late.

REFERENCES

Aikhenvald, A. 1998. "Physical properties in a gender system: a study of Manambu". *Language and Linguistics in Melanesia* 27 (1996). 175-187.

Ardener, E. (ed.). 1971. *Social anthropology and language*. London: Tavistock.

Bybee, J. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

Bloomfield, L. 1933. *Language*. New York: Holt.

Broch, I. and E. H. Jahr (eds.). 1996. *Language contact in the Arctic: Northern pidgins and contact languages*. Berlin: Mouton de Gruyter.

Comrie, B. 1988. "Coreference and conjunction reduction in grammar and discourse". In: Hawkins, J. (ed.). 186-210.

Comrie, B. and G. Corbett. (eds.). 1993. *The Slavonic languages*. London: Routledge.

Copeland, J. E. and P. W. Davis (eds.) 1981. *The seventh Lacus forum 1980*. Columbia: Hornbeam.

Corbett, G. 1991. *Gender*. Cambridge: Cambridge University Press.

Croft, W. 1994. "Semantic universals in classifier systems". *Word* 45. 145-171.

Crystal, D. 1971. "Prosodic and paralinguistic correlates of social categories". In: Ardener, E. (ed.). 185-208.

Dressler, W., Dziubalska-Kołaczyk, K. and M. Fabiszak. 1998. "Polish inflectional classes within natural morphology". *Bulletin de la Societé Polonaise de Linguistique* 53. 95-119.

Fisiak, J. 1975. "Some remarks concerning the noun gender assigment of loanwords". *Bulletin de la Societé Polonaise de Linguistique* 33. 59-63.

Fodor, I. 1959. "The origins of grammatical gender". *Lingua* 8. 1-41; 186-214.

Foley, W. A. 1997. *Anthropological linguistics: An introduction*. Oxford: Blackwell.

Foley, W. A. and R. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Givón, T. 1979. *On understanding grammar*. New York: Academic Press.

Givón, T. 1984. *Syntax: A functional typological introduction*. Amsterdam: Benjamins.

Greenberg, J. 1966. "Some universals of grammar with particular reference to the order of meaningful elements". In Greenberg J. (ed.). 73-113.

Greenberg, J. (ed.). 1996. *Universals of language*. Cambridge, Mass.: MIT Press.

Grosjean, F., J.-Y. Dommergues, E. Cornu, D. Guillelmon, and C. Besson. 1994. "The gender-marking effect in spoken word recognition". *Perception and Psychophysics* 56. 590-598.

Haas, M. R. 1944. "Men's and women's speech in Koasati". *Language* 20. 142-149.

Harris, A. and L. Campbell. 1995. *Historical syntax in a cross-linguistic perspective*. Cambridge: Cambridge University Press.

Hawkins, J. (ed.). 1988. *Explaining language universals*. Oxford: Blackwell.

Heath, J. 1975. "Some functional relationships in grammar". *Language* 51. 89-104.

Heilmann, L. (ed.). 1975. *Proceedings of the 11th international congress of linguists*. Bologna: Mulino.

Hendrick, R. A., C. A. Masek and M. F. Miller (eds.). 1981. *Papers from the seventeenth regional meeting, Chicago linguistic society*. Chicago: Chicago Linguistic Society.

Herbert, R.K. and B. Nykiel-Herbert. 1986. "Explorations in linguistic sexism: a contrastive sketch". *Papers and Studies in Contrastive Linguistics* 21. 47-86.

Hickey, R. Forthcoming. "The phonology of gender in Modern German".

Holm, J. 1997. "Passive-like constructions in English-based and other creoles". In Schneider, E. (ed.). 71-86.

Jahr, E. H. (ed.) 1992. *Language contact: Theoretical and empirical studies*. Berlin: Mouton de Gruyter.

Jaworski, A. 1986. *A linguistic picture of women's position in society: A Polish-English contrastive study*. Frankfurt: Peter Lang.

Jaworski, A. 1989. "On gender and sex in Polish". *International journal of the sociology of language* 78. 83-92.

Jespersen, O. 1924. *The philosophy of grammar*. New York: Norton.

Keller, R. 1994. *On language change: The invisible hand in language*. London: Routledge.

Kellerman, G. and M. D. Morrissey (eds.). 1992. *Diachrony within synchrony: Language history and cognition*. Frankfurt: Peter Lang.

Kryk-Kastovsky, B. 1999. "Norm vs. use: On gender in Polish". In Unterbeck, B. and M. Rissanen. (ed.). 729-748.

Labov, W. 1975. "On the use of the present to explain the past". In Heilmann, L. (ed.).

Langacker, R. W. 1987. *Foundations of cognitive grammar: theoretical prerequisites*. Stanford: Stanford University Press.

Lass, R. 1990. "How to do things with junk: Exaptation in linguistic change". *Journal of Linguistics* 26. 79-102.

Lass, R. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

Laycock, D. C. 1965. *The Ndu language family*. Canberra: Pacific Linguistics.

Lee, M. 1988. "Language, perception and the world". In Hawkins, J. (ed.). 211-246.

Lewandowska-Tomaszczyk, B. 1992. "Cognitive and interactional conditioning of semantic change". In Kellerman, G. and M.D. Morrissey (eds.). 229-250.

Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.

Miller, D. G. 1977. "Tripartization, sexism, and the rise of the feminine gender in Indo-European". *The Florida Journal of Anthropology* 2. 3-16.

Rothstein, R. 1993. "Polish". In Comrie, B. and G. Corbett. (eds.). 686-758.

Schneider, E. (ed.). 1997. *Englishes around the world*. vol. 1. Amsterdam: Benjamins.

Stone, G. 1993. "Cassubian". In Comrie, B. and G. Corbett (eds.). 759-794.

Stucky, S. U. 1978. "How a noun class system can be lost: Evidence from Kituba lingua franca Kikongo". *Studies in the Linguistic Sciences* 8. 217-233.

Sullivan, W. J. 1981. "English as a (non-)sexist language". In Copeland, J. E. and Ph. W. Davis (eds.). 444-456.

Trudgill, P. 1974. *Sociolinguistics. An introduction*. London: Penguin.

Trudgill, P. 1978 "Creolisation in reverse: Reduction and simplification in the Albanian dialects of Greece". *Transactions of the Philological Society 1976-77*. 32-50.

Trudgill, P. 1982. *On dialect: Social and geographical perspectives.* Oxford: Blackwell.
Trudgill, P. 1986. *Dialects in contact.* Oxford: Blackwell.
Trudgill, P. 1992. "Dialect typology and social structure". In Jahr, E. H. (ed.). *Language contact: Theoretical and empirical studies.* Berlin: Mouton de Gruyter. 195-212.
Trudgill, P. 1996. "Dual source pidgins and reverse creoles: Northern perspectives on language contact". In Broch, I. and E. H. Jahr (eds.). 5-14
Trudgill, P. 1998. "Typology and sociolinguistics: Linguistic structure, social structure and explanatory comparative dialectology". *Folia Linguistica* 31/3-4: 349-60.
Unterbeck, B. and M. Rissanen. (ed.). 1999. *Gender in grammar and cognition.* Berlin: Mouton de Gruyter.
Weist, R., H. Wysocka and P. Lyytinen. 1991. "A cross-linguistic perspective on the development of temporal systems". *Journal of Child Language* 18. 67-92.
Wierzbicka, A. 1992. *Semantics, culture and cognition: Universal human concepts in culture-specific configurations.* Oxford: Oxford University Press.
Zubin, D. A. and K. M. Köpcke. 1981. "Gender: A less than arbitrary grammatical category". In Hendrick, R. A., C.A. Masek and M. F. Miller (eds.). 439-49.