4 Gender and noun classes

Greville G. Corbett

0 Introduction

Gender is a fascinating category, central in some languages, absent in others. The term gender is normal in some traditions, in Indo-European and Dravidian studies for instance, while others use the term noun class (sometimes preferred by Caucasians and Australianisms). A language may have two or more such classes or genders. 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. Furthermore, the classification may correspond 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 genere, and originally meant ‘kind’ or ‘sort’).

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 through agreement. For example, in Russian we find novyi dom ‘new house’, novaja gazeta ‘new newspaper’ and novoe taksi ‘new taxi’. These examples demonstrate the existence of three genders, because the adjective nov- ‘new’ changes in form according to the noun. There are numerous other nouns like dom ‘house’, together making up one gender. Since this gender includes many nouns denoting males, like tata ‘father’, it is known as the ‘masculine gender’. There are many too like gazeta ‘newspaper’ (the feminines, since nouns denoting females, like mamy ‘mother’, typically belong here) and there are 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. 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.
Since agreement is taken as the criterion for gender, there are no grounds for drawing a distinction between languages in which the groups of nouns identified in this way correlate with sex and those where they correlate with some other feature, such as human/non-human or animate/inanimate. This is why languages described as having noun classes can normally equally well be said to have genders. Given these different semantic criteria for gender systems, the number of genders is not limited to two, nor to three: four is common and twenty is possible. 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 most logical approach, but, since it is not universally accepted, such systems are best labelled pronominal gender systems.

We shall first consider the problems the linguist/field worker faces when confronted by a gender system, that is, the analytical problem of determining the number of genders and the tests for deciding the gender of a given noun (section 1). We then consider the ‘speaker's problem': the speaker, of course, has to know the gender of a noun in order to use it (and to produce the data which the analyst uses). The way in which the speaker assigns nouns to genders is discussed in section 2; it is a topic which is often of interest to speakers as well as to linguists. Then we consider the question of default gender (section 3), followed by gender resolution, which is a complex area of gender in some languages—those which have the right configuration of the agreement system (section 4). Finally (section 5) we survey the prospects for research into the category of gender.1

1 Terms and analysis

Much of the literature on gender is confusing. In many languages the gender pattern appears straightforward. In others, linguists present the pattern as though it were equally uncontroversial, but we find that similar situations are described differently by those working on different language families. Or the number of genders in a particular language can be the subject of considerable dispute.

1 This account draws on Corbett (1991), where much more detail and a substantial bibliography can be found. Where possible, material published since that work went to press is cited here. The research was supported in part by the ESRC (UK) under grants RES0033062 and RES00338224; this support is gratefully acknowledged. I also wish to thank Alexandra Akerlund and Tulli Stjepanovic for helpful comments on a draft; Nilsen Giles, Jr., for data on Kao; and Marianne Mithun, Frans Plank, Hanna Tammola, and Larry Trask for bringing useful references to my attention.
Gender and noun classes

Evidently we need a consistent approach to analysing gender. The best approach derives genders from agreement classes, which are set up solely on syntactic evidence. This syntactic approach provides the first step in a procedure for deciding the number of genders to which nouns can be assigned in a given language. We shall consider the agreement class approach (section 1.1), and then look briefly at other related systems of classification (section 1.2).

1.1 Analysis based on agreement classes

We shall follow the widely accepted view that the existence of gender can be demonstrated only by agreement evidence. We cannot demonstrate the existence of a gender system just by looking at the nouns themselves. The presence of markers on the nouns, as prefixes or suffixes, does not of itself indicate that a language has genders (or noun classes); if we accepted this type of evidence, then we could equally claim that English had a gender comprising all nouns ending in -tion. In the case of gender, the evidence comes from agreement markers on other sentence elements, whose form is determined by the gender of the head noun of the controller. The range of items which may show agreement in gender is considerable, including adjectives, articles, numerals, possessives, verbs, various pronouns, adverbs (in languages like Lako), adpositions (Abkhaz), and even complementizers (West Flemish): see Corbett (1991:106-15) for examples of all of these. The form of gender agreement varies considerably too.

This approach to gender based on agreement owes a good deal to Zaliznjak (1964). It requires the notion of agreement class, which we define as follows:

An agreement class is a set of nouns such that any two members of that set have the property that

whenever

(i) they stand in the same morphosyntactic form

and

(ii) they occur in the same agreement domain

and

(iii) they have the same lexical item as agreement target

then

their targets have the same morphological realization.

The intuitive content of the definition is that two nouns are in the same agreement class provided that, given the same conditions, they will take the same agreement form. The three numbered clauses of the definition spell out what is involved in 'the same conditions'. Being in 'the same morphosyntactic form' (clause (i)) means that the nouns have the same specifications for all relevant syntactic features. The features most commonly involved are number and case. We rely on the notions of number and case being given, since they are simpler
notions, which can often be justified simply on morphological evidence. Identity of morphosyntactic form does not imply morphological identity. Two nouns may be in the same morphosyntactic form and yet differ morphologically: for example Russian mět 'mother' and sestra 'sister' can both be in the same morphosyntactic form, the accusative singular. Yet their morphological realizations are different: they take different endings (they belong to different declensional classes). Conversely, two morphosyntactic forms may have a single morphological realization: for example Russian očko 'window' may be the nominative singular or the accusative singular (two morphosyntactic forms for which many other nouns have distinct morphological realizations). Provided that the nouns meet the requirement of clause (i), that is, they stand in the same morphosyntactic form, then they start out, as it were, on level terms.

Clause (ii) requires that the nouns occur in the same agreement domain. This means that the configuration in which agreement applies must be identical in each case: it might be the agreement of modifiers with the head of a noun phrase, subject-verb agreement, and so on. Thus the two nouns must be in the same environment.

Clause (iii) requires that the lexical item which stands as the agreeing element or target must be the same. Since not all lexical items have the same agreement possibilities, it would not do to use, say, in one instance an adjective which distinguished gender and in the other an adjective which did not, nor adjectives which distinguished different numbers of genders. The possibilities for gender agreement can vary according to the syntactic construction, and so for comparison this variable must be held constant. And within the same syntactic construction, lexical items may differ as to whether or not they show agreement in gender or as to the number of gender forms they distinguish. In all instances we are interested in agreement domains and lexical items which allow the largest number of forms; by specifying that identity must be found 'whenever' the conditions listed are met, we ensure that the domain most favourable to gender agreement and the most differentiated agreement target will be included. Clause (iii) ensures that the nouns are tested in an identical way. Then if the same result follows with the two nouns, they must be in the same agreement class.

Let us first consider French, for a straightforward illustration of agreement classes:

(1) un grand garçon (compare: *une grande garçon)
    a big boy
(2) un grand jardin (compare: *une grande jardin)
    a big garden
Gender and noun classes

In these examples we have ensured that the nouns being tested occur in identical conditions: they stand in the same morphosyntactic form (the relevant feature specification is singular), in the same agreement domain (agreement of modifiers within the noun phrase), and the lexical items involved as agreement targets are the same (un- 'a' and grand- 'big'; either would be sufficient). The nouns garçon 'boy' and jardin 'garden' require the article and the attributive adjective to stand in the same form ((1) and (2)). If we consider other possible agreement targets, or if we change to the plural, we still find that the agreements required by garçon 'boy' and jardin 'garden' are identical. They therefore belong to the same agreement class. Now compare these examples:

(3) une grande femme (compare: un grand homme)
    a big woman

(4) une grande fleur (compare: un grand fleur)
    a big flower

The nouns femme 'woman' and fleur 'flower' differ from garçon 'boy' and jardin 'garden' and require the same agreements as each other ((3) and (4)). They belong to the second agreement class. There are many thousands of nouns which behave like garçon in the first test frame. Many of them denote male humans and so the gender which they form is conventionally called the 'masculine gender'. However, there are also many nouns like jardin 'garden', which denote inanimate but which take the same agreement as garçon 'boy', and so are also members of the masculine gender. Similarly, there are many thousands of nouns like femme 'woman', some denoting females and some not (like fleur 'flower'), which make up the feminine gender. We thus have two genders. To establish the gender of a given noun, we can try it in the frames in (1) and (3). This will work provided that, for instance, the meaning of the noun allows us to use grand(e) 'big' felicitously; for some nouns it will be necessary to change the test to allow for such factors. Similarly, some nouns typically do not occur with the article in French.

While the agreement class approach deals easily with a language like French, another Romance language, namely Romanian, provides a more interesting challenge. Its gender system has given rise to a considerable literature reflecting continuing debate. Consider the following data (from Mallinson (1984:441)) showing adjectives (in the predicate this time, and we assume an appropriate context for each) agreeing with the nouns bărbat 'man', scaun 'chair', and fătă 'girl'.

(5) bărbatul e bun
    man.mnr. is good
    "the man is good"
246  Givotte G. Corbett

(6)  scaunul  e bun
cabin.DEF is good
    'the chair is good'

(7)  fata  e bună
girl.DEF is good
    'the girl is good'

(The definite article is postponed; in nouns like fata its effect is to change the quality of the final vowel, mainly by lowering, to fat.) The evidence so far demonstrates the existence of two agreement classes, one including nouns like bărbați and scaun, and the other comprising nouns like fata. There is a second case (genitive-dative), but in the singular bărbați and scaun always take identical agreements while fata differs. But the situation is more complex, as is revealed when we consider the same examples in the plural:

(8)  bărbații  sint  buni
     men.DEF are good
     'the men are good'

(9)  scaunul  e bun
     chairs.DEF are good
     'the chairs are good'

(10) fetele  e bună
     girls.DEF are good
     'the girls are good'

If we had only the data of (8)–(10), then we would postulate two agreement classes: one for nouns like bărbați and one for nouns like scaun and fata (the oblique case in the plural shows the same pattern). The argument, a long-standing one, is whether we have two genders or three. The reason for the dispute is that nouns like scaun have no agreement forms which are used uniquely for them. In terms of agreement classes, however, the situation is clear -- we should set up three classes as follows.

I. nouns taking the agreement -î in the singular and -i in the plural (bărbați)
II. nouns taking the agreement -f in the singular and -e in the plural (fata)
III. nouns taking the agreement -î in the singular and -e in the plural (scaun)

Thus we have an unambiguous answer: there are three agreement classes, and there is no reason not to recognize each as a gender. However, just saying that Romanian has three genders might suggest that it is like German, Russian, or Tamil, though these languages have a rather different gender system. All of them

2 We use zero for convenience but this does not imply acceptance of zero morphemes.
have some agreement forms which are unique to each gender. The agreement
class approach leads us to the number of sets into which nouns are to be divided
(alternatively, the number of values of the category gender for noun phrases). It
is certainly the case that bârbat, seama, and femei (and the hundreds of other nouns
similar to each of them) require three different labels. Nevertheless, the agreeing
forms (suffixes) are simpler in their morphology than is implied by the statement
that Romanian has three genders. We should therefore differentiate controller
genders, the genders into which nouns are divided, from target genders, the
genders which are marked on adjectives, verbs, and so on. The distinction is
illustrated in figure 4.1. This figure shows that Romanian has two target
 genders in both singular and plural: it has three controller genders, indicated by
the lines and labelled I, II, and III. I is usually called 'masculine', II is the 'feminine'
and III is the disputed gender sometimes called 'neuter' and sometimes
'ambigeneric'.

Diagrams like figure 4.1 can be labelled in various ways and we should con-
sider the alternatives. The first target gender is designated 'o' in the singular
on the basis of adjectives like bun 'good'. However, not all adjectives take this
form: aspr-a 'rough' has -a, as shown by comparison with aspr-ă (feminine)
corresponding to bun-ă. We have chosen to give typical allomorphs for each
target gender. This labelling avoids the danger of premature naming of genders:
on the other hand, problems can arise when the typical forms chosen suggest
similarities which are not general through the system. (Taking Latin adjectives,
we might suppose that the feminine singular -a is equivalent to the neuter plural
-ă; however, although many adjectives have identical morphological realiza-
tions for these two morphosyntactic forms, not all do.) A way of avoiding the
latter problem is to list all the allomorphs, but this can become unwieldy. Since
it is hardly practical to keep referring to strings of allomorphs, names such as
masculine, feminine, and neuter tend to be preferred. Let us consider French
again in this form (figure 4.2). In a language of this type, it is natural to use
the same labels for the sets into which nouns are divided (controller genders)

3 Cartsal-Ar Curty (1994:771) makes the interesting claim that such instances always involve
number, never any other category.
and the sets of agreeing forms (target genders). When this usage is carried over into more complex systems, then difficulties arise. Indeed, although the distinction between controller and target genders may seem an obvious one, there are several examples in the literature of the number of genders being given for a particular language, in cases where the situation is complex, without any indication as to what is meant. While there are many languages where the number of controller and target genders are the same, mismatches of the type we have seen in Romanian are not uncommon.

We have seen that the agreement-class approach gives useful results, for an analysis of controller genders, and that it must be supplemented by the notion of target genders. A further issue which arises is that the number of agreement classes may be considerably larger than the traditional (and often intuitively satisfying) number of genders generally accepted for a given language. While it is important to identify all the classes of nouns which differ in their agreement possibilities, the initial analyses which result can be unsatisfactory for two reasons: first, they miss generalizations; and second, they make similar systems appear more different than they really are. We shall, therefore, look at methods by which the number of agreement classes may be reduced, in principled ways, to give a lower number of genders (for much more detail, see Corbett (1991: 161–88)). To give an accurate account of the data of a given gender system, it is important that these steps, and the justification for them, should be made clear, rather than that the genders be presented as self-evident, a mere set of labels for dictionary entries.

First, then, there are subgenders. These are agreement classes which control minimally different sets of agreements (agreements differing for at most a small proportion of the morphosyntactic forms of any of the agreement targets). These are well attested in Slavonic languages. If we take the South Slavonic language Slovene as an example, we find three agreement classes whose nouns control very different sets of agreements. But within one of the classes there is a further distinction, found in one form only (agreement with masculine nouns in the accusative singular). According to our definition we have four agreement classes. Looking at the system as a whole it makes more sense to talk of three
Gender and noun classes

genders (required for the statement of most agreement rules, including subject-predicate agreement) and two subgenders, animate and inanimate, which have a much smaller place in the agreement rules. (For recent work on subgenders in the Network Morphology framework, see Brown (1998).)

Inanimate genders are agreement classes which comprise a small number of nouns, whose agreements can be readily specified as an unusual combination of forms available for agreement with nouns in the normal genders. For instance, in the Daghestanian language Lak, there are four genders. But there is one noun which does not fit into the four-gender system, namely ġala ‘house’. This noun takes gender III agreement in the singular and gender IV in the plural.

We should treat it as an individual exception, an inanimate gender rather than a fifth gender. Other languages have a similar situation, with a few such nouns behaving as in one gender in the singular and another in the plural. In all these cases the nouns should be lexically marked as exceptional. The situation is different from that found in Romanian, where the neuter/ambiguous genres, which take masculine agreements when singular and feminine when plural, are counted in hundreds and not in ones and twos.

While the cases discussed may appear uncontroversial, it is worth pointing out that if the first published analysis of a language takes a different approach, the existence of a larger number of genders can be perpetuated through the literature. The decision as to whether to treat a particular group of nouns as a full gender, or to mark them as exceptional, (an inanimate gender) is less important than the need for explicitness. It is profoundly misleading to say that a language has eight genders, if four contain many thousands of nouns and four count a handful each. So long as this situation is made clear, the label is secondary.

It is important to note that, for inanimate genders, the nouns can be given an exceptional marker (for an unusual pairing of singular and plural target gender forms) which would allow the normal agreement rules to determine the required markers. Thus Lak ġala ‘house’ is gender III/singular and gender IV/plural. It does not follow that any agreement class with a small number of members is necessarily inanimate, since it may not be possible to give all the nouns an irregular marker in this way. An interesting case is Leleri (a Togo Ramnai language, in turn part of Niger-Kordofanian, spoken in the Volta region of Ghana by 14,000 people at the 1960 census), where there are small numbers of nouns which require unique agreement forms and must be recognized as

* An inanimate meeting is one at which there are insufficient appropriate persons present to take decisions; hence an inanimate gender is an agreement class with insufficient nouns to deserve being labeled a gender. But note, as the text makes clear, that the number of members is not the only criterion; there is also the question of whether the agreements can be characterized as an unusual combination of forms available for agreement with nouns in the normal genders.
Table 4.1 Agreement patterns in Russian

<table>
<thead>
<tr>
<th>attributive adjective</th>
<th>predicate</th>
<th>relative pronoun</th>
<th>personal pronoun</th>
<th>traditional gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>-yj</td>
<td>-a</td>
<td>-ajk</td>
<td>on-ajk</td>
<td>masculine</td>
</tr>
<tr>
<td>-yju</td>
<td>-a</td>
<td>-ajk</td>
<td>on-ajk</td>
<td>feminine</td>
</tr>
<tr>
<td>-yke</td>
<td>-o</td>
<td>-ne</td>
<td>on-o</td>
<td>neuter</td>
</tr>
</tbody>
</table>

comprising full genders (for these see Heine (1968:114–15, 1982:197–8); and Corbett (1991:173–5)).

The question of differences in targets has already been mentioned. In some languages, all targets mark the same distinctions; we may then take only one target type and use it to establish the agreement classes. In other languages, different targets make a greater or lesser number of distinctions; in such cases, we include the target type which marks most distinctions when establishing agreement classes (for more complex cases, see Corbett (1991:176–7)). Given that different targets may show gender agreement, an important part of the analysis, and one which is normally passed over in silence, is the linking of these forms into consistent agreement patterns. Russian can again serve as an example. Earlier, three main target gender forms were noted. For the attributive adjective, we have the nominative singular endings: -yj, -ajk, and -yke. The past tense verb, taken here as the representative of the predicate, has three possibilities in the singular: the bare stem, and the endings -a and -o. The relative pronoun has the same endings as the attributive adjective, and the personal pronouns are on, ona, and ono. For the vast majority of nouns, the agreements are as given in table 4.1, which represents a simplification; the full version would include the other cases, the plural number, and the animate and inanimate subgenders. From the data given, relating to the main genders, the distinctions made by each target gender type are identical. The question is how this analysis is done. It is not as obvious as it appears, since there are nouns which take other combinations of agreements, as we shall see shortly. There are two important factors. The first is that the vast majority of nouns which take agreements with -yj, also take -a and on; the second point is that these are nouns for which we can give absolute rules: they always take the same agreements. Each horizontal line of table 4.1 represents a consistent agreement pattern, which we define as follows:

A consistent agreement pattern is a set of target gender forms such that:
(i) the agreement class it induces is as large as possible;
(ii) agreement rules relating to this agreement class will be simple and exceptionless.
Gender and noun classes

A consistent agreement pattern links all the target gender forms of a given gender. The notion of consistent agreement pattern thus gives us a principled way of capturing the intuition that, for example, a feminine marker on an attributive modifier is 'the same as' a feminine marker on a verb, even if they are phonologically different. Thus, it is needed even for languages where different targets mark similar distinctions, as well as for languages where they do not (like Yimas, for which see Foley (1986:66–7, 89, 199):119–62; Corbett (1991:176–7)).

It is the notion of consistent agreement pattern which allows us to understand and distinguish two types of noun which complicate the analysis of gender systems in interesting ways. First, there are nouns which can take all the agreements of more than one gender. These are often called nouns of common gender, particularly where the noun denotes a human and may take masculine or feminine agreements depending on the sex of the human referred to. While nouns which can be masculine or feminine according to the sex of the referent are widely attested, some languages allow nouns to take two genders according to the size and shape of the referent (one such is Manambu, a New Guinean language: Aiikhevald (2000:42)). Second, there are nouns which do not simply take the agreements of a single consistent agreement pattern nor belong to two or more genders. Rather, the agreement form used with them depends in part on the type of target involved. Such nouns are termed hybrid nouns. One such case is the Russian noun vraè in the meaning 'female doctor', which can occur in the following constructions among others:

(11) nov-yj vraè
    new-MASC doctor
    'the new (female) doctor'

(12) nov-qaja vraè
    new-FEM doctor
    'the new (female) doctor'

Examples like (11) are more common than (12): recall that we are considering cases where it is a female doctor. In the predicate, the feminine is somewhat more common:

(13) vraè rabotal-s
    doctor worked-MASC
    'the (female) doctor worked'

5 It is possible too for certain nouns to be of different genders according to the sex of the speaker; this is found in Garifuna, a member of the Arawak family spoken in Belize, Honduras, Nicaragua, and Guatemala (Taylor (1972:69); Munro (1998)).

6 See Dahl (2000) for a different approach to the problem.
(4) vrac rabešal-a
    doctor worked-FEM
    'the (female) doctor worked'

The relative pronoun is usually feminine and the personal pronoun is normally feminine (though even here the masculine is possible); in summary the agreements are like this:

<table>
<thead>
<tr>
<th>Type</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>attributive modifiers</td>
<td>usually masculine, feminine possible</td>
</tr>
<tr>
<td>predicate</td>
<td>both possible</td>
</tr>
<tr>
<td>relative pronoun</td>
<td>normally feminine, masculine rare</td>
</tr>
<tr>
<td>personal pronoun</td>
<td>normally feminine (masculine just possible)</td>
</tr>
</tbody>
</table>

In this remarkable case, the agreement required is variable for all the different types of target. Vrac denoting a female is a 'hybrid' noun since it does not take consistently feminine agreements, nor consistently masculine agreements, nor both. The notion of consistent agreement pattern thus allows us to separate nouns like vrac from ordinary nouns like, say, ženičina 'woman' and mučina 'man'; the agreements taken by the latter two each form a consistent agreement pattern whereas those of vrac do not. The distinctive feature of hybrid nouns (like vrac) is that the choice of form to be used with them depends in part on the target type. The possible patterns of agreement with such nouns are constrained by the Agreement Hierarchy (Corbett 1991:225-60).  

As with inanimate genders, it is important that our descriptions which include details of nouns with double or multiple gender and of hybrid nouns should give an indication of how prevalent these types are. Moreover, the class of hybrid nouns may be, not surprisingly, rather disparate: the different agreements may occur with very different frequencies from noun to noun.

We have seen that, in establishing the number of genders in a particular language, Zajenjak's approach, based on the notion of agreement class, is a

---

7 The Agreement Hierarchy consists of four target types, which have been introduced earlier in the section. Possible agreement patterns are constrained as follows: For any controller that permits alternative agreement forms, as we move right to left along the Agreement Hierarchy, the likelihood of agreement forms with greater semantic justification will increase monotonically (that is, with no intervening decrease). The agreements found with Russian vrac 'woman' occur in this pattern, since syntactic agreement (masculine) is the more common option in attributive position, with semantic agreement (feminine) becoming steadily more likely as we move along the hierarchy.

attributive < predicate < relative pronoun < personal pronoun

---

Figure 4.1 The Agreement Hierarchy
Gender and noun classes

useful starting point. It is important to bear in mind, however, that this approach
leads us toward controller genders: the other side of the coin is the system of
different forms of the agreeing elements: we termed these forms target genders.
While in some languages there are numerous agreement classes, not every
agreement class is necessarily recognized as a gender. To find how many genders
a language has, we begin with agreement classes, separating out all the different
sets of nouns according to the different agreements they take. However, we do
not automatically accept each set as a gender, as we have seen in this section.
This minimalist position in some cases leads us to traditional analyses.

1.2 Classifiers and complex systems

Since agreement is a prerequisite for a gender system, classifiers are to be seen
as a different phenomenon. Classifiers are of various types, the best known
being numeral classifiers. In a language with prototypical numeral classifiers,
noun phrases including a numeral and a noun will normally have a third ele-
ment: the classifier. Thus in the Sino-Tibetan language Burmese, ‘one river’
might in the appropriate circumstances (where the context involves a river
on a map) be translated as myiθ to lwe, literally ‘river one line’ (the last ele-
ment is the classifier). Such classifiers are free forms, often appearing also as
fully fledged nouns. This is illustrated in the following example (the unmarked
case for ‘river’): myiθ to myiθ ‘river one river’, here the noun is repeated as
a classifier. In such systems classifiers frequently do not co-occur with cer-
tain nouns, and for others classifiers may be obligatory or their use may vary
according to speech style. Often, different classifiers are possible with the same
noun, and the choice depends on meaning. Thus, apart from the two classifiers
already given with the Burmese noun myiθ, there are several other possibil-
ties, such as myiθ to hmai, ‘river one section’, when treating it as a fishing
area, and myiθ to pa, ‘river one sacred object’, when discussing mythology
(Hecker (1975:113)).

Some languages have classifiers which are not restricted, but occur freely
in ordinary noun phrases (‘noun classifiers’ rather than ‘numeral classifiers’).
One such is Karo, a member of the Ramarama branch of the Tupi family, with
about 150 speakers in the Brazilian Amazon (data from Nilson Cabas, Jr.). There
are eleven classifiers, mainly classifying according to shape. Again different
classifiers may occur with a given noun:

\[
\begin{align*}
(15a) & \quad \text{iyá kap} & (15b) & \quad \text{iyá peθ} & (15c) & \quad \text{iyá ?a?} \\
\text{stone ct.together} & \quad \text{stone ct.flat} & \quad \text{stone ct.round} & \quad \text{‘gravel’} & \quad \text{‘digging stick’} & \quad \text{‘stone’}
\end{align*}
\]

The Australian language Yidiŋ also has noun classifiers; furthermore, in Yidiŋ
two classifiers may be found together with a single noun (Dixon (1982:192)):
bama wuguja wurgun
person man pubescent boy
'teenage boy'

Here the first two elements are both classifiers. These are not part of an agreement system, and so are a different phenomenon from gender. These classifiers do not show variation of a formal property (as is the case when, say, an adjective marks agreement in gender), rather the selection of one classifier as opposed to others is involved. The types of classifiers we have discussed are independent items, selected largely according to semantic criteria, while gender markers typically appear attached to agreement targets. These differences between classifiers and genders (or noun classes) are drawn clearly by Dixon (1982, 242–18); see Labov (2000) for further discussion. There are some similarities too; the selection of classifiers is based on principles which partly resemble the assignment rules investigated in section 2.

Dixon (1982) gives a helpful account of the extremes (which are actually quite commonly found). But things have turned out to be much more complex. Various new types of classifier have been identified, since Allan’s analysis (1977), including incorporated classifiers (Mithun 1986); revised typologies of classifiers are proposed in Craig (1997) and Aikhenvald (1994, 2008–14, 2009). Again, it is worth noting that terms are used in confusing ways in the literature. What is a classifier system for one author would be a noun-class system for another.

Between the poles of clear gender systems and clear classifiers, there are classifier-like elements which are affixes rather than free forms; thus in Tuyuka, a Tucanoan language of Colombia and Brazil, classifiers form a single phonological word with the item to which they attach (Barnes (1990, 273)). So classifiers are more heterogeneous than was thought, and some are rather similar to gender/noun-class systems. Derbyshire and Payne (1990) give a survey of systems found in Amazonian languages, showing both that there are various intermediate types of classifier and that different systems of classification, including gender systems, can coexist in the same language. A striking case is Tariana, a North Arawak language of Brazil (Aikhenvald 1994) which has three sub-types of classifier and a gender system. This type of coexistence is not restricted to the Amazon area: Donga (a member of the Mba group, which belongs to the Ubangi branch of Niger-Kordofanian) has a complex gender system and is in addition developing a system of possessive classifiers (Pasch 1985, 1986, 245–55). There are examples in Australia too: Reid (1997) shows how in Ngan’gityemneri (a southern Daly language with...

---

8 This particular combination of classifiers is unusual in Yida, consisting of two 'inherent nature' classifiers; more often one is an 'inherent nature' classifier and one is a 'function/use' classifier. See Dixon (1977, 485–86) for details.
Gender and noun classes

about 150 speakers some 300 kilometres southwest of Darwin there are both
gender-like agreements and classifiers, with evidence that the classifiers are
in the process of developing into gender markers (as suggested in Greenberg
(1978)).

Ngan'gityemirri has arguably fifteen genders, with semantic assignment
according to criteria such as male, female, canine, non-human animates (other
than canines), striking instruments, and separate genders for two types of spear.
Nine genders are distinguished by the bound agreement markers found on
agreement targets, such as adjectives (Reid (1997:181)):

(17) a-syensyerrgini
     a=tyentenmmuy
     ANIMATE-white.rock.wallaby ANIMATE-tame

'a tame white rock wallaby' ("a" is used for clitics and "=" for affixes)

Six genders have (optional) freeform generics/classifiers (Reid (1997:177)):

(18) (syiri) magulu
     (syiri) margu
     STRIKE cylindrical.fighting.stick STRIKE new

'Syiri is the freeform generic for weapon-like objects which have a striking type
of contact. In its first use in (18) it is analogous to a classifier. In its second use it
is more like an agreement marker. At first sight we might think the language has
two different systems, but this is not the case, since in some genders there is a
generic available in addition to a marker on the noun and to a bound agreement
marker. Moreover, while the use of the generic is optional, so too is agreement
(Reid (1997:468)). Ngan'gityemirri provides a clear window on the rise of
gender systems and of agreement systems. Reid charts the likely development
from freeform generic to bound agreement marker, in a system in which the
generics are still feeding the gender system (pp. 211-22).

Surprisingly, perhaps, two systems of the gender type can indeed coexist in
a single language. This rare situation is found in the Mba group (Ubangian
branch of Niger-Kordofanian; data from Tucker and Bryan (1966:110, 114-23,
131-40); Pusch (1985:69-71, 1986)). Here we find a system with several
distinctions but none based on sex (which is somewhat similar to the type found
in Bantu languages) and a second system distinguishing up to four members:
male human, female human, animal, and inanimate. The latter system, based on
semantic criteria, is a later development in the Mba group. The four languages
of the group show four different possibilities. Ndunga has only the Bantu-like
system and so is straightforward – the new development has not affected it. At
the other extreme, Ma has lost the earlier system and has only a semantically
based system; this four-gender system is found in pronouns, and elsewhere
there is an animate–inanimate distinction. Ma too is therefore unproblematic.

A similar situation, with an extensive
system, is found in Mitaña, a
Witotoan language of Colombia,
Dongo, however, has both types of system. For example, verbs agree in animacy (an animate–inanimate distinction) while adjectives preserve a Bantu-like set of agreements. When we look more carefully at adjectival agreement, we find that agreement according to the gender which can be predicted from the morphological class of the noun occurs only with inanimates. Animate, irrespective of their morphology, take the agreements of one specific gender. When we analyse Dongo using the agreement-class approach we arrive at precisely the same genders which could be identified by looking solely at the adjectival agreement forms. Thus, though the gender system of Dongo has two separate origins, the systems have fused and can readily be described using the approach we have developed. (It also has classifiers, as noted above.) The most interesting of the Mba languages in terms of gender is Mba itself. Mba has several inanimate genders, and one or two whose size is not fully clear. To avoid exaggerating the problem, let us concentrate on the well-established genders. If we examine agreement within the noun phrase, we can distinguish six agreement classes. There is also a personal pronoun, used only for animates, which distinguishes male human from other animate in the singular, and has one form for the plural. These pronoun forms can be used optionally as agreement markers, which, it can be argued, gives a three-way distinction: male human versus other animate versus inanimate (for which no optional pronoun is available). Since attributive modifiers show a six-class system and the optional markers divide nouns into three classes, we might expect to find eighteen possibilities (eighteen agreement classes). However, some of the six classes established on the basis of attributive modifiers include only inanimates, so that there are eleven rather than eighteen possibilities. That is to say, the two systems are not fully independent; animacy is a determining factor in both of them. (For a full analysis in terms of agreement classes, see Corbett (1991: 84–8.) Thus, there are indeed two systems, but they are not fully independent of each other.

Pumari is a language which comes closer to having two independent gender systems. It belongs to the Arawá family, and has around 500 speakers living in the state of Amazonas, Brazil. The language has been described by Chapman and Derbyshire (1991), and its gender system specifically by Aikhenvald (unpublished ms.). Pumari distinguishes masculine from feminine. All nouns denoting females are feminine; so are body parts and the majority of artifacts. Most insects and fish are masculine, and so on. There is a second binary distinction: some nouns (under circumstances we will come to) control a ka-form (sometimes *a-, *ko- or *ki-) and others do not. The ka-class is the minority. Again, assignment is complex: larger and flatter objects tend to be in the ka-class; for instance, *kururi 'padle', *kajiriri 'island'. 'Substances which consist of smaller particles, or are thick in texture' (Aikhenvald (unpublished ms.)) take ka- *jokri 'salt', *kajahri 'banana mash'. The ka-class does not include any nouns denoting humans.
Gender and noun classes

Table 4.2 Interaction of genders in Faumani

<table>
<thead>
<tr>
<th>masculine</th>
<th>feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka-</td>
<td>karita 'crocodile'</td>
</tr>
<tr>
<td>non-ka</td>
<td>aravaha 'rooster'</td>
</tr>
<tr>
<td></td>
<td>araho 'land, ground'</td>
</tr>
</tbody>
</table>

The description of the domains in which these agreements operate is complex (Alkhovenaid [ms.]); demonstratives and some adjectives agree with the head noun in gender; adjectives also take ka- when appropriate, as do suasive verb forms used attributively. Certain types of possesive construction involve both types of agreement. In verb agreement, some verbal affixes require gender agreement and verbs also take ka- (both according to an ergative pattern, though this is being eroded). Consider this example (Chapman and Derbyshire 1991:255):

(19) o-ka-rofi-ki oni vanami ka-haro
     I.SG-ka-verb-non.thematic DEM.FEM paddle.FEM KA-big
     'I want the big paddle'

The noun vanami 'paddle' is feminine, and belongs to the ka-class. The demonstrative shows agreement in gender, and the adjective has a ka-marker. The verb also has a -ka- marker, showing agreement with the direct object (the suffix -ki is not one of those which marks gender). Thus the domains of gender and ka-class agreement are rather different. A further difference is that the masculine and feminine agreement markers are opposed to a plural marker (and so gender is not differentiated in the plural). Ka-agreement, on the other hand, does not interact with number; it is retained in the plural.

The independence of the two systems is shown in assignment too: nouns are found showing each of the four logical possibilities (see Table 4.2, data from Alexandra Alkhovenaid). However, though all four types of noun are attested, the nouns are distributed unequally over the four types, in a way which shows that there are connections between the two criteria. Thus nouns in the ka-class are normally feminine (there are few masculines in the ka-class).

An equally remarkable case is Michif, the language of the descendants of French Canadian fur traders and Cree speakers in western Canada (Bakker 1997; Bakker and Papen 1997:315-16). The language has both the masculine–feminine distinction of French, and the animate–inanimate distinction of Cree (which is typical for Algonquian languages, of which Cree

*Indicates a glottal stop, which is a voiced implosive.
is one). Nouns are independently assigned to masculine or feminine and to animate or inanimate; broadly speaking, agreement within the noun phrase is according to the French-type masculine–feminine distinction, while agreement of the verb (and of demonstratives) follows the Cree-type animate–inanimate split.

It remains to be seen what types of different systems can coexist, and how independent they can be (see discussion in Aikhenvald (2000:61–77)). The data from languages like Paunmir and Michif, which have become available in recent years, are fascinating and tantalizing; we must hope for further detailed descriptions of languages with systems of comparable complexity.

2 The speaker's problem: gender assignment

We have considered the analytical problems of determining the number of genders in a particular language. This analysis requires us to establish the syntactic configurations which allow us to separate nouns of different genders and so it provides the test situations for determining the gender of a given noun. Let us now imagine that we have the nouns of a given language all allotted to the appropriate gender (an easy task in some languages, a more difficult one in others). The question is then whether we have a random collection of nouns in each gender or not. From the point of view of the speaker of the language (who must clearly 'know' the gender of a noun in order to produce the examples of agreement that serve as input to our analysis), is it necessary to remember/store the gender of each noun individually? Despite statements in the literature about the apparent arbitrariness of gender in some languages, I claim there is always a system behind the distribution of nouns over the genders. This system, an assignment system, is a model of the native speaker’s ability to allot nouns to genders on the basis of information which must in any case be stored as part of the lexical entry. This is not to claim that assignment systems are exceptionless; however, even in the most complex cases, when carefully analysed, have shown that gender is predictable for at least 85 percent of the nouns, and the figure is normally much higher than that.

Assignment may depend on two types of information: semantic and formal (the latter being a cover term for morphological and phonological information). In one sense all assignment systems are semantic, since genders always have a semantic core (there are no purely formal systems). However, there are systems in which semantic information is sufficient for gender assignment and it is these which we call semantic assignment systems.

16 For an extended discussion see Corbett (1991:1–65), and for additional references see Chitt (1993:409, 1995).
Gender and noun classes

Table 4.3 Gender assignment in Godoberi

<table>
<thead>
<tr>
<th>criterion</th>
<th>gender</th>
<th>example</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>male rational</td>
<td>masculine (m)</td>
<td>ima</td>
<td>father</td>
</tr>
<tr>
<td>female rational</td>
<td>feminine (f)</td>
<td>ila</td>
<td>mother</td>
</tr>
<tr>
<td>other</td>
<td>neuter (n)</td>
<td>lemanXa</td>
<td>dogkey</td>
</tr>
</tbody>
</table>

2.1 Semantic assignment

Let us consider the gender system of Godoberi, a language of the Andic subgroup of the Avaric group of Daghestanian (Northeast Caucasian) languages, with about 2,500 speakers living in the Botlikh area of Dagestan. The data are primarily from Kibrik (1990); a sketch of the language can be found in Gudava (1969). There are three genders, which can be labelled 'masculine', 'feminine', and 'neuter'; as mentioned earlier, because these classes of nouns are so transparent semantically in comparison with Indo-European genders, Caucasianists sometimes prefer the term 'noun class', and they then use numbers (I, II, III) instead of names. Nouns are assigned to the three genders / noun classes as shown in Table 4.3. This assignment system is simple and operates consistently. 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.

2.2 Predominantly semantic assignment

We now move on to languages which have semantic assignment rules which appear to allow sets of exceptions. These exceptions may not be a significant proportion of the nouns in the languages, but they cannot be dismissed as mere sporadic exceptions. A good example is provided by Archi, which has four main genders (data from Kibrik, Kodzasov, Olovjanikova, and Sargsyan (1977:55–66)). The first two genders are straightforward: male rationals make up gender I: ala ‘father’, doga ‘grandfather’, allah ‘God'; and female rationals constitute gender II: arota ‘grandmother’, hana ‘aunt’, qaraj ‘witch’. There are no non-rational in these genders. Genders III and IV are more complex, as we see summarized in Table 4.4. In this system we find overlapping semantic criteria. Note that for animals, sex is of no importance: the words for ‘cow’ and ‘bull’ are both in gender III. There is a division between domestic animals and birds

---

11 Rational is a term frequently found in accounts of gender classification; it is almost equivalent to ‘human’, but often includes various mythical beings and often excludes infants.
<table>
<thead>
<tr>
<th>Gender III</th>
<th>Gender IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>dromitic animals and birds</td>
<td>young animals and birds (wild and domestic)</td>
</tr>
<tr>
<td><em>gúny</em> 'cow'</td>
<td><em>bul</em> 'calf', <em>lú'or</em> 'foal (of donkey)'</td>
</tr>
<tr>
<td><em>gụp</em> 'donkey', <em>gá</em> 'goose'</td>
<td>smaller wild animals and birds</td>
</tr>
<tr>
<td>larger wild animals and birds</td>
<td><em>ejem</em> 'bark', <em>njirnúta</em> 'monkey',</td>
</tr>
<tr>
<td><em>ló</em> 'elephant', <em>jum</em> 'wolf', <em>liñ</em> 'eagle', <em>išu</em> 'owl'</td>
<td><em>bíd-bíd</em> 'hare', <em>bíhba</em> 'swallow'</td>
</tr>
<tr>
<td>all insects</td>
<td>musical instruments</td>
</tr>
<tr>
<td><em>fík</em> 'fly', <em>nimh</em> 'moth'</td>
<td><em>pip</em> 'spade', <em>dák</em> 'awl', *kó' 'knife'</td>
</tr>
<tr>
<td>mythical beings</td>
<td>cloth, most clothing</td>
</tr>
<tr>
<td><em>bí</em> 'gente', <em>líbí</em> 'devil'</td>
<td><em>sí</em> 'suit', <em>palí</em> 'trousers', *kó'</td>
</tr>
<tr>
<td>musical instruments</td>
<td><em>shi</em> 'shawl', *yi' 'sleeve'</td>
</tr>
<tr>
<td><em>pá</em> 'drum', <em>mók</em> 'tambourine'</td>
<td>metals</td>
</tr>
<tr>
<td>cereals</td>
<td><em>li</em> 'iron', <em>qal</em> 'tin'</td>
</tr>
<tr>
<td><em>gôó</em> 'wheat', <em>máá</em> 'barley'</td>
<td>liquids</td>
</tr>
<tr>
<td></td>
<td><em>ń</em> 'water', <em>ést</em> 'wine'</td>
</tr>
<tr>
<td></td>
<td><em>fáy</em> 'tears', <em>x</em> 'lel* 'rain'</td>
</tr>
<tr>
<td>trees</td>
<td>abstracts (including some temporal concepts)</td>
</tr>
<tr>
<td><em>bád</em> 'tree', <em>ká</em> 'fir'</td>
<td><em>ká</em> 'high summer', <em>sé</em> 'autumn',</td>
</tr>
<tr>
<td></td>
<td><em>kó</em> 'day', <em>kó</em> 'year', <em>kú</em> 'beauty',</td>
</tr>
<tr>
<td></td>
<td><em>bá</em> 'tiny'</td>
</tr>
<tr>
<td>water phenomena</td>
<td>stars</td>
</tr>
<tr>
<td><em>já</em> 'sea', <em>bá</em> 'lake', <em>bí</em> 'whirlpool', <em>wá</em> 'ice'</td>
<td></td>
</tr>
<tr>
<td>astronomical and meteorological phenomena</td>
<td></td>
</tr>
<tr>
<td><em>bá</em> 'moon', <em>bá</em> 'sun', <em>má</em> 'snow',</td>
<td></td>
</tr>
<tr>
<td><em>gá</em> 'sunrise', <em>gá</em> 'sunrest'</td>
<td></td>
</tr>
</tbody>
</table>

(all III except their young) and wild animals and birds. The latter divide into larger (III) and smaller (IV), though *naq* 'cat' is unexpectedly in gender III. The young of animals and birds are in gender IV. There is thus a correlation between large (III) and small (IV), which is confirmed by examples of nouns denoting concrete objects (many of which are otherwise problematic). We find pairs like *fáh*n (III) 'town', *x* 'or' (IV) 'village', and *x* 'it (III) 'scotop' and *x* 'it (IV) 'spoon'.

A second correlation is that concrete objects tend to be in gender III and abstracts in gender IV, as seen from the last categories in table 4.4. There are further possible connections: there are similarities between cloth and liquids (both IV), both being non-count and non-rigid (though the same could be said of cereals, which belong in gender III). And the items listed under 'water phenomena' (III) are typically larger, more specific instances of liquids than the general terms in IV. We may say that prototypical members (that is, best or most central instances) of gender III are concrete and large (pil 'elephant', ká 'fir',...
Gender and noun classes

buc 'moon'). The typical member of IV is not. In other instances the motivation for assignment to gender III or IV is not straightforward; this is particularly true for nouns denoting inanimates not covered by the criteria given.

In systems of the predominantly semantic type, assignment is based on semantics, but not in the absolutely straightforward way which we observed in Godoberi. Languages of this type may be hard to analyse when approached with conventional semantic features in mind; an understanding of the world view of the speakers may make the gender system more comprehensible. A well-known case of this type is Dyirbal, for which see Dixon (1982).

2.3 Morphological assignment

We have looked at Godoberi, in which semantic criteria are sufficient for assignment, and Archi, which allows various numbers of exceptions (and many more instances of each language type could be given). We now come to languages in which large numbers of nouns fall outside the semantic assignment rules. These nouns may be handled instead by formal assignment rules, rules which depend on the form of the nouns involved rather than on their meaning. These latter rules are of two types, morphological and phonological, which we will consider in turn (sections 2.3 and 2.4). Whereas the distinction between semantic and formal assignment rules is clear (though their effects may overlap), the distinction between morphological and phonological rules is not always clean-cut. As a starting-point, we may say that phonological rules refer to a single form (typically the most basic form) of a noun, for example, 'nouns ending in a vowel are feminine'. Morphological rules, on the other hand, require more information; they need to refer to more than one form, though this is not always obvious. A typical assignment rule of the morphological type might be: 'nouns of declension II are feminine'; establishing that a noun is of declension II might require information about, say, the nominative singular and the genitive singular. (While declensional classes are abstract notions, for our purposes the important point is that they relate to forms—the inflections—rather than to semantics.)

A clear example of a morphological assignment system is Russian, an East Slavonic language with three genders. The masculine and feminine genders have a semantic core, as can be seen from the semantic assignment rules.

Semantic assignment rules

(i) Sex-differentiable nouns denoting males (humans and higher animals) are masculine: otec 'father', syn 'son', brat 'brother', lev 'lion'.

(ii) Sex-differentiable nouns denoting females are feminine: mats 'mother', doč 'daughter', tetja 'aunt', frica 'lions'. (Note that ' transliterates the Russian soft sign, which normally indicates palatalization of the preceding consonant.)
Sex-differentiable nouns are those where the sex matters to humans (as in the case of humans and domesticated animals) and where the difference is obvious (as with lions).

While these rules operate with very few exceptions, they do not cover a large proportion of the nouns, those which make up what we call the 'semantic residue'. It is not the case that all the nouns in the semantic residue are neuter. Rather they are distributed over the three genders. This situation, a common one in Indo-European languages, is shown schematically in table 4.5. To confirm that the nouns of the semantic residue are indeed found in all three genders, consider those in table 4.6. It does not seem possible to account for the gender of these nouns by reference to their meaning. Nevertheless, gender in Russian is highly predictable; for many nouns it is determined not by semantic but by formal factors, namely by the declensional class of the noun, as we shall see. From some of the examples given already, it might appear that simple phonological rules would be sufficient; for example, concerning the final segment of a noun. Unfortunately, there are numerous examples for which no such rule works, pairs such as *porst* (masculine) ‘briefcase’ and *pyf* (feminine) ‘dust’. The forms

### Table 4.5 Gender assignment in Russian (semantic criteria only)

<table>
<thead>
<tr>
<th>gender</th>
<th>criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>masculine</td>
<td>male + residue</td>
</tr>
<tr>
<td>feminine</td>
<td>female + residue</td>
</tr>
<tr>
<td>neuter</td>
<td>residue</td>
</tr>
</tbody>
</table>

### Table 4.6 Examples from the semantic residue in Russian

<table>
<thead>
<tr>
<th>masculine</th>
<th>feminine</th>
<th>neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>zametka</em> 'notebook'</td>
<td>*zamets' 'notebook'</td>
<td><em>zamenny</em> 'notebook'</td>
</tr>
<tr>
<td><em>izba</em> 'house'</td>
<td><em>izba</em> 'house'</td>
<td><em>izba</em> 'house'</td>
</tr>
<tr>
<td><em>voda</em> 'water'</td>
<td><em>voda</em> 'water'</td>
<td><em>voda</em> 'water'</td>
</tr>
<tr>
<td><em>malina</em> 'plum'</td>
<td><em>malina</em> 'plum'</td>
<td><em>malina</em> 'plum'</td>
</tr>
<tr>
<td><em>not</em> 'night'</td>
<td><em>not</em> 'night'</td>
<td><em>not</em> 'night'</td>
</tr>
<tr>
<td><em>kaga</em> 'bone'</td>
<td><em>kaga</em> 'bone'</td>
<td><em>kaga</em> 'bone'</td>
</tr>
<tr>
<td><em>roda</em> 'hand'</td>
<td><em>roda</em> 'hand'</td>
<td><em>roda</em> 'hand'</td>
</tr>
<tr>
<td><em>belyak</em> 'arm'</td>
<td><em>belyak</em> 'arm'</td>
<td><em>belyak</em> 'arm'</td>
</tr>
<tr>
<td><em>shto</em> 'thing'</td>
<td><em>shto</em> 'thing'</td>
<td><em>shto</em> 'thing'</td>
</tr>
<tr>
<td><em>kaz</em> 'law'</td>
<td><em>kaz</em> 'law'</td>
<td><em>kaz</em> 'law'</td>
</tr>
<tr>
<td><em>glasov</em> 'oppeness'</td>
<td><em>glasov</em> 'oppeness'</td>
<td><em>glasov</em> 'oppeness'</td>
</tr>
<tr>
<td><em>deemo</em> 'trust'</td>
<td><em>deemo</em> 'trust'</td>
<td><em>deemo</em> 'trust'</td>
</tr>
</tbody>
</table>
Gender and noun classes

Table 4.7 Noun paradigms in Russian

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMINATIVE</td>
<td>салон</td>
<td>гостиница</td>
<td>касса</td>
<td>вино</td>
</tr>
<tr>
<td>ACCUSATIVE</td>
<td>салон</td>
<td>гостиница</td>
<td>касса</td>
<td>вино</td>
</tr>
<tr>
<td>GENITIVE</td>
<td>салона</td>
<td>гостиницы</td>
<td>кассы</td>
<td>вина</td>
</tr>
<tr>
<td>DATIVE</td>
<td>салону</td>
<td>гостиници</td>
<td>кассой</td>
<td>вином</td>
</tr>
<tr>
<td>INSTRUMENTAL</td>
<td>салонами</td>
<td>гостиницами</td>
<td>кассой</td>
<td>вином</td>
</tr>
<tr>
<td>LOCATIVE</td>
<td>салоне</td>
<td>гостиниче</td>
<td>кассе</td>
<td>вине</td>
</tr>
<tr>
<td>NOMINATIVE</td>
<td>салон</td>
<td>гостиница</td>
<td>касса</td>
<td>вино</td>
</tr>
<tr>
<td>ACCUSATIVE</td>
<td>салон</td>
<td>гостиница</td>
<td>касса</td>
<td>вино</td>
</tr>
<tr>
<td>GENITIVE</td>
<td>салонов</td>
<td>гостиниц</td>
<td>касс</td>
<td>вина</td>
</tr>
<tr>
<td>DATIVE</td>
<td>салону</td>
<td>гостиниц</td>
<td>касс</td>
<td>вина</td>
</tr>
<tr>
<td>INSTRUMENTAL</td>
<td>салонами</td>
<td>гостиницами</td>
<td>кассами</td>
<td>винами</td>
</tr>
<tr>
<td>LOCATIVE</td>
<td>салоне</td>
<td>гостиниче</td>
<td>кассе</td>
<td>вине</td>
</tr>
</tbody>
</table>

Note. Forms are given in a transliteration of the standard orthography, which is largely phonemic. Palatalization of the preceding consonant is indicated by /j/.

discussed so far are those of the nominative singular. Attempts using any other case form, which would in any case be harder to justify since the nominative is the basic case, are less successful than those using the nominative, so we shall not pursue them. It could be argued, however, that, since Russian has at least six cases, a phonological rule should be used not upon a particular case form but upon the stem; this more consistent approach actually fares rather worse, since the two nouns above have stems identical to the nominative singular, while other nouns such as неделя ‘week’ (feminine) also have palatalized stems (/neːdlʲ/) which cannot be distinguished from the two above.

The assignment rules require access to more than one case form of the noun, in other words, to its declensional class; they are therefore morphological assignment rules. Russian has four main noun paradigms, which account for all but about twenty of the declinable nouns (this analysis is justified in detail in Corbett (1982: 300–11)). Examples are given in table 4.7. Given these declensional classes, it is relatively simple to predict the gender of a noun.

Morphological assignment rules

(i) Nouns of declensional class I are masculine.

(ii) Nouns of declensional classes II and III are feminine (but see below).

(iii) Others are neuter.

These morphological rules are highly predictive; we might therefore ask whether the semantic rules given earlier are superfluous. The nouns which were assigned to gender because of their semantics might instead come under
the morphological rules. For instance, *oteq* 'father' would be assigned to the masculine gender because it is in declension i, and *nej* 'mother' to the feminine because it is a member (an irregular one) of declension iii. But this approach would not account for nouns like *didad* 'uncle' and *dedušot* 'grandfather'. They denote males and so should be masculine by the semantic assignment rules; at the same time, they belong to declension i, and so would be expected to be feminine. In fact they are masculine (just as in Latin, nouns like *agricola* 'farmer' are masculine). Thus there can be different predictions from the two sets of rules, and when this occurs the semantic rules take precedence.

There are further complexities, involving acronyms, indeclinables, and the animate subgender (for which see Corbett (1994)). The essential point is that many nouns are assigned to gender by the semantic assignment rules. For the semantic residue, the nouns for which there is no semantic assignment rule, the gender can be predicted by the morphological assignment rules. A more formal analysis, demonstrating that the proposed assignment rules do in fact make the correct predictions, is given in the Network Morphology account in Fraser and Corbett (1999). Such assignment systems are common in Indo-European languages; they are also found widely elsewhere, for instance in Bantu languages.

2.4 Phonological assignment

We now turn to examples of phonological systems of gender assignment, those in which gender can be established by reference to a single form. These vary in complexity. A simple case is the two-gender system of Qatar (Afar). This East Cushitic language (Cushitic forming part of Afro-Asiatic) has approximately

---

12 Our Network Morphology account is expressed in DATR, a lexical knowledge representation language devised by Roger Evans and Gerald Gazdar (see Gazdar (1998); Evans and Gazdar (1999)). This fully explicit language has a compiler, so that it is possible to check that an inheritance network expressed in DATR captures the intended generalizations. This means that we are able to demonstrate that our analysis does indeed assign Russian nouns to the appropriate genders. The importance of this is that, while the semantic and phonological assignment systems discussed are relatively clearcut, the morphological systems have sometimes been analysed differently (when the number of declensional classes and the number of genders is similar, this leads one to try to predict declensional class on the basis of gender). There is a set of arguments in Corbett (1998) showing the difficulty with such an approach. Being able to express an analysis in DATR takes the debate a step forward in that we can prove that the approach assigning gender on the basis of declensional class does give the right results. No such analysis is at present available for the alternative, which would not, of course, fit in with the typology proposed. This is an instance of the usefulness of computational techniques for typologists.

13 The complex interactions of gender and morphological form in Mayali, a non-Pama-Nyungan language of northern Australia, are analysed in Evans (1997) and Evans, Brown, and Corbett (2002). Again the analysis is implemented to demonstrate that the nouns are indeed assigned to the appropriate gender.
Gender and noun classes

250,000 speakers in northeastern Ethiopia and in Djibouti (data are from Parker and Hayward (1986: esp. p. 225)). Nouns denoting male humans and the males of sexually differentiable animals are masculine; for example: Ḗq̲αq̲a la 'husband'. Females (human and animal) are feminine: barrī 'woman, wife'. Note that ' indicates the accent position, which marks potential high tone. These semantic assignment rules are unremarkable. The phonological rules are of greater interest:

*Phonological assignment rules*

(i) Nouns whose citation form ends in an accented vowel are feminine: catə 'help', karmə 'autumn'.

(ii) Others are masculine: There are two possibilities:

(a) those ending in a consonant: cedər 'supper time', gilləl 'winter';

(b) those with a citation form ending in a vowel but with non-final accent: tinu 'taste', bainta 'trumpet'.

These phonological rules have few exceptions; an example is dobnik 'sail-boat', which 'ought' to be masculine but is feminine.

When the two sets of rules are in conflict, the semantic rules take precedence (as in Russian). Thus abba 'father' is masculine because of its meaning, even though it ends in an accented vowel, normally an indicator of feminine gender. Conversely, gabitukēro 'slender-waisted female' is feminine, though the accent is non-final, which is a masculine pattern. Such phonological assignment systems are also found all over the world, as in the Kru language Gode and, in a much more complex form, in French.14

Before leaving assignment systems there are several general points to note. We have considered clear examples of the different types of assignment system. Sometimes we find complex overlapping of features: in these cases the gender of nouns is still largely predictable, but the rules and their interactions are more difficult to establish.15 A language like this is German, for which see Zubin and Köpcke (1986), and references there. Another complex and interesting case is Luvakaleve, a Papuan language spoken on the Solomon Islands (Terrill (1999)). A different type of complexity is that there are languages which at first sight appear to have phonological assignment systems, since gender can often be predicted from a single form, but which on closer examination turn out to have morphological systems. In these, the simplest analysis results when we see that phonology determines declensional class, and declensional class in turn determines gender. A good example of this type of system is found in the Papuan language Arapesh, for which see Aronoff (1976, 1999:89–121)

---

14 For which see Crockett (1997:7–61), and for work stressing the importance of the final syllable see Hardison (1997).

15 Interesting complications in Australian languages with body parts, which may have intrinsic gender or take the gender of the whole (or the strategies may be mixed), are described in Evans (1999).
Fraser and Corbett (1997) all following the description in Fortune (1942). It should be borne in mind that the assignment systems we have called formal systems still have a place for semantic rules; these provide the semantic core for (some of) the genders but have to be supplemented by formal type rules. There are no purely formal gender assignment systems: there is no language in which, say, all nouns beginning in a vowel are assigned to one gender, all those in a consonant belong in the other, with there being no correlation of the two genders to some semantic feature. Finally, though some still claim that gender is arbitrary, more and more of the difficult languages in this regard are being successfully analysed: French on three occasions, independently (Bidon 1925; Melčuk 1959; Tucker, Lambert, and Rigault 1977); German is becoming clearer, notably in the work of Zubin and Köpcke mentioned above; and more recently Swedish has proved to be substantially predictable (Källström 1996; Frantrud 2000).

3 Default genders

The notion of ‘default’ is current in linguistics, in large measure through the influence of Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, and Sag 1985). References specifically to ‘default gender’ can also be found in increasing numbers (for instance, in Hayward 1989). But this term has been used in various different ways, sometimes by a single author, so there is a danger that a potentially valuable notion could become debased. There are distinct uses, which we shall examine briefly (for more detail see Fraser and Corbett 1997), and especially Corbett and Fraser 2000). In particular, it is important to avoid the trap of assuming that a given language will have one default gender; different types of default sometimes coincide and sometimes do not. Naturally the notion of default is connected to markedness (Gazdar et al. 1985:29–31; Zwicky 1986:206–7). However, gender has proved problematic for traditional accounts of markedness. An early paper on the topic is that of Schane (1979) who discusses French and shows how the masculine may be considered unmarked according to different criteria. But since French has only two genders, this is not remarkable, and the patterns suggested tend to break down when more complex systems are considered. Greenberg too (1966:38–40) considers the question of the markedness of gender ‘less clear’ than with other categories. Both unmarked and default cases are in some sense ‘normal’ (though we shall need to revise this view for some instances of defaults discussed below): markedness theorists typically look for language-independent criteria.

Gender and noun classes
to establish unmarked values, while defaults are worked out on a language-
internal basis.

3.1 Types of default
An obvious area in which to begin a discussion of default genders is that of
gender assignment. The assignment rules given earlier can be thought of as
defaults; thus, by default, Russian nouns denoting males are masculine; also by
default, nouns of declensional class II are feminine. Note already that we must
allow for defaults to interact, since a noun can denote a male and be a member of
declesional class II (like дядя 'uncle', in which case it is masculine). If
we hanker after identifying one gender as the default, in the case of Russian
we would have to say that it is the masculine, for various reasons (for instance,
class I is the default declensional class, and masculine is the gender associated
by default with this class; masculine has the largest number of nouns, it attracts
the most borrowings). This may convince some, provided we remain with the
problem of gender assignment. Now consider this example:

(20) На вечеринке был интересно
    at party was-NEUT.SG interesting-NEUT.SG
    'it was interesting at the party'

Here there is no overt subject, but the verb and adjective must still take a
particular agreement form and they take not the masculine but the neuter. The
masculine may then be the default gender for nouns, but it is not the default
gender throughout the grammar of Russian. A higher-level default for gender is
needed for items, other than nominals, which may head syntactic constituents
with which gender agreement is required. The situation arises if, as in (20),
there is no overt subject, or if an infinitive phrase stands in subject position,
or if agreement is with an interjection or other quoted material. Intuitively the
two cases of default we have considered are somewhat different (see Fraser
and Corbett 1997); in the first type, the default accounts for the cases when
'everything goes right' (as with nouns denoting males being masculine). We
shall call instances of this type normal case defaults. In the second use of the
term, a default is something which applies when the normal system breaks down,
when 'something goes wrong' (as in the verb having to show agreement with a
non-prototypical noun phrase). We shall call instances of this type exceptional
case defaults.

Consider now the forms used. In the case of the verb it is straightforwardly
a neuter singular form. But the adjective is more complex. Russian adjectives
have two forms available for predicate use, the long form and the short form.
The short form is being lost in most uses; however, in examples like (20), the
short form is required. Thus it is not sufficient to say that these forms are neutrers.
For these and other reasons we need to distinguish neutral agreement forms, as required for agreement with non-protopypical controllers, from other agreement forms. In Russian, by default, these neutral forms are the same as the neuter.

A clear case of the neutral form having special properties is found in Romanian, where it varies according to the particular agreement target (data from Donka Parkus, see Corbett (1999:213-14)).

(21) *e evident că a venit, și asta o știe toată lumea*
    *is clear.MASC.SG that has come and this.FEM.SG il.FEM.SG
    *knows all the world*
    *It is clear that s/he came and everyone knows this*

Here we have a clause as subject (some might prefer to say there is no subject); the predicative adjective has to mark agreement, and is masculine (evidentă, the feminine form, is unacceptable). Asta ‘this’ can stand for ‘that s/he came’ or ‘it is clear that s/he came’. Either way it must be feminine (the masculine astă is unacceptable). Thus the form used for neutral agreement in Romanian varies according to the type of target involved. The next example includes attributive modifiers:

(22) *Un burmă puternică a fost auzită*
    *a loud boom was heard*
    *a.MASC.SG ‘boom’ strong.MASC.SG has been heard.MASC.SG*

Here an ‘a’ is masculine, like the agreeing predicate. Consider now the relative pronoun:

(23) *a admis că a venit, ceea ce nu e surprinzătoare*
    *he admitted that has come which.FEM.SG not is surprising.MASC.SG
    *s/he admitted that s/he came, which is not surprising*

Ceea ce is a complex relative, the first part of which shows feminine gender agreement. We thus have masculine agreement for attributive modifiers and the predicate, and feminine for the relative pronoun and the demonstrative (which replaces the personal pronoun when neutral agreement is required). However, though cea ce is feminine, its predicate surprinzător is masculine (*surprinză toate feminine*). This shows that cea ce is a remarkable neutral form: though morphologically feminine, it must carry a feature to distinguish it from ordinary feminines. The reason for this is the fact that its antecedent is a clause, and the evidence for the special feature is that cea ce controls masculine predicate agreement (as clauses do). Let us try the demonstrative in a similar environment:

reason: other speakers say

There is a problem with

ceea ce
Gender and noun classes (2a)

\[\text{asta e uliisor} \quad \text{this is amazing masc} \quad \text{‘this is amazing’}\]

Here *asta* refers to a situation not a specific object. While it is morphologically feminine, its takes a masculine predicate. Thus *asta fef* is a special neutral form, since it controls a different agreement from the *asta* which can stand for a noun of feminine gender.

3.2 Defaults in gender systems

Given that the notion of default appears valuable, we now sketch informally the areas of gender where this notion might be applied.

We have already considered assignment systems, and perhaps the most straightforward examples of defaults are found in gender assignment systems of the semantic type. A clear instance is found in Diyari, an Australian Aboriginal language which had about a dozen speakers at the last report, living near Lake Eyre in the north of the state of South Australia. One gender is for ‘all animals whose reference is distinctly female, for example, women, girls, bitches, doe kangaroos etc.;’ the other is for ‘all others, that is, male animals, non-female animals, non-sexed animals and all inanimates’ (Austin (1961:60)). By default in Diyari nouns are masculine. The converse system, in which nouns denoting males are singled out as masculine and all others are feminine, occurs in Kala Lagaw Ya, the language of the western Torres Strait Islands (Bardi (1987)). Here by default nouns are feminine. (Note, however, that the moon is also masculine, as is generally the case in the languages of Australia.) These are obvious cases of normal case defaults.

If we move on to gender agreement, we find three broad types of problem, all caused by agreement controllers other than straightforward noun phrases. The problems arise because if a particular target type can mark agreement in gender, then in many languages it must, whether or not there is a normal noun phrase to act as controller.

The first type of problem is that there are constructions in which the target has to agree in gender with a controller which is not specified for gender. The obvious examples here are those of the type we have already discussed, namely the ‘neutral’ agreement which results from agreement with non-prototypical controllers. The range of non-prototypical controllers varies from language to language. It may include clauses, infinitive phrases, nominalizations, interjections and other quoted phrases, noun phrases in particular cases (for example, subject noun phrases in an oblique case), dummy elements, and certain null elements.
Languages may solve the problem of agreement with non-prototypical controllers by pressing one of the regular gender/number forms into service. The form may be termed the 'neutral agreement form' as above or the 'default agreement form'. Though neutral agreement forms may appear to be identical to some other form, they are usually odd in some ways. Thus they typically appear identical to singular markers but they lack plural counterparts. Moreover certain target types may be avoided. And as we saw in Romanian, the form to be used can vary according to the target type. Some languages have unique neutral agreement forms (examples are Spanish, Portuguese, the Surselvan dialect of Romansh, Ukrainian, and the Selo Fara dialect of Slovene). However, no language has yet been found with a full set of unique neutral forms: regular gender/number forms are used for some targets. An interesting development occurs when neutral forms are used when the controller is an apparently straightforward noun phrase. This phenomenon is well attested in Scandinavian languages (Faarlund 1977); Hellan (1977:102-8); Erikkson (1979); Niksson (1978); for extensive discussion see Källström (1992:188-240) and Hedlund (1992:95-111). Our example is Norwegian, taken from Faarlund (1977).

(25) Ein ny utanriksminister ville ikkje vere så dum
    a new foreign secretary would not be so stupid
    a new foreign secretary would not be a bad idea

Norwegian predicative adjectives in the singular distinguish two genders, neuter and common. In (25) the neuter form is used, and the interpretation is that having a new foreign secretary would not be a bad idea. If the adjective were in the common form dem, then it would agree directly with the subject noun phrase and the interpretation would be less complimentary.

Even if the agreement controller is a noun phrase headed by a noun or pronoun, there may still be problems involving gender agreement, caused by reference difficulties. These are the second type of gender agreement problems to consider. These have mainly been investigated relative to human referents, though there can be similar problems (usually in larger gender systems) with non-humans. There are at least three sub-areas to investigate (it is not even clear whether all the different types have yet been identified): the gender required may be unknown, unclear, or mixed.

Suppose we ask Who said that? in a language which requires agreement in gender on the verb. We cannot determine the gender, since we cannot identify the referent of who; thus the gender required is unknown. Similarly, when we ask What was that? we may have theoretically possible referents of more than one gender. As a variant of this type we may have a noun, like English manager or friend, which can be used of a person of either sex. Again, in a specific instance, we may not know the sex of the referent. Second, there are cases where the gender required is unclear because the referent is non-specific.
Gender and noun classes

If a patient wishes to change doctors, he/she/he or she should advise the receptionist. A third area of difficulty here is agreement with a noun denoting a group of referents which would separately be referred to with nouns of different genders. The most obvious examples involve humans of both sexes (villagers, athletes). Here again the sex cannot be uniquely determined, but if the language distinguishes gender in the plural, then clearly one form must be selected for agreement purposes.

We shall see that there are two main approaches to dealing with these problems. First, one of the possible alternative agreement forms may be used by convention – an obvious type of default. If the ‘reasonable possibilities’ are genders A and B, then either A or B is chosen. The second possibility is for an ‘evasive’ form to be used. If the ‘reasonable possibilities’ are genders A and B, then gender C is chosen.

It is often assumed that in a single language, all problem types are dealt with in the same way (for example, it may be stated or implied that a particular gender is the unmarked or default one and so used in all these cases). But in fact languages may handle the three parts of the problem (gender unknown, unclear, or mixed) differently. This is an area where there has been a good deal of research on one small part of the topic but where much of the problem is only poorly understood.

Consider first the case where the appropriate gender is unknown. Suppose we have a language in which there is at least a masculine gender (containing nouns denoting males and other nouns) and a feminine gender (for females and other nouns). For the problem cases above, one set of target gender forms, say the masculine set, could be used by convention. This situation is found in many Indo-European languages. Let us take a Russian example:

(25) Kto eto sdelal?
    who this did,MASC
    ‘who did this?’

The speaker does not know the sex of the person responsible, but the masculine is used. Surprisingly, even in a setting in which the person must be one of a group of women, masculine agreement is still normal. Though the literature might suggest otherwise, it is not the case that the masculine is always used. In the Nilotic language Massal, we find the feminine used for questions when the person involved could be male or female (Tucker and Mpanye, 1955:27); for more on this interesting gender system, including the role of derivational morphology, see D. L. Payne (1958). If we now consider nouns which can be used in reference to a male or a female, then we find that in Russian, nouns like vrah ‘doctor’, take masculine agreements if the sex is not known. In Archi, however, we find an ‘evasive’ form. As we saw in section 2.2, Archi has four genders, 1 and 11 for humans, male and female, 11 and 111 less clearly defined.
semantically but with the larger animates in iii and most abstracts in iv. In 
Ari, nouns like lo ‘child’, adan ‘person’, cobra ‘thief’, mizgin ‘poor per-
son’ take gender iv agreements in the singular if the sex of the referent is 
unimportant or unknown (Kibrik (1972:285)). This shows a particularly clear 
example of an evasive form, since gender iv does not contain any nouns denoting 
humans.

The second type of reference problem, non-specific referents, has created 
a considerable literature, but generally with reference to a small number of 
languages. English can use the masculine here (Everyone loves his mother) 
or the plural can be used to evade the gender choice (Everyone loves their 
mother).

The third reference problem involves mixed groups. Usually mixed groups of 
humans are investigated, but in large gender systems there could be analogous 
problems with inanimates. Given, however, a mixed group of humans, in Serbo-
Croat (South Slavonic) we find the masculine plural oni ‘they’ in such cases. We 
may take the problem back into derivational morphology: Amerikanac (masu-
cline) is a male American, while Amerikanica (feminine) is a female American 
in Serbo-Croat. To refer to Americans in general, the plural of the masculine 
noun is used, that is Amerikanici, and it takes masculine plural agreements. This 
instance of the way in which gender is assigned to nouns denoting mixed groups 
links directly to the analysis of agreements used with conjoined noun phrases, 
which we consider in the next section. The ‘opposite’ system is found in the 
Khoisan language Dama, spoken in northern Namibia; here mixed groups of 
persons are referred to using the feminine pronouns (John Payne (p.c.)). In this 
problem area too, ‘evasive’ forms are an alternative strategy. Polish for instance 
uses the neuter singular. This usage is described by Goteri (1984), who took 
up the term evasive following a suggestion by Doroszewski. An example of the 
Polish neuter in evasive use is the following:

(27) Kôr-eś z małżonków jest winne zarzucane mu 
one-neut from spouses is guilty-neut imputed it
dat 
hzrodni 
crime 
‘one of the spouses is guilty of the crime he or she has been accused of’

Małżonkowie is masculine personal and means ‘husband and wife’; when either 
the husband or the wife is potentially the referent, then the evasive neuter is 
used. The neuter cannot be used in all the situations we have considered; in 
most the masculine is used (for examples, see (Herbert and Nykiel-Herbert 
(1986:87); see also Weiss (1993)). Most interestingly, the evasive neuter seems 
to be used in the sort of contexts which also preclude the use of generic 
he in English, that is where there are implied disjuncts, one of which is
Gender and noun classes

specifically female. Though a fuller analysis would be required in order to be certain, it appears that the cases that use one of the expected genders should be treated as normal case defaults, while the evasive forms are exceptional case defaults.

A final area, different to those described so far but related to the last, is that of gender resolution. It has been suggested that when agreement is required with conjoined noun phrases, then we may expect to find rules invoking the default gender. This, as the reader will probably guess, is an over-simplification; there is not ready sustainable prediction here. However, agreement with conjoined noun phrases, in those languages which have both natural conjoining of noun phrases and agreement in gender in non-singular numbers, is a fascinating problem which will be discussed in outline in the next section.

Before moving to that topic it is worth stating that defaults of different types may or may not line up together (and when there are only two possibilities, as with two-gender systems, then the coincidence cannot be assumed to be of any great significance). For instance, in the Russian data discussed earlier, at one level the default gender is neuter, and at another it is masculine. Then for nouns of particular declensional classes the default is feminine. Even in a two-gender system the defaults need not coincide. Recall that in Kala Lagaw Ya nouns are assigned by default to the feminine gender (only nouns denoting males are masculine). However, for a single human of unknown sex, the masculine is used (Alpher (1987:76)).

4 Gender resolution

This term is due to Givón (1979) and it refers to a rule which specifies the form of an agreeing element (or target) when the controller consists of conjoined noun phrases. Resolution is generally not obligatory; instead agreement may be with one conjunct only. Where this occurs, resolution is not involved and we shall not be concerned with that construction here. Gender resolution rules are of three different types: some languages have rules which are basically semantic, others rely on a syntactic (or ‘formal’) principle, while yet others show interesting combinations of the two.

4.1 Semantic gender resolution

Gender resolution by the semantic principle involves reference to the meaning of the conjoined elements, even if this implies ignoring their syntactic gender. Examples can be found in Bantu languages. These usually have several

---

17 A remarkable default is found in Jawa, a language of the Amor family (Dixon (1965:265)); here all pronouns take feminine agreement, irrespective of the sex of the referent(s).
genders, which correspond to semantic classifications only partially: nouns of the 1/2 gender typically denote humans, but not all nouns denoting humans belong to the 1/2 gender (Bantuists use labels such as 1/2 to indicate the agreements taken for singular and plural – a clear way of specifying the agreement class). For gender resolution, the important thing is whether a noun denotes a human or not, irrespective of its gender. This point is illustrated in data from Luganda (Givón (1970:253-4; 1977:38-9)). The resolved form for conjoined noun phrases headed by nouns denoting humans is the class 2 marker – the one used for agreement with plural nouns of the 1/2 gender. In (28) none of the conjuncts belongs to the 1/2 gender, but as all denote humans the resolved form is the class 2 marker:

(28)  
a-ka-ka na olo-ka-ja ba-ka-labwa  
5-fat woman 12-small child and 11-tall man 2-were seen  
'the fat woman, the small child and the tall man were seen'  

Clearly the use of the class 2 form as the resolved form is motivated by semantic considerations. If none of the conjuncts denotes a human, then the class 8 form is used, as in (29):

(29)  
e-ka-ta, olo-ka-ju, aki-ke na sku-ku bi-ka-labwa  
9-cow 3-wild cat 7-jackal and 5-canoe 8-were seen  
'the cow, the wild cat, the jackal and the canoe were seen'  

Conjoining nouns denoting a human and a non-human produces an unnatural result; the preferred alternative is the conative construction:

(30)  
o-ka-ja y-agwa ne en-bwa-ye  
1-man 1-fell with 9-dog-his  
'the man fell down with his dog'  

A similar situation obtains in several other Bantu languages, but there may be complications (see, for example, the analysis of Chichewa by Corbett and Mtenje (1987)).

4.2 Syntactic gender resolution

Gender resolution according to the syntactic principle means that the gender of the nouns involved is what counts, rather than their meaning. French has two genders, and if conjoined noun phrases are headed by nouns of the same gender then that gender will be used (examples from Grevisse (1965:806-7); further examples in Heybye (1944:232-17)).

18 Gender 7/8 is arguably the gender which includes nouns of the widest semantic range, hence the use of the class 8 marker is understandable.
Gender and noun classes

(31) un livre et un cahier
    a book.masc and an exercise.book.masc new MASC.PL
    'a new book and exercise book'

(32) la misère et la ruine général-es
    the poverty.fem and the ruin.fem general.fem.PL
    'the general poverty and ruin'

In many instances the marking of gender is purely orthographic in French; examples (31) and (32) have been chosen because the particular adjectives have phonetically distinct forms. They also illustrate that resolution can operate for various target types and not just for predicates.

When the conjuncts are headed by masculine and feminine nouns, then a masculine form is used:

(33) un père et une mère excellent-s
    a father.masc and a mother.fem excellent-masc.PL
    'an excellent father and mother'

(34) un caractère et une énergie particulier-s
    a nature.masc and an energy.fem special-masc.PL
    'a special nature and energy'

Here the rules apply with the same effect to animate (33) and inanimate (34) nouns.19 though the relative frequency with which they apply is likely to differ.

The rules are similar of the syntactic type:
(i) if all the conjuncts are feminine (syntactically), the feminine form is used;
(ii) otherwise the masculine is used.

Languages with similar resolution rules are common; they include Italian, Spanish, Slovene, Latvian, Hindi, Panjabi, and modern Hebrew (for sources and further examples, see Corbett: (1991) 261–306) and (2003)),20

19 Alain Chissool points out that, though the attested examples (33) and (34) are acceptable, they are awkward and likely to be avoided. The problem is that the plural does not have any phonetic effect in these two adjectives and so there is a feminine noun immediately followed by an adjective which is clearly marked as masculine but not as plural. In careful style the conjuncts would be reordered to place the masculine conjunct adjacent to the masculine adjective. This awkwardness is a side effect of French phonology and does not undermine the validity of the resolution rule. The problems are avoided if we look at resolution in the predicate, rather than in attributive position:
(i) son caractère et son énergie sont particulier-s
    his nature.m and his energy.f be.pl special.m.PL
    'his nature and his energy are special!'

Here the verb signals plurality clearly and it separates the masculine adjective from the feminine conjunct; there is no problem with examples like (i).

20 Early Germanic had an interesting set of rules, preserved in Old High German and Middle High German, and indeed in Modern Icelandic, according to which if all conjuncts were masculine, then masculine agreement was used, if all feminine, then feminine, and in all other cases neuter. Thus the neuter covered the instances of mixed gender. For the historical data, see Askold (1973).
4.3 Mixed semantic and syntactic gender resolution

The semantic and the syntactic principles of gender resolution coexist in Latin. When resolution occurs in Latin, conjuncts of the same syntactic gender take agreeing forms of that gender. Thus if all conjuncts are masculine, then masculine; if all feminine, then feminine; and if all neuter, then neuter. This is resolution by the syntactic principle. However, when conjuncts are of different genders, then the resolved form to be used depends on whether the nouns denote persons or not. For persons, the masculine is used (examples are from Kühner and Stegmann (1955:44-52)):

(35) quam pridem pater mihi et mater mortu-i essent
dead-MASC.PL were
how long ago father.MASC DATIVE and mother.FEM
was "how long ago my father and mother had died"

For other conjoined elements (that is, when the head nouns are of different genders and do not all denote humans), when gender resolution operates, the neuter is used:

(36) murus et porta de caelo tacta erant
wall.MASC and gate.FEM from sky were
stuck-NEUT.PL were
"the wall and the gate have been struck by lightning"

Here we have resolution according to a semantic principle. Thus Latin shows both semantic and syntactic principles at work.

4.4 The relation between resolution and assignment

The type of gender resolution system found in a particular language is not random. In Corbett (2000) it is claimed that the type of gender resolution system depends in part on the assignment system. The evidence is as follows:

Languages with strict semantic assignment systems (like Godoboi) have semantic resolution systems, as do those with predominantly semantic systems (like Archi).
Languages with formal assignment systems may have semantic resolution (Luganda), mixed (Latin) or syntactic (= formal) resolution (as in the case of French).

Thus: ‘gender resolution may not be determined by semantic considerations to a lesser degree than is gender assignment’. This claim is based on yet on limited data, since the number of languages for which the assignment system has been investigated is rather small. There is fruitful research to be done in this area.

This basic typology covers the ‘standard’ examples, but there are interesting complications if a conjunct is headed by a hybrid noun. Such examples show that semantic resolution is required, even for the languages discussed in section 4.2 (Wechsler and Zlatić (2003: 171-95)). This means that resolution rules follow closely the assignment rules of a language (see Corbett (2006:259-63) for discussion).
5 Prospects

Gender continues to be a live topic in linguistics. While the literature is considerable, there are many languages whose gender system is not adequately described; in some cases, there is little time left, since languages are disappearing fast. It is particularly urgent to document complex systems like those discussed in section 1.2. Even for apparently well-studied languages, the accounts are often only partial.

Ideally a description should include the types of agreement in gender, that is, the evidence which demonstrates the presence of a gender system. Then we would expect an account of the number of genders and of any problem cases (indefinite genders, hybrid nouns, and so on). Once given the number of genders, there should be a set of rules for assigning nouns to these genders. If gender resolution occurs, then there should be an account of the rules.

Assignment rules can be verified by investigating the assignment of invented words, by observing gender assignment by children, and by studying the allocation of borrowings to gender. In the last case, it is safest to select nouns borrowed within a specified period. For certain languages, dictionaries of new words are available, which contain only words which appeared after a certain date (though not all will necessarily be borrowings). Alternatively, both period and subject area can be limited by examining, for example, all loans relating to a specific technology or activity.

As we gain more detailed and complete descriptions of the gender systems of a wider range of languages, so we can propose more restrictive typologies. We could hope to establish the possible semantic features on which assignment can be based, the possible ways in which factors may overlap in assignment, the possible alignments of default genders, and the possible relations of resolution to assignment and to defaults.

As more gender systems are adequately described, we can also expect progress in the study of how children acquire gender systems. There is already interesting work in this area (see, for example, the work of Mills (1980) and Müller (1999, 1994, 2000)) but many studies are vitiated by an inadequate understanding of the system which the child is learning. The way in which children acquire gender systems can help us to understand better how such systems change over time (as shown by the work of Polinsky and Jackson (1998); for development of the work on modelling change in assignment systems, see Polinsky and Everbrooke (2005)).

Our account here has focussed on what gender systems are; this is a prerequisite for investigating the function of gender. There is great potential for the investigation of authentic spoken language material to establish what gender actually does, whether by itself or together with other linguistic subsystems.
There are two studies which demonstrate that gender has a major role in the languages described. The first concerns the Australian language Nunggubuyu. In terms of its syntactic structure, Nunggubuyu appears remarkably simple: subject and object are usually not differentiated and there is almost no cross-clause relational syntax. Here the gender system 'appears to constitute the glue which holds the system together' (Heath (1983:139)). A text provided shows how the verb, by indexing the different participants according to the seven genders, allows the language to function without many of the syntactic devices which are sometimes thought to be essential. In a second study, Foley and Van Valin give an account of the Papuan language Yimas and claim that the gender system 'carries most of the load of referential tracking' (1983:527). Thus gender may have a central role: in some languages, and there are many more similar to the two just noted, the reference-tracking function depends largely on gender; in others this function is shared with other devices, and in some languages gender has no place.

Besides this major function, gender may have other secondary functions in showing the attitude of the speaker. It may be used to mark status, to show respect or a lack of it, and to display affection. The use of a particular gender may be fixed for this purpose, or it may be available for 'switching' in particular circumstances according to the speaker's attitude. As an instance of its use to show status, in some Polish dialects the feminine gender is used only for women who are married. The neuter (or masculine according to dialect) is used of unmarried women (Zareba (1984:5)). Affection is shown in baby talk in Arabic by shifting gender (masculine for a girl, feminine for a boy), according to Ferguson (1964:196). In Grebo the use of non-human agreements for humans is insulting (Innes (1966:53)); for further examples, see Head (1978:125-7).

A related question is what gender does is where it comes from. In the last century there was a good deal of speculation on this topic, but more recently data have become available which give a more plausible picture. Gender systems can arise from classifiers, hence ultimately from nouns. Greenberg (1963) suggested that this can be observed in the Daly languages of Australia; Reid (1991) shows this clearly. Again it would be of great interest to have accounts of other languages in which the gender system is at the very early stages of development. Rather more is known about later stages in the development of gender systems. These changes may be complex and subtle, as...

---

22 This is not the only route, however. For instance, Mosel and Sprigg (2001) suggest that gender has arisen in Troop (Sehati North Bougainville network, Northwest Solomonic Group, Melanesian Cluster, Oceanic, Austronesian) by three distinct spatial demonstratives coming together into a system of gender-distinguishing articles.
Gender and noun classes

shown, for instance, by Klein-Andreu (1996) in an analysis of various Spanish dialects. Gender offers exciting research prospects for linguists of various types; the inherently puzzling nature of the phenomenon has been recently stressed by Aronoff (1998). There are also many fascinating opportunities for collaborative work. Anthropologists and ethnographers have already contributed to our understanding of assignment systems, notably those which are primarily semantic but where the semantic criteria are not fully clear. Joint work on such languages is still possible, though time is running out. Assignment systems offer scope also for collaboration with psycholinguists and psychologists. For many languages, especially those with formal systems, we can now describe the assignment of nouns quite accurately (and as noted in section 2.3, we can use computational techniques to check that our analyses make the correct predictions). There is then the question of how this information is represented in the brain, as part of the goal of understanding how the internal lexicon is structured. 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 might have been expected. Provided we examine a wide range of languages, we discover that it is not at all straightforward to establish links between grammatical gender and the relative status and treatment of those classified by the different genders (notably men and women, though the other classifications also deserve study). This is an area where cross-linguistic work must be combined with cross-cultural research. A fourth type of collaboration is with computational linguists. We have seen how gender can provide a means for reference tracking in a language, yet it may also be absent. If this is so, then the strategies for parsing must reflect this difference. And we should be able to implement parsers to demonstrate how the different strategies work in different languages. There is much still to be learned about gender.

6 Suggestions for further reading

An account of classification more generally, including a substantial chapter on gender systems, is found in Aikhenvald (2000). Corbett (1991) contains a typology of gender systems, with reference to over 200 languages. Craig (1986) is a set of papers on noun categorization, including gender as one means of categorization. An essay which helps to distinguish gender from other means of classification is found in Dixon (1982: 167–233).

23 Another area of interest to psycholinguists, the interpretation of pronouns, is analysed in Garasch, Oakhill, Ehrlich, and Careiras (1999), while van Berken (1996) looks at the place of gender in word recognition and in speech production.
Evans *et al.* (2002) provides a detailed account of a complex system, with an implementation to validate the analysis. A set of papers on languages which had previously been under-represented in work on gender systems is found in Harvey and Reid (1997). Senft (2000) is a collection devoted mainly to other systems of classification, but it includes work on gender (pp. 293–325). Unterbeck, Rissanen, Nevalainen *et al.* (2000) is another varied collection.