1 Introduction

Gender is the most puzzling of the grammatical categories. It is a topic which interests non-linguists as well as linguists and it becomes more fascinating the more it is investigated. In some languages gender is central and pervasive, while in others it is totally absent. One of its attractions for linguists is that there are interesting aspects of the study of gender in each of the core areas of linguistics. And work on it promises practical benefits, even in the short term, in meeting the problems which gender causes in second-language learning. In the longer term, research into gender will be important for at least two other areas: first, it can shed light on the way in which linguistic information is stored in the brain; and second, it has implications for natural language processing, notably for the elimination of local ambiguities in parsing. To understand what linguists mean by ‘gender’, a good starting point is Hockett’s definition: ‘Genders are classes of nouns reflected in the behavior of associated words’ (1958: 231). A language may have two or more such classes or genders. The classification frequently corresponds to a real-world distinction of sex, at least in part, but often too it does not (‘gender’ derives etymologically from Latin genus, via Old French genire, and originally meant ‘kind’ or ‘sort’). The word ‘gender’ is used not just for a group of nouns but also for the whole category; thus we may say that a particular language has, say, three genders, masculine, feminine and neuter, and that the language has the category of gender.

1.1 Gender in the languages of the world

Discussions of gender as a category have tended to centre on relatively small numbers of languages, and often on selections which are not typical of the systems found in the world’s languages. In contrast, we shall look at over 200 languages. Some will appear only briefly, because of some special point of interest, others will run like threads through the book showing how the different aspects of gender systems relate to each other. Grammatical gender is certainly widespread, and so a brief account of its distribution may
Introduction

prove helpful. Europe is dominated by the Indo-European language family, which also extends well into Asia. Many Indo-European languages show gender (some with three genders, others having reduced the number to two); a few have lost gender, while others, notably the Slavonic group, are introducing new subgenders. Uralic has some members in Europe (like Hungarian), and others in the northern area where Europe and Asia meet, and is devoid of grammatical gender. Joining Europe and Asia in the south we find the Caucasus, where the languages of the northern Caucasus, some thirty-five in number, show particularly interesting gender systems, which contrast markedly with those of Indo-European. Several of the major families of Asia provide no material for our investigation, but in south India we find the Dravidian family, which includes languages like Tamil and Telugu, which are of great importance for the typology of gender. Bridging Asia and Africa, the Afro-Asiatic family offers numerous two-gender systems, some of which are of special importance. The other three families of Africa, namely Nilo-Saharan, Niger-Kordofanian and Khoisan, all have languages with gender systems. Niger-Kordofanian provides some of the most extensive examples, in terms of the numbers of genders and the degree to which gender is reflected in syntax. Heine (1982: 190) estimates that 600 African languages (some two-thirds of all African languages) are gender languages. New Guinea has around 1,000 languages, a substantial proportion of the world's languages, and gender is widespread here too. In Australia gender is found in various languages, mainly in those of Arnhem Land and the North Kimberleys. Finally, in the Americas, the examples of gender languages are few and are generally isolated. The most important exception is the Algonquian family, whose two-gender systems will figure prominently in our study. In comparing data from languages of such variety we must be careful to ensure that we are comparing like with like. To do this we shall be explicit about the techniques used, the sources of data and, of course, the definitions of the terms we use.

1.2 General approach and outline of the book

The book is designed for various types of readers. First, for the student of linguistics, it is an introduction to an area of obvious interest, one which is poorly represented in the standard texts. And through this topic the book attempts to give an insight into the richness and variety of the world's languages. Second, it is intended to help those doing research on specific languages or groups of languages, whether for an undergraduate dissertation or a major research project. Seeing a familiar language analysed in the broader context of languages with comparable but different systems can give a new perspective on familiar material. For some languages, the accounts of gender
1.2 General approach and outline

set in their particular grammatical tradition obscure similarities to other
genetically distant or unrelated languages. An overview of this type seeks to
highlight such similarities and to suggest new ways of approaching old
problems. References to work on specific languages can be found by checking
relevant sections identified using the language index. Of those researching
individual languages, field-workers are a special category. It is hoped that the
definitions provided will help to ensure that the invaluable work done in the
field – particularly on languages with uncertain futures – will not be under-
mined through the contradictory use of terms (which has hampered this topic
in the past) or the failure to obtain data which are of special value for
understanding gender more generally. There will also be readers, from various
disciplines, concerned with sexism in language. This is a topic on which several
interesting studies have appeared recently, but not one which is central to this
book. However, it is hoped that material presented here will contribute to that
debate in two ways. First, the systematic presentation of linguistic data from
many different languages may help to broaden a discussion which has tended
to centre on English. It will also show how divisions into animate and
inanimate, or human and non-human, function in language exactly as does
the division into female and male. Second, by drawing attention to languages
where the feminine rather than the masculine is in some sense favoured, it may
suggest possible comparative approaches. The book is therefore planned to be
a source book as well as a textbook, with extensive references for those who
wish to go further, whether into particular topics or into particular languages.

Different readers will have different requirements, so it will be useful to
outline the structure of the book, to make clear which parts will be most
relevant to particular needs. Chapters 2–5 are all concerned with gender
assignment, that is, the way in which native speakers allocate nouns to
genders. The type of question at issue is how speakers know that, for example,
the word for ‘house’ is masculine in Russian, feminine in French and neuter
in Tamil. In chapter 2 we analyse languages where the meaning of a noun is
sufficient to determine its gender; thus ‘house’ in Tamil is neuter because it
does not denote a human. Then in chapter 3 we move to languages where
meaning is not adequate to determine gender on its own, but has to be
supplemented by formal criteria. These additional criteria may be mor-
phological, that is, relating to word-structure: ‘house’ can be assigned to the
masculine gender in Russian, given the declensional type to which it belongs.
Or the criteria may be phonological, relating to sound-structure: hence
‘house’ is feminine in French because of the phonological shape of the word.
In these two chapters we look at the straightforward linguistic evidence, and
we find that the regularities which justify the analyses offered are striking. In
Introduction

chapter 4 we go on to examine other types of evidence which support the rules proposed, suggesting that they do indeed form part of the native speaker’s competence. The evidence comes from the way in which nouns borrowed from one language into another gain a gender, children’s acquisition of gender, psycholinguistic experiments, the curious effects of the residual meaning of gender, and from the investigation of the way in which gender systems change over time. This fourth chapter will be of central interest to some readers but can safely be skipped by those wishing to gain an initial outline of the subject.

In chapters 2–4 we assume that we can determine analytically the number of genders in a given language and the gender of a particular noun; our task is to determine how the native speaker assigns nouns to genders (and so can produce the examples which form our data).

In many languages there is no dispute as to the number of genders, but there are other languages where the question is far from straightforward; consequently it is important to investigate how we solve such cases. While nouns may be classified in various ways, only one type of classification counts as a gender system; it is one which is reflected beyond the nouns themselves in modifications required of ‘associated words’. For example, in Russian we find: novyj dom ‘new house’, novaja gazeta ‘new newspaper’ and noevo taksi ‘new taxi’. These examples demonstrate the existence of three genders, because the adjective nov- ‘new’ has to change in form according to the gender of the noun. There are many other nouns like dom ‘house’, making up the masculine gender, many too like gazeta ‘newspaper’ (the feminines) and numerous nouns like taksi ‘taxi’ (the neuters), each requiring the appropriate ending on the adjective. There are various other ways in which nouns could be grouped: those denoting animals, those which are derived from verbs, those whose stem has three syllables or more, those whose stress changes from singular to plural. These groupings are not genders in Russian because they do not determine other forms beyond the noun; they are classifications internal to the class of nouns.

All this means that the determining criterion of gender is agreement; this is the way in which the genders are ‘reflected in the behavior of associated words’ in Hockett’s definition given earlier. Saying that a language has three genders implies that there are three classes of nouns which can be distinguished syntactically by the agreements they take. This is the generally accepted approach to gender (other suggestions prove unsatisfactory, as we shall see).

Given its importance for the analysis, agreement in gender is considered in detail in chapter 5, and it turns out to be varied and complex. It is not only adjectives and verbs which can show agreement in gender, but in some languages adverbs agree, in others numerals and sometimes even conjunctions
1.2 General approach and outline

agree in gender. Since agreement is taken as the criterion for gender, there are no grounds for drawing a distinction between languages in which nouns are divided into groups according to sex, and those where human/non-human or animate/inanimate are the criteria. Thus many languages described as having ‘noun classes’ fall within our study. The number of genders is not limited to three: four is common and twenty is possible. On the other hand, classifiers fall outside our study because they do not show agreement; but they are discussed in chapter 5, because they are a source for gender systems. A further consequence of having agreement as the criterion is that the definition of agreement itself becomes important. Most scholars working on agreement include the control of anaphoric pronouns by their antecedent (the girl…she) as part of agreement. If this is accepted, then languages in which pronouns present the only evidence for gender should be recognized as having a gender system. This is the approach we shall adopt but, because it is not universally accepted, we shall call such systems ‘pronominal gender systems’. Since it raises these problems and illustrates the possible divergence in gender agreement, chapter 5 is particularly important for those whose knowledge of languages is predominantly in the field of ‘standard average European’.

Chapter 5 also serves as an introduction to chapter 6, where the major definitions and procedures are given and illustrated, allowing us to determine the number of genders in a given language and giving terms to describe the complex gender systems which occur surprisingly frequently in the languages of the world. Chapter 6 is central, forming the basis for the analysis of gender systems which underpins the entire study. Though we find gender systems which appear to differ radically, we see how they can be analysed within a common framework. The next two chapters are devoted to topics highlighted by chapter 6. In chapter 7 we examine problems concerned with the ‘agreement target’, the item which shows agreement in gender. There is the question of syncretism, where particular agreeing items have ‘too few’ forms in that they do not distinguish as many gender forms as might be expected. There is also the situation in which the target has ‘too many’ forms; the existence of gender agreement can impose a problematic choice if, for example, the speaker has to choose between masculine and feminine even though the sex of the referent is not known. Chapter 8 is devoted to ‘hybrid’ nouns; these are nouns which are not assigned to a single gender by the assignment rules. They therefore take different gender agreement according to what is agreeing with them (an example is German Mädchen ‘girl’). The different patterns of agreement with hybrid nouns seem confusing at first sight, but we discover that their distribution is predictable to a large degree and that they offer a route for change in gender systems. Chapter 9 examines


Introduction

the rules which determine agreement in gender when there is more than one noun phrase; for example, if the subject consists of two noun phrases, one headed by a feminine noun and the other by a neuter noun, there must be a rule to determine which gender the verb or predicative adjective will stand in. The data here are varied and unexpected. Finally, chapter 10 attempts to draw together various threads. In particular, the question of change in gender systems, which is dealt with at appropriate points throughout the book, is surveyed more generally. We concentrate on the analysis and description of changes for which firm data can be found, since much of the earlier work on the origin of gender in particular was largely speculative. In the final chapter we also look forward to how the study of gender may progress, looking at possibilities for further work, both elementary and advanced.

In broad outline, the first part of the book, chapters 2–4, is concerned with the genders into which nouns can be divided, and the middle, chapters 5–7, is more concerned with elements which agree in gender. The two sides of the analysis come together in chapters 8 and 9, while chapter 10 looks back over the topics covered and forward to possible advances.

1.3 Presentation of data

The orthography used in examples normally follows that of the original sources, so that the interested reader can refer back to them easily. However, if they are in a non-Roman script, a standard transliteration is used. The languages of the Caucasus pose special problems and, where possible, minor amendments have been made in transliteration in order to be consistent with Kibrik, Kodzasov, Olovjannikova & Samedov (1977: 41). Descriptions of Algonquian languages indicate length in a variety of ways: to avoid confusion we shall use a colon to mark length of the preceding vowel for all Algonquian examples. Occasionally italics are used to draw attention to part of an example for discussion in the text. Examples are followed by morpheme glosses, which are intended solely to help with understanding the point at issue; they are not full glosses. When words are segmented in the example, the same segmentation is used in the gloss, for example laugh-s laugh-3RD.SG, in which the s is glossed as '3RD.SG'. Since the s cannot be segmented into constituent morphs representing third person and singular number separately, the glosses for these morphemes, abbreviations in this case, are joined by a stop. Abbreviations are listed on page xx. A translation is also given, unless the meaning of the example is fully clear from the gloss.
2

Gender assignment I: semantic systems

An intriguing question, which interests non-linguists as well as linguists, is the way in which nouns are allotted to different genders. The linguist who wishes to establish the gender of a given noun can use agreement as a test (for details see chapter 6). However, the native speaker of the language must know the gender of a noun in order to produce the correct agreements (the evidence which the linguist uses). The amount of information is substantial, since native speakers know the gender of many thousands of nouns. For foreign learners of the same language, in contrast, this knowledge often proves elusive in the extreme. How then does a native speaker know the gender of a particular noun? One possible answer would be that the speaker simply has to remember the gender of each noun. This suggestion would involve a considerable feat of memory. It seems an unlikely answer, though many linguists have been ready to accept it. For example, in an often quoted remark, Bloomfield (1933: 280) claimed that:

There seems to be no practical criterion by which the gender of a noun in German, French, or Latin could be determined.

This pessimism now appears misplaced in view of the following evidence. First, native speakers typically make few or no mistakes in the use of gender; if the gender of every noun were remembered individually, we would expect more errors. Second, words borrowed from other languages acquire a gender, which shows that there is a mechanism for assigning and not just remembering gender. And third, when presented with invented words, speakers give them a gender and they do so with a high degree of consistency. Thus native speakers have the ability to ‘work out’ the gender of a noun; models of this ability are called ‘assignment systems’. Convincing accounts of gender assignment in French have in fact been offered and, while German gender appears more complex than French gender, recent analyses have gone a long way towards establishing practical criteria for gender assignment in German too.

Assignment may depend on two basic types of information about the noun:

7
**Gender assignment I**

its meaning (semantics) and its form. Information about form may in turn be of two types: word-structure, comprising derivation and inflection (morphology), and sound-structure (phonology). In this chapter we concentrate on semantic factors, and in chapter 3 we turn to morphological and phonological factors. Languages may use different combinations of these factors and may also permit varying numbers of exceptions. Mel'čuk (1958 [1974: 33]) makes the point that rules are valuable even if there are exceptions; a rule which assigns a large proportion of the nouns correctly is of theoretical interest and practical use. From the theoretical point of view, assignment systems have important implications for attempts to determine the structure of the lexicon. And given the ease with which native speakers assign nouns to genders, and the difficulty experienced by foreign learners of many gender languages, an understanding of gender assignment systems is of considerable practical importance.

In a sense all gender systems are semantic in that there is always a semantic core to the assignment system (Aksenov 1984: 17–18). However, we shall consider here those languages where semantic factors are sufficient on their own to account for assignment. We shall first examine strict semantic systems, then move to those which are primarily semantic but which allow varying numbers of exceptions.

2.1 **Strict semantic systems**

These are systems in which the meaning of a noun determines its gender and in which, equally, given the gender of a noun we can infer something about its meaning. This is the sort of system we might have expected to find as the normal case. While there are several examples in the Dravidian family (section 2.1.1) and various others scattered around the world (section 2.1.2), overall this type of system is not particularly common.

2.1.1 **Tamil and other Dravidian languages**

Gender is found in most Dravidian languages and nouns are assigned to gender according to their meaning. We take as our first example Tamil, one of the major Dravidian languages. It has some 50 million speakers, mainly in Tamil Nadu in south-east India, but also in Sri Lanka and various other parts of the world. The data on colloquial Tamil are from Asher (1985: 136–7), supplemented by Arden (1942: 74) and Andronov (1966: 54–5), who describe written Tamil; their examples have been retransliterated according to Asher's system (1985). Nouns may be divided into rational and non-rational (neuter). The rationals may in turn be divided into two groups, masculine and feminine. Nouns are assigned to these three genders as shown in table 2.1.
2.1 Strict semantic systems

Table 2.1 Assignment in Tamil

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Gender</th>
<th>Examples</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>god or male human</td>
<td>masculine (= male rational)</td>
<td>aaŋ</td>
<td>man</td>
</tr>
<tr>
<td>goddess or female</td>
<td>feminine (= female rational)</td>
<td>civan</td>
<td>Shiva</td>
</tr>
<tr>
<td>human</td>
<td></td>
<td>pen</td>
<td>woman</td>
</tr>
<tr>
<td>other</td>
<td>neuter (= non-rational)</td>
<td>kaalı</td>
<td>Kali</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maram</td>
<td>tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>viitı</td>
<td>house</td>
</tr>
</tbody>
</table>

This assignment system operates with a high degree of consistency. Given the meaning of a noun, its gender can be predicted without reference to its form. Thus, for example, one can be confident that a noun denoting a female will be feminine, and that a noun which is feminine will denote a female. Such systems are sometimes called 'natural gender systems'. There are a very few apparent exceptions to the rules given above. The words cuuriyan ‘sun’ and cantiran ‘moon’ are both treated as masculine, as are other heavenly bodies; this is explained by the fact that they are also the names of gods. Words for child, such as makau, are usually neuter but may also be masculine or feminine. There is some scope for metaphorical use of gender; vaanai ‘elephant’ with masculine or feminine gender would refer to a man or woman with some elephant-like qualities. Animals may be treated as persons in fables; normally, however, even when there are distinct words for the male and female of animals, all nouns referring to animals are neuter. In some instances there is also a morphological clue as to gender, but this is an additional regularity since the meaning of a noun is sufficient in itself.

At this stage, it is worth pointing out that the names used for different genders are not significant. The traditional Tamil terms for the two main classes of nouns are ‘high-caste’ and ‘no-caste’ (rational and non-rational in more modern terms). This primary division is reflected in morphology (see figure 6.8) and will be of importance too when we analyse gender resolution (section 9.3). As far as assignment is concerned, however, nouns must be divided into three classes and nothing rests on the actual labels used. Equally, different languages may have similar systems but linguists working on them may use different labels. For instance, several North-East Caucasian languages have three genders and assign nouns to them using the same semantic factors as does Tamil. They include Akhvakh, Bagval, Godoberi and Karata (all Andi languages of the Avar-Andi-Dido group of North-East Caucasian). Linguists working on these languages normally use the labels I, II and III. While names for genders are helpful, there is much to be said for the
Gender assignment I

numbering system, since it prompts us to spell out exactly which types of nouns are included. For languages where the use of names like ‘masculine’ and ‘feminine’ is normal, it is important to remember that, say, the feminine gender in one language may contain a rather different set of nouns from the feminine gender in another: the feminine gender in Tamil does not include inanimates, but there are large numbers of them in the feminine gender of French. It should also be said that, while in Dravidian linguistics it is normal to talk of ‘genders’, those working on Caucasian languages usually talk of ‘noun classes’ rather than ‘genders’. The use of ‘gender’ or ‘noun class’ is also more a matter of tradition than of substance, as we shall see in section 6.1. The choice is not important; for consistency, we shall normally use the term ‘gender’.

Turning now to other members of the Dravidian family, we find that in Kannada the situation is similar to that in Tamil. There is a very small number of exceptions to the semantic principle: basava ‘bull’ and koonja ‘buffalo’ are masculine (Andronov 1969: 29). These are not arbitrary exceptions; rather, their gender reflects the special status of these higher animals. However, though the nouns given are masculine in the singular, they are neuter, as expected, in the plural; these nouns therefore form an ‘inanimate’ gender (see section 6.4.3). Like Kannada, other Dravidian languages such as Telugu have the same semantic assignment rules as Tamil (Arden 1873: 46) although, as we shall see in section 6.3, the morphological structure of gender in Telugu is rather different. Once again male humans are masculine, female humans feminine and others neuter. The gender of divine beings depends on their role in mythology (Malathi Rao, personal communication): thus ganga (the river Ganges) is feminine, hanumantuDu (Hanuman, a monkey) is masculine and kaamadheenu (divine cow) is neuter, and so on, because of the parts these divine beings play in myths. Gender according to role in mythology is something we shall find in languages from all over the world.

Some Dravidian languages have two genders rather than three. The languages involved include Kolami (Emeneau 1955: 73), Ollari (Bhattacharya 1957: 19) and Parji (Burrow & Bhattacharya 1953: 9), all members of the Kolami-Parji subgroup. These languages show what is probably a development of the situation found in Telugu; the feminine and neuter genders have coalesced (but see section 6.4.2) so that masculine is now opposed to non-masculine (more details in section 7.1.3). Thus the masculine gender includes nouns denoting male humans and the non-masculine gender includes all others. The situation of gods is not clear, but in Parji, according to Burrow & Bhattacharya (1953: 9), all supernatural beings, including gods and goddesses, are treated as neuter. These languages all have semantic assignment