Trends in Linguistics
Studies and Monographs 124

Gender in Grammar and Cognition
I Approaches to Gender
II Manifestations of Gender

Barbara Unterbeck
Matti Rissanen (Editors)

Offprint

MOUTON
DE GRUYTER
Default genders

Greville G. Corbett and Norman M. Fraser

1. Introduction

The notion of default has been invoked on several occasions in the linguistic literature with reference to gender. It has been used in varying ways, sometimes quite loosely, with the result that a potentially valuable notion may become debased.

In this paper we introduce the notion of default (section 2) and move on to discuss default inheritance (section 3); then we proceed to look at a single language, Russian, and show how different defaults for gender can be identified, depending on the level examined (section 4). We shall see that the defaults are of different types, which we investigate further in section 5. Following this we consider gender defaults informally in a range of languages (section 6), and show how different types of default may or may not coincide in different languages (section 7). In the course of the paper we take a good deal from work done within the framework of Network Morphology (Corbett and Fraser 1993).

2. The notion of default

References to defaults in the context of gender are fairly common in the literature. For instance, Corbett (1979: 16) mentioned defaults when discussing Russian agreement in number and gender, and this usage was followed by Dziwirek (1990) in her account of Polish. Hayward’s (1989) paper on East Omo to languages has "default gender" in the title, and Gazdar (1992: 51) writes:
... the lexicon is ridded with interdependencies. It is not uncommon, for example, for the default syntactic gender of a lexeme to be a complex function of the sex of the referent and the phonology of the final syllable.

Hedlund (1992, see especially pp. 95-111) and Doleschal (1993: 35-36, 52) both use the term "default". A careful reading of these various sources reveals that the term is used with reference to gender (including gender agreement) in different senses, sometimes by the same author.

Straightforward defaults are not problematic; the difficulties arise when defaults are interdependent, that is, when a default for a particular feature depends on the value of some other feature. Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum and Sag 1985) tackled this problem explicitly and in considerable detail. Zwicky (1986: 307) refers to the "enormous success of GPSG" in promoting the default approach to syntax, and considers the implications in morphology as well as syntax in a later paper (Zwicky 1989).

Naturally the notion of default is connected to markedness (Gazdar, Klein, Pullum, and Sag 1985: 29-31, Zwicky 1986: 306-307). However, gender has proved problematic for traditional markedness accounts. An early paper on the topic is that of Schane (1970), 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 tend to break down when more complex systems are considered (see Corbett 1991: 290-291 for further discussion). And Greenberg (1966: 38-40) considers the question of the markedness of gender "less clear" than with other items. 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 to establish unmarked values, while defaults are worked out on a language-internal basis.

Given this, tackling gender in a framework with interacting defaults looks promising. But as Gazdar himself points out (1987: 43), the solutions within GPSG to the problems of combinations of features were difficult. He went on to import from Artificial Intelligence the notion of default inheritance, and together with Evans, built the lexical-knowledge representation language DATR around this notion (Evans and Gazdar 1989a; b, 1995).

3. Default inheritance

The basic idea of default inheritance is very simple: for any entity which can be analysed as an instance of some general type or class, it is sufficient to encode only exceptional features in respect of that entity, since the regular features can be inferred by virtue of its ontological type. This idea will be familiar to anyone used to devices such as lexical redundancy rules which are used to "fill in" predictable information in lexical entries in accounts which view the lexicon as a repository for exceptional information (Jackendoff 1975). However, default inheritance supports a richer, more structured understanding of defaults than traditional lexical redundancy rules, since the types which entities instantiate can themselves be said to instantiate more general types, and so on.

Any taxonomic classification can be viewed as a default inheritance system. For example, consider the taxonomic hierarchy in Figure 1. The lines in this taxonomy indicate instantiation. So, eagles, robins and penguins are birds; Edwina is an eagle, Rupert is a robin, and Percy is a penguin.

![Figure 1. A simple default inheritance structure](image-url)

Given an instantiation hierarchy of this kind, default inheritance allows all attributes of a given node in the hierarchy (such as BIRD) to be inherited by a node which instantiates it (such as EAGLE), except in cases where the lower node already has a value for the attribute in question and thus overrides the default value. In our example, a BIRD has feathers and can fly. These facts are inherited
by EAGLE and ROBIN and, indirectly, by Edwina and Rupert. The attribute of having feathers is also inherited by PENGUIN and thus by Percy. However, specific information about the flying abilities of PENGUINs blocks inheritance of the more general information about BIRDS. Thus, despite the fact that Percy is a BIRD, he cannot fly.

Default inheritance allows generalizations to be expressed once, at a high level, and then automatically to apply to everything which inherits from there. Using this approach, it is simple to encode information, be it regular, subregular or completely exceptional. It has the added advantage of indicating exceptions explicitly as such, as in the case of PENGUIN in Figure 1. Penguins are exceptional by virtue of being flightless, while they are normal (or regular) in other ways, such as laying eggs. Appropriately, it is their flightlessness which is specified in relation to the PENGUIN node.

The relevance of such notions to linguistics should now be obvious; we are used to exceptions, but we do not normally find total exceptions. Typically exceptions are a matter of degree, from those which have a single unusual feature to those which have several such features.

It is now widely acknowledged that the ability to separate substance from form in linguistic theories offers at least two benefits: first, it focuses attention on the substantive content of theories rather than on their notation; second, it facilitates cross-theoretical comparisons (Shieber 1987). Our approach to morphology, which we call Network Morphology, takes default inheritance as a central notion (Corbett and Fraser 1993, Brown and Hippiusly 1994, Fraser and Corbett 1995). It could, in principle, be formalized in a number of different notations. For this purpose, we have chosen to use the DATR knowledge-representation language 3, which was developed by Evans and Gazdar as a declarative formalism for describing inheritance networks (Evans and Gazdar 1989a,b).

The information in Figure 1 would be expressed in DATR as follows:

(1) BIRD:

<has feathers> = yes
<can fly> = yes.

EAGLE:

<> = BIRD.

ROBIN:

<> = BIRD.

PENGUIN:

<> = BIRD
<can fly> = no.

Edwina:

<> = EAGLE.

Rupert:

<> = ROBIN.

Percy:

<> = PENGUIN.

The labels preceding colons are "nodes"; the angle bracket expressions to the left of the "=" symbol are "paths"; the words to the right of non-empty paths are "values". Thus, the value of the <can fly> path at the PENGUIN node is 'no'.

It should be immediately apparent how this kind of formalism might be useful in expressing linguistic generalizations. Consider the data on Russian inflectional paradigms shown in Table 1:

<table>
<thead>
<tr>
<th></th>
<th>zakon ('law')</th>
<th>komnata ('room')</th>
<th>kost ('bone')</th>
<th>vin ('wine')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINGULAR</td>
<td>nom (nominative)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
</tr>
<tr>
<td>acc (additive)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>gen (possessive)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>dat (dative)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>inst (instrumental)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>loc (locative)</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>PLURAL</td>
<td>nom</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
</tr>
<tr>
<td>acc</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>gen</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>dat</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>inst</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
<tr>
<td>loc</td>
<td>zakon</td>
<td>komnata</td>
<td>kost</td>
<td>vin</td>
</tr>
</tbody>
</table>

Notes: (i) Forms are given in phonemic transcription. (ii) Polurlatization (or "softening") is indicated by . (iii) There is no overt ending in the nominative/accusative singular in classes I and III, nor in the genitive plural of classes II and IV.
Figure 2 shows the inheritance structure for Russian nominals which we presented in our earlier work (Corbett and Fraser 1993: 126).

\[ N_{\text{I}}, N_{\text{II}}, N_{\text{III}}, \text{ and } N_{\text{IV}} \] are declensional classes. (Notice that we posited the existence of a node \( N_{\text{O}} \) from which \( N_{\text{I}} \) and \( N_{\text{IV}} \) both inherit, thereby finding an harmonious resolution for the long-standing debate over whether Russian has three or four noun-declensional classes.) The following (incomplete) fragment is taken from our earlier Network Morphology analysis.

\[ \text{Figure 2. A default inheritance structure for Russian nominals} \]

\[ (2) \]

\[ \text{NOUN:} \]
\[ <\text{mor loc sg}> \text{=} "\text{stem}" \text{ } _e \]
\[ <\text{mor nom pl}> \text{=} "\text{stem}" \text{ } _i. \]

\[ N_{\text{III}}: \]
\[ <> \text{=} \text{NOUN} \]
\[ <\text{mor loc sg}> \text{=} "<\text{mor dat sg}>". \]

\[ \text{Kost'}: \]
\[ <> \text{=} N_{\text{III}} \]
\[ <\text{stem}> \text{=} \text{kost'}. \]

The first sentence at \text{NOUN} should be read as saying that the locative singular consists of the stem followed by an \text{-}\text{e} ending. A path enclosed in double quotes in a DATR sentence is used to retrieve the specified value (in this case, \text{<stem>}) for the item (in this case, a lexical item) inheriting from it. If we wanted to find the nominative plural of \text{kost'}, we would inherit the sentence \text{<mor nom pl>} \text{=} "<stem>" \text{ } _i. \] Before going any further we would have to find out what the \text{<stem>} of \text{kost'} is. Since the answer is stated in (2) to be \text{kost'}, the nominative plural must be \text{kost'\text{i}}. If we wanted to know the locative singular of \text{kost'}, we would never inherit the definition of locative singular at the \text{NOUN} node because it is overridden at \( N_{\text{III}} \), from which \text{kost'} inherits first, thereby blocking the more general ending. The definition of locative singular at \( N_{\text{III}} \) establishes an asymmetric identity between the locative singular form of an \( N_{\text{III}} \) noun and its dative singular \text{<mor loc sg>} \text{=} "<mor dat sg>".

It is worth noting some salient features of this approach. First it is declarative; we do not give underlying forms, and then transform them into other forms. We make a set of (partly conflicting) statements, and by embedding them in a network we specify the relations which hold between them. Since the approach is declarative we do not have ordering of rules. Second, defaults may be seen as hierarchically related: we make default statements about nominals, about nouns, about the \( N_{\text{O}} \) class, and so on. Third, since computer interpreters are available for the DATR language, it is possible to check that an inheritance network expressed in DATR captures the intended generalizations. All our Network Morphology analyses have been checked in this fashion.

More detailed introductions to default inheritance can be found in Gazdar (1987), and Daellemans, De Smedt and Gazdar (1992). Briscoe, de Paiva and Copestake (1993) and Calder (1994) are good places to look for recent formally explicit research in linguistic defaults. Introductions to DATR can be found in Gazdar (1990; forthcoming).

4. Defaults in a single language

We now return to our central topic, namely gender. Russian has three genders, masculine, feminine, and neuter. We could enter a value for gender in the lexical entry of each noun. However, this would miss generalizations at two levels: first, the gender of Russian nouns does not appear to be random, and second, more generally, we have argued that languages never have to specify gender for the majority of
nouns. We have claimed that the gender of the overwhelming majority of nouns can always be predicted, either from semantic information which must, in any case, be stored in the lexical entry, or from semantic information supplemented by formal information, which may be morphological or phonological (Corbett 1991: 68).

4.1. Defaults for gender assignment

In Russian, the gender assignment rules refer to semantic and to morphological information:

Semantic assignment rules for gender
1. Sex-differentiable nouns denoting males (humans and higher animals) are masculine: for example, student 'male student';
2. Sex-differentiable nouns denoting females are feminine: for example, učitel' nica 'teacher'

Nouns which are sex-differentiable are those denoting beings whose sex matters to humans (that is, other humans and domesticated animals) and where the difference is striking (as in the case of lions).

Very few exceptions are left by these rules, but there are many nouns which are not covered by them. Those remaining are distributed over the three genders and their distribution is accounted for by the morphological assignment rules. The major morphological assignment rules are as follows:

Morphological assignment rules for gender
1. nouns of declensional class I are masculine;
2. nouns of declensional classes II and III are feminine;
3. nouns of declensional class IV are neuter.

As we shall consider in more detail later, there is some overlap between the two sets of rules: nouns which denote males are often in declensional class I. We might try to do away with the semantic assignment rules. However, the two sets of rules can make conflicting assignments, and when they do it is the semantic rules which dominate. The crucial case is muzhchina 'man' (and similar nouns), which ought to be masculine according to its semantics (male), but feminine according to its morphology (declensional class II). In fact, it belongs in the masculine gender.  

There are also nouns which do not decline, and to which the above morphological assignment rules do not apply. However, their lack of declension is itself a matter of morphology, and this interacts with the semantic feature of animacy to allow gender assignment. We treat indeclinable nouns as having their own declensional class (V); nouns of this class may be subject to the normal semantic assignment rules. Failing this, they are masculine if animate and neuter if not.

In our Network Morphology account (the detail of which can be found in Fraser and Corbett 1995), every noun inherits from the NOUN node. Since the following path equation is specified at NOUN, every noun inherits it unless it is overridden by a more specific equation declared in the lexical entry:

(3) NOUN:
   <syn gender> == GENDER:< "<sem sex>" >
   ...

In order to find a value for the path <syn gender>, the path <sem sex> is evaluated (i.e., the sex of the noun's denotatum is retrieved) and then a path consisting of only the sex is evaluated at the node GENDER (the node below):

(4) GENDER:
   <male> == masc
   <female> == fem
   <undifferentiated> == "<mor formal_gender>".

The interpretation of this fragment of DATR is straightforward in the case of sex-differentiated nouns. If the denotatum of the noun is male then the gender is masculine, and if female then the gender is feminine. In these two instances there is a simple default, which holds of nouns in general. However, if the sex is undifferentiated, it is necessary to consider additional criteria concerning form, and this is done by evaluating the path <mor formal_gender>. Formal gender is defined for each of the five main declensional classes. Thus, for example, the following equation is located at the node for declensional class II nouns, N_II:

(5) N_II:
   <formal_gender> == fem
This is used in the evaluation of the path <mor formal_gender> (the function of the initial mor will be spelt out below). This is a lower level default, which holds just for class II nouns. The effect of these defaults is that a class II noun denoting a male (such as mužčina 'man') will have masculine gender; a class II noun denoting a female (such as učitel' nica 'female teacher') will have feminine gender; and a class II noun denoting a non-sex-differentiated denotatum (such as komnata 'room') will have feminine gender, the default gender for class II nouns.

As we noted above, there is an additional complication with declensional class V, where animacy must also be taken into consideration. Some equations from the node N_V, from which all class V nouns inherit, are given below.

(6) \[ N_V; \\
<\text{formal_gender}> = <\text{sem animacy}> \] \\
<\text{animate}> = \text{masc} \\
<\text{inanimate}> = \text{naut} \\
...

In order to assign a formal gender for a given class V noun it is necessary to evaluate the <sem animacy> path for that noun and use it to select masculine gender if the noun denotes an animate and neuter gender if it denotes an inanimate.

Our analysis so far accounts for the gender of the vast majority of Russian nouns. It is an advance on Corbett (1982) in that it is formally explicit, and since it is encoded in DATR we can see that the right predictions are indeed made.

4.2. Defaults for inflectional class

To illustrate our theme of defaults we will look briefly at the question of declensional class. So far we have assumed that the declensional class of each noun will be specified in its lexical entry. This seems to be missing a generalization, since there are common, though by no means exceptionless, correspondences between meaning, gender, and declensional class. Indeed, some earlier analyses attempt to predict declensional class from other information which was specified (for discussion see Corbett 1982, Fraser and Corbett 1995). For many nouns, it is the case that declensional class is predictable from semantic or formal information. The semantic correspondences are as follows:

**Semantic assignment rules for declensional class**

1. Sex-differentiable nouns denoting males (humans and higher animals) are of declensional class I: for example, student 'male student';
2. Sex-differentiable nouns denoting females are of declensional class II: for example, učitel' nica 'female teacher'.

There are substantial numbers of nouns whose declensional class must be specified to override rule one. These are nouns like mužčina 'man', which denote males but which decline according to declensional class II. (Lazova 1974: 942–943 puts the figure at 273, but the number involved is actually larger because there are many hyponyms of this type, like Sača 'Sasha' which do not appear in dictionaries like Lazova's.) There are fewer instances of overrides to rule 2, but we find a small number of nouns like svekrov 'mother-in-law' in declensional class III.

**Formal assignment rule for declensional class**

1. Nouns whose stem ends in a vowel are of declensional class V.

The effect of this rule is to make nouns whose stem ends in a vowel indeclinable (class V are the indeclinables). Our rule follows Worth (1966), though as he points out the idea goes back much further. A noun like taks 'taxi' is entered as such in the lexicon and this guarantees its indeclinability.

These generalizations can be formalized fairly easily. Consider the following DATR fragment, which is positioned at the NOUN node:

(7) NOUN:
<mor> = "<declensional_class>"  
<declensional_class> = DECLENSTION;  
<"<infl_root_final>" "<sem sex>" >  
...

The first equation indicates that in order to find a value (or values) for the <mor> path (and its extensions), it is necessary to evaluate the <declensional_class> path. The second equation is somewhat more complex: a value can be retrieved for the <declensional_class> path by evaluating a path consisting of the value of
the path <infl_root final> followed by the value of the path <sem sex>. Paths beginning <infl_root> make available information about the inflectional root (i.e., the stem) of a word. The <infl_root final> path is used to store information concerning the final segment of the inflectional root, specifically whether it is a consonant or a vowel. We may assume that this information would, in reality, be supplied by a phonological component, though for convenience we simulate this crudely by means of the following equation recorded at the NOUN node:  

\[ (8) \quad \text{NOUN:} \quad <\text{infl_root final}> = \text{consonant} \]

This generalization holds for all Russian nouns, except certain indeclinable nouns of relatively recent foreign origin. In the case of these exceptions the default generalization is overridden in the lexical entries.

We have already noted how values for the <sem sex> path are supplied either in the lexical entries or by means of a default assignment of undifferentiated sex. Thus, where declensional class is predictable, it is found by evaluating a path at the DECLENSION node. This path consists of information on the final segment of the inflectional root of the word followed by the sex of the word's denotatum. The DECLENSION node is given below:

\[ (9) \quad \text{DECLENSION:} \]
\[ <\text{consonant male}> = \text{N_I:<>} \]
\[ <\text{consonant female}> = \text{N_II:<>} \]
\[ <\text{vowel $sex$}> = \text{N_V:<>}. \]

The first two paths encode the semantic assignment rules for declensional class. The first element of each path is consonant so these paths potentially apply to the bulk of Russian nouns. The second elements of the paths narrow down on two subsets of these. The first path assigns core nouns (typical native nouns) denoting males to class N_I; the second assigns core nouns denoting females to class N_II. The third path picks out those nouns whose inflectional root ends with a vowel and assigns them to class N_V, the class for indeclinables. In DATR, every symbol which begins with a dollar sign ($) is a variable. The variable $sex$ is defined to range over all possible values for the <sem sex> path, namely male, female and undifferentiated. One group of nouns is not covered by the equations at this node, namely those whose stem ends in a consonant and which have non-sex-differentiated denotata. Such nouns must typically specify declensional class in their lexical entries. However, as we shall see shortly (in the discussion leading to (10)), a substantial proportion of them can be assigned to a declensional class by the setting of a default.

It is worth contrasting declensional class assignment with gender assignment. A clear difference is the outcome when semantic and formal rules make different predictions. For gender assignment, quite generally, it is the semantic rule which takes precedence. As we noted earlier, with nouns like *mužčina* 'man', which ought to be masculine according to its semantics (male), but feminine according to its morphology (declensional class II), it is the semantic rule which "wins" and the noun is masculine. When we find a similar clash in declensional class assignment, in Russian at least, the formal factor dominates. Thus where *ataše* 'attaché' denotes a male and so "should" be in declensional class I, it nevertheless ends in a vowel and so "should" be in declensional class V. It is actually in declensional class V. Conversely, as far as gender is concerned, where it denotes a male and so "should" be masculine, it is nevertheless in class V and so "should" be neuter; as already noted, semantic factors take precedence in gender assignment and so it is masculine.

A second difference between gender assignment and declensional-class assignment, is that the gender assignment rules make a prediction for every noun (which has to be overridden in an extremely small number of instances), while the declensional-class assignment rules leave large numbers of nouns with no prediction. A non-sex-differentiable noun with a stem ending in a soft consonant could decline according to class I, II, III or IV. However, assignment to the four classes is by no means equally likely. Consider data on the numbers of nouns in each class, to the nearest fifty (derived from Lazova 1974, especially pp. 942-943), given in Table 2 (next page).

These raw statistics give a rough picture of what is going on. The figures do not take account of the effect of derivational morphology; for instance, over 4,300 of the nouns in class III have the suffix -ost' which forms abstract nouns from adjectives (*star'-jy 'old', *star'-ost' 'old age'). If the suffix is labelled as belonging to class III, then the number of distinct members of the class is substantially reduced. Similarly nominalizations in -anie/-enie (like razrušenie 'destruc-
tion', derived from razrušit' 'destroy') inflate the figure for IV (see Schupbach 1984 for discussion).

Table 2. Number of nouns in the different declensional classes

<table>
<thead>
<tr>
<th>Class</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>zakon</td>
<td>komnata</td>
<td>kost'</td>
<td>vino</td>
<td>taksi</td>
</tr>
<tr>
<td></td>
<td>'law'</td>
<td>'room'</td>
<td>'bone'</td>
<td>'wine'</td>
<td>'taxi'</td>
</tr>
<tr>
<td>No of nouns</td>
<td>20,850</td>
<td>16,050</td>
<td>5,150</td>
<td>11,050</td>
<td>450</td>
</tr>
</tbody>
</table>

Table 2 shows that I and II have substantially more members than the other three. We can therefore set a default which will put nouns in the major class I. Clearly this will be overridden in many instances, but equally it will simplify a very substantial number of lexical entries.

This is a considerable simplification, and one which can be justified in three interrelated ways. First, there is the numerical preponderance of class I. Second, it is the declensional class which takes the majority of borrowings. There is no need to appeal to any markedness considerations to explain this, it is simply that class I has no ending in the citation form, the nominative singular, so that foreign words ending in a consonant are typically borrowed into this class. The third argument is most important for our purposes. If the declensional class is I, and masculine is the gender associated by default with this class, then we make masculine the default gender for nouns, without specifying it directly. Sources vary as to the numerical preponderance of the genders. All agree that the neuter is easily the weakest; Lazova (1974: 942-943) has only marginally more masculines than feminines, while Mučnik (1971: 196-197) calculates that the masculine has 46% to the feminine 41% (on a sample of 33,952 nouns) and Zaliznjak's dictionary has 46% masculine and 38% feminine (sample 47,030; see Ilola and Mustajoki 1989: 9); the masculine is growing fastest, from the assignment of borrowings (see the sources reported in Corbett 1991: 78).

The only change required in our DATR theory to include all of these interacting defaults is given below, as a revision to the fragment shown above as (9).

(10) DEF_DECL:
    <phonological_type $sex> == N_I:<>

DECLENSION:
    <> == DEF_DECL
    <consonant female> == N_II:<>
    <vowel $sex> == N_V:<>

Here, the variable $phonological_type ranges over consonant and vowel, and $sex ranges over male, female and undifferentiated, as before. In all cases other than the two explicitly described in the paths at the DECLENSION node, the maximally general path specified at the DEF_DECL node is inherited and, thereby, class I becomes the default declensional class for nouns.

4.3. A higher-level default

We argued above that if class I is the default declensional class, and masculine is the gender associated by default with this class, then we make masculine the default gender. While this result fits with the intuitions of some investigators, it seems unsatisfactory to others. If masculine is the default gender for Russian, then we would expect it to appear, for instance, in examples like the following:

(11) Byl-o xolodn-o
    was-NEUT.SG cold-NEUT.SG
    'It was cold.'

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 resolution of this apparent paradox is that we are claiming only that the masculine is the default gender for nouns. Taking a broader view, we would claim that there is a default for gender at a higher level than the nodes relating directly to nouns. This higher default is necessary for items other than nominals which may head syntactic constituents with which gender agreement is required. The situation arises if, say, an infinitive phrase stands in subject position (and there is a past tense verb), or there is an interjection or other quoted material. Here we normally find the neuter.
Thus in an analysis of the full lexicon, at a high level, perhaps at the level of word, the default gender is the neuter; at the lower level—that of nouns—it is the masculine. Of course, the lower default is much the more important: gender is a central category for Russian nouns, while for the various items (such as infinitives) covered by the higher default, it is of much less importance. This hierarchy of defaults would thus allow us to capture the intuition that in one sense the masculine is the default gender, while in another it is the neuter.

Slavists will recognise, however, that to say that the top level default is the neuter is an oversimplification. Before returning to see why this is so, we should first consider more carefully the nature of defaults, as applied to linguistic phenomena.

5. Types of default

Consider the following situation. Mary and John both work for a firm in Clacton. Mary is the personnel manager and has her office in Clacton. Occasionally, when there are problems or training courses she spends the day at head office in Truro. By default, then, Mary works in the office at Clacton. John is a salesman. He normally spends Mondays in the south of England, Tuesdays in Wales, and Wednesdays and Thursdays in the north. If, however, clients cannot see him, or his car is unserviceable, or there is a department meeting, he goes to the office in Clacton. Fridays he usually spends exhausted in bed, but during school holidays he goes to the office. By default, then, John also works in the office in Clacton.

Intuitively the two cases are rather different. Mary is "normally" at the office, John is not. And yet at a higher level of abstraction it is true to say that the office is the default workplace for both. It is these two types of default, both reasonable uses of the term, which have led to the differences in usage in the literature, both generally and specifically in relation to gender.

In our analysis of the gender system of Arapesh (Fraser and Corbett forthcoming) based on Aronoff (1992, 1994: 97-103) and like him following Fortune (1942)) we distinguish these two types of default which, though conceptually related, are nonetheless formally distinct. In the first type, the default accounts for the cases when "everything goes right" (as in Mary working in the office). 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 John working in the office). We shall call instances of this type "exceptional case defaults".

There is a common conceptual core running through both usages of the term "default": the default is the last thing you get to. However, a normal case default is retrieved after failing to find any more specific value; an exceptional case default is retrieved after finding too much information - information which blocks normal retrieval and causes a backstop value to be accessed instead. One form of default is concerned with typicality, the other with exceptionality. It is therefore particularly important that conceptual and terminological confusion be avoided by proper definition of terms.

One of the payoffs of working in a formally explicit framework such as Network Morphology is that it lays bare the differences between these otherwise confusable notions. In instances where normal case defaults apply, lexical entries are characterized by their brevity. Because a word is fairly typical, many of its parts can be left underspecified, to be filled in by default inheritance. On the other hand, there is an inverse correlation between radically underspecified lexical entries and exceptional case defaults. An exceptional case default is unlikely to apply unless the lexical entry includes some idiosyncratic information.

With all this in mind, let us return to example (11), repeated here for convenience:

(11) Byl-o was-NEUT.SG
    xolodn-o cold-NEUT.SG
    'It was cold.'

There are various circumstances where similar forms are required: when there is no agreement controller as in (11), when the controller is a clause or infinitive phrase and so on. It is reasonable therefore to invoke a default here. The circumstances are, roughly speaking, all those where agreement is not controlled by a prototypical noun phrase (one headed by a noun or pronoun). This can be seen as things
going wrong (agreement is normally controlled by a prototypical noun phrase), and so we have an exceptional case default.

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 is most uses; however, in examples like (11), the short form is required. Thus it is not sufficient to say that these forms are neuters. For these and further reasons discussed at greater length elsewhere (Corbett 1980) we need to distinguish "neutral" agreement forms, as required for agreement with non-prototypical controllers, from other agreement forms.\(^{13}\) 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 Farkas, see Corbett 1991: 213-214).

\[(12)\]  
\[e\text{ evident }\, c\,\text{ că }\, a\text{ venit,}\]
\[\text{is clear.MASC.SG that has come}\]
\[și\,\text{ asta }\, o\]
\[\text{and this.FEM.SG it.FEM.SG}\]
\[știe\,\text{ toată }\, lumea\]
\[\text{knows all the.world}\]

'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, which has to mark agreement, is masculine (the feminine evidentă, "is unacceptable). Asta 'this' can stand for 'that s/he came' or 'it is clear that s/he came'. What concerns us is that it must be feminine (the masculine a, 'sta 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:

\[(13)\]  
\[Un\,\text{ a.MASC.SG. }\, bum\,\text{ puternic}\]
\[\text{"boom" strong.MASC.SG}\]

\[a\text{ fost auzit}\]
\[\text{has been heard.MASC.SG}\]

'A loud boom was heard.'

Here un 'a' is masculine, like the agreeing predicate. We now try the relative pronoun:

\[(14)\]  
\[a\text{ admis că }\, a\text{ venit, ceea ce}\]
\[\text{has admitted that has come which.FEM.SG}\]
\[nu\,\text{ e surprinzător}\]
\[\text{not is surprising.MASC.SG}\]

'S/he admitted that s/he came, which is not surprising.'

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

\[(15)\]  
\[asta\,\text{ e uluitor}\]
\[\text{this.FEM is amazing.MASC}\]
\[\text{'this is amazing'}\]

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

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

6.1. Gender assignment

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 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 animates, non-sexed animates and all inanimates" (Austin 1981: 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 Straits Islands (Bani 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.

6.2. Gender agreement

Here we shall discuss three broad types of problem, which are 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.

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 noun phrases. The second type of problem involves cases where the choice of gender agreement in the normal way would force greater specificity than is possible (or perhaps desirable) for the speaker. A speaker may wish to refer to a child but be unable to select gender agreement based on sex. Finally there are complex noun phrases which are overspecified for gender.

6.2.1. Non-prototypical controllers

As noted earlier, if an agreement target can agree then typically it must agree, even if the agreement controller lacks the appropriate features. We termed this enforced agreement "neutral agreement". The idea of "neutral agreement" can be seen as a development of Jespersen's "conceputal neuter" (1924: 241–243); the wider term is preferred since the phenomenon is found in gender systems other than the Indo-European type, to which Jespersen restricted himself. 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 (see Klaon 1984–85: 351) for examples of non-prototypical controllers from various Indo-European languages.

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" or the "default agreement form". The first problem which arises is the motivation for the selection of a particular form.

In some languages a semantic account is plausible; thus in Russian, almost all nouns denoting humans are in the masculine and feminine genders, while inanimates are distributed across all three genders. The use of the neuter for default agreement could be understood as the selection of the gender which is most appropriate in semantic terms (thereby avoiding the semantic clash of neutral with human, which would arise with the other genders). Similar arguments can be given for languages as diverse as the Algonquian language Menominee, the Omotic language Zaye, the Niger-Kordofanian language Fula, the Mon-Khmer language Khmer, the North East Caucasian languages Archi and Khinalug and the *Australian language Ungarinj (see Corbett 1991: 206-207)

The fact that a semantic explanation works for examples which are so diverse genetically might make us expect it to apply without
exception. Nevertheless there are languages for which the semantic criterion fails. In some we find that a different criterion is at work, but we leave others as perplexing problems. Lak is a language which, though related to Archi and Khinalug, differs from them in an interesting way. Again there are four genders, and nouns are assigned to them as follows: I - male humans, II - some female humans (older females), III remaining female humans, most other animates, some inanimates, IV - a very few animates, some inanimates. Again we would expect gender IV to be used for neutral agreement, as in Archi and Khinalug; in fact gender III is used (Kibrik 1979: 13):

(16) g aniša d-uč'an b-uqlalaj b-u-r
she II-come III-can III-AUX.3RD
'she can come'

The gender II agreement marker d- on the infinitive d-uč'an is controlled by a deleted subject noun phrase g.a, coreferential with g aniša. The point of interest is the agreement on the modal and on the auxiliary; both have a gender III prefix, since they are controlled by the infinitive. Thus gender III is used for neutral agreement and there is no evident explanation in terms of the semantics of the controller genders as to why this should be (gender IV would be expected). However, as Aleksandr Kibrik points out (personal communication), an explanation is available if we take into account the pattern of the target gender forms, given in Table 3.

Table 3. Gender agreement in Lak

<table>
<thead>
<tr>
<th>Gender</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender I (male humans)</td>
<td>Ø/w</td>
<td>b/w</td>
</tr>
<tr>
<td>gender II (some female humans - older)</td>
<td>d/r</td>
<td>b/w</td>
</tr>
<tr>
<td>gender III (most remaining animates, some inanimates)</td>
<td>b/w</td>
<td>b/w</td>
</tr>
<tr>
<td>gender IV (residue - largely inanimates)</td>
<td>d/r</td>
<td>d/r</td>
</tr>
</tbody>
</table>

Forms before the slash are prefixal, those after it are internal or suffixal. As can be seen in Table 3, the forms for gender IV and gender II are identical in the singular. The use of the gender IV marker d/r for neutral agreement is ruled out by the fact that this marker is also the marker for gender II, which contains only nouns denoting females. This leaves b/w, the gender III agreement form as the only possibility for neutral agreement. This particular syncretism is not found in Archi and Khinalug, and so in those languages the gender IV forms are available for neutral agreement.

Thus far we have found an explanation for the choice of neutral form in several languages, according to the semantics of the controller genders (with additional complications in the case of Lak). We now turn to cases which are initially surprising and for which the semantics of the controller genders appears to provide no help. But in these cases we find an explanation available in terms of the gender agreement forms. For example, the East Cushitic language Bayso has two genders, masculine and feminine, and uses the masculine agreement forms for neutral agreement (Corbett and Hayward 1987: 11n):

(17) ibaaddo boc, aano
    person to beat
    ka - meelan - ya
    PARTICLE.MASC - ḥad - COP.3RD.SG.MASC

'To beat people is bad.'

In the last item, ka- is the associative particle in the masculine form and -ya is the clitic copula, third singular masculine. In Qafar, another Cushitic language, it is the feminine which is used. This can be shown with the so-called 'm-nominalization'. The clitic element -m attaches to a wide range of items, and the resulting form takes feminine agreement, despite ending in a consonant and so being of a phonological shape associated with masculines:

(18) gaddali kinnim yōt
    wealthy.man is. 3RD.SG + m to.me
    celta
    seems.3RD.SG.FEM

'He seems to me to be a wealthy man.'
Here the -*m* is attached to the phrase 'is a wealthy man' and this new element is a non–prototypical controller; as a result, the verb *celta* stands in the neutral form, which is the feminine (Hayward and Corbett 1988:266–268).

There are no obvious differences in the semantics of gender in the two languages to explain the different choice of neutral form (R. J. Hayward, personal communication). However, if we look at the morphology of agreeing forms there is a ready explanation. Table 4 gives the forms of the associative particle for the two of the four numbers of Bayso which are significant for differentiating gender, and the verbal agreement markers of Qafar which are found in agreement with simple noun phrases.

<table>
<thead>
<tr>
<th>Table 4. Patterns of syncretism in Bayso and Qafar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>MASULINE</strong></td>
</tr>
<tr>
<td>ka</td>
</tr>
<tr>
<td><strong>FEMININE</strong></td>
</tr>
<tr>
<td>ta</td>
</tr>
</tbody>
</table>

There is considerable syncretism: one form covers three of the four theoretical slots, being used for the singular of one gender and the plural of both. And in each case it is this form, the one with the wider range of use, which is the neutral agreement form. Thus the neutral form is here aligned with a default inflectional form.

It is worth clarifying the criteria we have employed so far. In the first languages discussed it was the semantics of the controller genders—the genders into which nouns are divided—which provided the reason for the choice of neutral agreement form. In these languages there is a gender for abstracts or inanimates, or at least one containing few if any nouns denoting humans, which appeared an appropriate gender for non–prototypical controllers to attach themselves to. In the case of Bayso and Qafar, however, there is no obvious choice in terms of the semantics of the controller genders, since humans and other animates are found in both of the available genders, as are inanimates. Here a second criterion comes into play: the choice of the neutral form can be understood in terms of the morphology of the agreeing elements (target gender forms). Lak can be seen as an intermediate type, in which the semantics of the controller genders and the pattern of syncretism of the target gender forms interact to determine the neutral agreement form.

Some instances are more difficult, as in the Bantu language Chichewa:

(19) "aaa" a - na - mw - eka
"aaa" 1 - PAST - hear - PASSIVE
'an "aaah" was heard'

Remarkably, class 1 agreement (as for humans) is used here, as the agreement marker prefixed on the verb suggests; to demonstrate this conclusively requires an example with object agreement (Corbett and Mienie 1987: 14):

(20) a - na - mu - mwa
1.SBJ - PAST - 1.OBJ - hear | "moyo" | crying sound
'he heard a crying sound'

The combination of *a*- as subject agreement marker in (19) and *-mu*- as object agreement marker in (20) means that we are dealing with class 1 markers. What makes the Chichewa case particularly surprising is that there is a special agreement form for infinitives, and this is not used for neutral agreement.

Though neutral agreement forms may appear (in some languages) 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. This can be illustrated by conjoining. 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 Sele 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 (for references see Corbett 1991: 216 and for extensive discussion see Källström 1993: 188-246 and Hedlund 1992: 95-111). Our example is Norwegian, taken from Faarlund (1977). Norwegian predicative
adjectives distinguish singular from plural and, in the singular, neuter from common:

(21) Een ny utanriksminister ville ikkje vere
a new foreign secretary would not be
sdז dumt.
so stupid.NEUT.SG

'A new foreign secretary would not be a bad idea.'

The interpretation is that having a new foreign secretary would not be a bad idea. If the adjective were in the common form dum, then it would agree directly with the subject noun phrase and the interpretation would be less complimentary.

Default agreement then may be required in a range of circumstances: agreement with a clause as subject, with an absent subject, and so on. Note that the range of circumstances varies from language to language. In Russian, as in many other Indo-European languages, an infinitive phrase requires default agreement, while in Chichewa, and many other Bantu languages, there is a special gender for infinitives and the default is not invoked here. We have seen too that the default form overlaps to a greater or lesser extent with "normal" forms, and that the gender form with which it overlaps can be successfully predicted in some languages at least. In these instances a careful analysis would be required to determine which are the normal case defaults and which are the exceptional case defaults.

Before moving on to our second major type of agreement default, we will consider two further types of case which might be expected to fit easily under neutral/default agreement but which in fact do not.

The first concerns missing antecedents of pronouns, and is a problem raised by Tasmowski-De Ryck and Verhuyten (1981; 1982). Consider the following French example. John is trying to get a desk into his car, then the pronoun would be masculine (bureau 'desk' is masculine).

Here the pronoun must be in the feminine form la (*le is unacceptable); table 'table' is feminine in French. However, if John were trying to get a desk into his car, then the pronoun would be masculine (bureau 'desk' is masculine).

The pronoun must be in the appropriate gender; since there is no antecedent present, the question is the source of the gender of the pronoun. These data have been one of the stimulants to an interesting debate on the nature of deixis and anaphora (see Corbett 1991: 244 for extensive references). Most agree that the gender of the pronoun in examples like (22) is determined by that of the default description of the referent in question. Normally this default description will be the basic level term, a notion which comes from Rosch (see, for example, Rosch 1978, and for discussion see Pulman 1983: 83-106). The basic level is the appropriate level. In (22) above, the basic level term for the object in question is table, which is feminine, hence the use of la. The fact that it is a piece of furniture (meuble, masculine) and indeed an object (objet, masculine) does not affect the gender since neither are basic level terms (Bosch 1987: 73). Hence we have another, rather different, sort of default, and one which need not coincide with the neutral agreement type.

The second type concerns citation forms and other uses where there is no possible antecedent. For instance, if speakers wish to discuss a particular word, in a language where items of that class always agree in gender, then they will have to use one of the gender forms. A specific instance of this general problem is the use of numerals in abstract counting. It would surely be a reasonable prediction, that in such instances the neutral form (as for, say, impersonal sentences) would be used. Reasonable but not necessarily true. If counting in the abstract, a Russian speaker says dvadcat 'odin 'twenty-one.MASC', even though the gender-differentiating odin 'one' has a neuter form (odno). We might also have expected that older defaults would be preserved in fossilized form in complex numerals like odinmadcat 'eleven' and dvenadcat 'twelve'. But again the situation is more complex: in odinmadcat 'eleven' the masculine odin is preserved, while in dvenadcat 'twelve' we apparently have the feminine dve, though the form dates from a period when dve was the form for neuter as well as feminine (see Suprun 1969: 51-52 for an argument for dvenadcat based on the number of examples of use with feminine nouns (including weights and measures) and Comrie 1992: 768 for discussion).
6.2.2. Reference problems

Even if the agreement controller is a noun phase headed by a noun or pronoun, there may still be problems involving gender agreement, caused by reference difficulties. 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 types of problem (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. In this first type 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 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; here there are various possibilities: *If a patient wishes to change doctors, he/she or she/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 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 Russian examples:

(23) *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 Maasai, we find the feminine used for questions when the person involved could be male or female (Tucker & Mpaaye 1955:27).

Returning to Russian, nouns which can denote a male or a female, like *vrae* 'doctor', take masculine agreements if the sex is not known. In Archi, however, we find an "evasive" form. Archi has four genders, I and II for humans, male and female, III and IV less clearly defined semantically but with the larger animates in III and most abstracts in IV. In Archi, nouns like *lo* 'child', *adam* 'person', *c'hor* 'thief', *misgin* 'poor person' take gender IV agreements in the singular if the sex of the referent is unimportant or unknown (Kibrik 1972:126). Archi shows a particularly clear example of an evasive form, since gender IV does not contain any nouns denoting humans.14

The second type of problem, non-specific referents, has created a considerable literature, but generally with reference to a small number of languages. Languages like Russian use the masculine here (see, for instance, Doleschal 1993:39).

The third 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 we find the masculine plural one 'they' in such cases. We may take the problem back into derivational
morphology: Amerikanac (masculine) is a male American, while Amerikanka (feminine) is a female American in Serbo-Croat. To refer to Americans in general, the plural of the masculine noun is used, that is, Amerikanci, 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 people are referred to using the feminine pronouns (John Payne personal communication).

Here too "evasive" forms are an alternative strategy. Polish for instance uses the neuter singular. This usage is described by Gotteri (1984), who took up the term "evasive" following a suggestion by Doroszewski. An example of the Polish neuter in evasive use is the following:

\[
(24) \text{Któr-es z małżonków jest one-NEUT from spouses is}
\]
\[
\text{winn-e zarzucanej mu zbrodni guilty-NEUT imputed it.DAT crime}
\]

'One of the spouses is guilty of the crime he or she has been accused of.'

Małżonkówie 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: 67). 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 specifically female. Weiss (1993) also discusses the Polish situation. The evasive neuter is found in Serbo–Croat too, but in rather limited use.

Though a full analysis would be required in order to be certain, it would appear that in this section, the examples that use one of the expected genders should be treated as normal case defaults, while the evasive forms are exceptional case defaults.

6.2.3. Gender resolution

Our last type involves controllers with too many gender specifications, namely those consisting of conjoined noun phrases. Languages with gender agreement in the plural require rules to determine the form to be employed with conjoined noun phrases, and these rules are termed "resolution rules".

As an example, we take Tsova-Tush, sometimes called Bats or Batshi, which is a Nakh language (a subgroup of North-East Caucasian). It has some 3,000 speakers in Zemo Alvani in Kakheti (in eastern Georgia); all are bilingual with Georgian, and children are no longer learning the language. The data come from Dee Ann Holisky (forthcoming, and unpublished field notes). Tsova-Tush has five main genders, and the target agreement forms make up a complex crossed system (shown in Figure 3; next page).

Nouns which take v – b (v in the singular and b in the plural) denote male humans and those which take j - d denote female humans. Thus far gender is semantic, and predictable. The basis for the remaining three major genders has not been established. There are just over 20 nouns which do not fit into the five main genders (Holisky forthcoming; see Corbett 1991: 171-172 for discussion of their status). They belong to three further classes: agreements b - j are taken by 15 nouns, all of which denote body parts (such as 'leg' and 'throat'), a further four nouns, also denoting body parts (this time all paired, like 'ear' and 'hand') take d - j, while three nouns take b - b (they are horag 'knit slipper', ërkam 'boot', and kakam 'wool cut in autumn').

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th></th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td></td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td></td>
<td>j</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The main genders in Tsova-Tush
The gender resolution examples are straightforward, though unusual. If all conjuncts are headed by nouns denoting male humans, then the male human form (b-) is used (Holisky forthcoming, example (74)):

(24) mit'o-E p'et'o-E ha^n važar-i b-a
   Mito-and Peto-and 2SG.GEN brothers-INT AG-be
   'Are Mito and Peto your brothers?'

Conjoining is marked by -E on each conjunct; variant forms are for phonological reasons (Holisky forthcoming; footnote 4). E is a reduced vowel; h is a voiceless pharyngeal fricative and e^n is a nasalized vowel. INT indicates the question marker, and AG is for agreement marker. (Holisky fieldnotes) reports checking various instances with conjuncts referring to male human and female human; in these mixed cases the d-form was used.

In all other circumstances (when not all conjuncts are headed by nouns denoting male humans) the d-agreement form is used (the form which is appropriate for female humans, and for two of the remaining major genders). There are a large number of possible combinations: female human plus female human, female human plus other, and the various types of noun not denoting humans, including those from the smaller genders. Here are just a few examples. The first has two nouns from the largest gender, d-d (Holisky fieldnotes):

(25) kotama-E mamala-E d-a
    chicken-and rooster-and AG-be
    'a chicken and a rooster are: It is a chicken and a rooster.'

The next example has two nouns from the large j-j gender (Holisky field notes; in (26) E represents a voiced velar fricative, and E is again a reduced vowel):

(26) ęga-E ezo-E d-a
    fence-and yard-and AG-be
    'a fence and a yard are: It is a fence and a yard.'

In the next example (Holisky forthcoming example (26b)) we have two nouns from the small genders b-j and b-b:

(27) kok'a-E borga-E d-a
    leg-and knit slipper-and AG-be
    'a leg and a knit slipper are: It is a leg and a knit slipper.'

The resolution rules are again based on semantics:
1. By default, the gender d-form is used.
2. This is overridden if all conjuncts are headed by nouns which denote male humans; in this case the male human form (b-) is used.

For a typology of resolution rules, see Corbett (forthcoming). We might expect resolution rules to be examples of normal case defaults. However, there are systems of some complexity in which it is at least not obvious that they can be handled in this way.

7. Correlation of defaults

It is important to note that grammatical 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). Let us return to Tsowa-Tush. We saw that the default form in gender resolution is the d-gender. Of the numerals in Tsowa-Tush, just the numeral 4 (and numbers which contain it) agree. When counting in the abstract, it takes the d-agreement form (Holisky forthcoming, section 2.8). The d-gender is the largest one in terms of the number of nouns included and so a reasonable candidate for the default for gender assignment (section 2.1.1.), and verbal nouns, which might be thought of as non-prototypical nouns, are assigned to this gender. Consider finally what happens when the appropriate gender is unknown:

(28) mwx d-a?
    what AG-is
    'What is it?'

Again the d-gender form is used. Thus we have a clear example of several defaults lining up.
An equally impressive case is Arapesh, as described by Aronoff (1992, 1994: 97-103, following Fortune 1942). The same default gender is used for null heads (1992: 26), interrogatives, including the interrogative for humans when sex is not known (1992: 27), for agreement with conjoined noun phrases when the head nouns are of different gender (1992: 27). Nouns which fall outside the regular gender assignment rules are assigned to this same default gender (1992: 28) as are "words that designate persons in a sex-neutral fashion" (1992: 29).

These examples from large gender systems might lead us to expect that gender defaults regularly line up. However, there are other examples where this is certainly not the case. Moreover, we can show this even in relatively small systems. If we return to the Russian data discussed in detail earlier, we recall that at one level the default gender is neuter, and at another it is masculine. Then for nouns of particular inflectional classes the default is feminine. Russian of course has three genders, but even in a two-gender system the defaults need not line up. 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: 173).

8. Conclusion

The notion of default is of considerable potential value in the analysis of gender systems; indeed it could be so valuable that it is important to be clear in its use. When we attempt to cover a complete system we find that the defaults at different levels may well be different. Just as they may line up fairly consistently, they may also diverge, even in small systems. Since they can diverge, we need more descriptions of complete systems, so that we can attempt a typology of configurations of defaults in gender systems.

Notes

1. Addresses for correspondence: Greville Corbett, Department of Linguistic and International Studies, University of Surrey, Guildford, Surrey, GU2 SXH, UK. Norman Fraser, Vocalis Ltd, Chaston House, Mill Court, Cambridge CB2 5LD, UK.

email: g.corbett@surrey.ac.uk, norman@vocalis.com.

This research was supported in part by the Economic and Social Research Council (grants R000233633 and R000233603) and the Leverhulme Trust (grant F.242M); this support is gratefully acknowledged.

Some of the data in this paper have been presented previously: sections 3 and 4 use material from Fraser and Corbett (1995b) while section 6 takes data from Corbett (1991, chapter 7). What is new here is the central attention given to the notion of default, and the attempt to understand familiar data in this light. In addition, in the sections based on earlier work, new references to recent work have been added. A version of this paper was read at the workshop 'Approaches to Gender' held at the Forschungsschwerpunkt Allgemeine Sprachwissenschaft, Typologie und Universalforschung, Berlin 26-28 May, 1994. We are grateful to those present for interested discussion, and especially to Wolfgang Wurzel and David Zabin for helpful suggestions at the meeting, and to Osien Dahl, Ursula Deleschul, Gerald Goldar and Andrew Hippisley for subsequent useful comments.

2. For further discussions of markedness see Schwartz (1980) and Greenberg (1988).

3. The selection of DATR as the language of formal expression for Network Morphology is no accident. The theoretical framework emerged out of our exploratory work with DATR, informed by our earlier experience of default inheritance in Artificial Intelligence and in Word Grammar (Fraser and Hudson 1992). DATR offers a conceptually straightforward syntax and a well-defined semantics. In addition, the existence of a number of computer interpreters for the language allow theories encoded in DATR to be validated automatically.

4. The following automatic phonological correspondences are assumed:
a) /h/ is retracted to its allophone [i] after non-back hard (unpalatalized) consonants. Thus the nominative plural form /zakon/ will be realized with [i] but /host/ retains [i] since [j] is soft.
b) All consonants which can be palatalized are automatically palatalized before /el/. Thus the locative singular of /zakon/, namely /zaken/ will be realized with a palatalized [n'']. If the consonant is already palatalized as in genitive plural /host'-el/, it simply remains palatalized. Some consonants are always hard (/s, z, ch/), and remain so before /el/. On the other hand, /k'/ and /g'/ are always soft (palatalized), and naturally remain so before /el/. We have chosen to mark softening redundantly for greater clarity in this instance. In addition, the gutturals /h, g, x/ are palatalized before /h/, so that the genitive form /kniq/, from /kniq/ "book", will be realized with palatalized [y'] (which then demands the front allophone [i]);
c) There are complex patterns of reduction of vowels in unstressed position, which can safely be omitted from the transcription since our focus is on morphology. In particular, the unstressed /o/ ending of nouns with soft stems such as /pol/ 'field' (orthographically pole) is realized as either [a] or [i]. Although this /o/ is never realized as anything approximating to a mid
rounded back vowel, positing /o/ is justified by the stressed /e/ which occurs in [ˈvɪnə] ‘wine’ and [ˈpɪtʃ] ‘drink(ing)’.

For an informative sketch of Russian phonology, see Timberlake (1993: 828-832).

5. We do not examine the complexities of hybrid nouns here, for which see Corbett (1991:39, 231-232), and for interesting data see Weiss (1991) and Dolezichl (1992; 1993:40-46, 138-144).

6. There is a small number of exceptions, for which see Smirnova (1979). We are not concerned here with the special case of acronyms where derivational information may be relevant.

7. A more extended account, to include cases where nouns may head noun phrases referring to either sex, would require us to state the first equation as: GENDER: ∼ ∼ ∼ male. Instances of this kind would be correctly assigned masculine gender so long as the indeterminate sex of the referent were identified by an explicit marker such as either.

8. There is no reason why such a phonological component could not also be formalized using DATR, as the work of Reinhart and Gibbon (1991) and Gibbon (1992) demonstrates.

9. While the claim relating to gender assignment is a general one, we have not investigated declensional-class assignment beyond Russian. It is a plausible hypothesis that in declensional class assignment generally, formal factors will take precedence over semantic.

10. If the stem ends in a hard (not palatalized) consonant, then type III can be eliminated. We have no rule to predict membership of class III; stems are by default “hard” (they end in a hard consonant). The III class is relatively small (see discussion after Table 2).

11. This figure includes the small number of nouns (32 in Lazov’s table) which are indelible but which do not end in a vowel. They are typically borrowings. Borrowings like miss ‘miss’, which denote females but end in a consonant are indelible. These require an exceptional marker just to indicate that they are indelible: the semantic gender assignment rule will then correctly assign them to the feminine gender. Rare borrowings which are indelible, consonant final, do not denote an animate, and are masculine or feminine, require two irregular markers (for declensional class V, and for gender). These marginal cases tend to be integrated over time or lost.

12. An important paper which includes consideration of what we would call an exceptional case default is Marcus—Brinkmann—Chaksen—Wiese—Woo—Pink (1993) where -s is discussed as the default plural of German. They quote Van Dam (1940) who called it the *plural ending 'emergency plural ending'*. See also Chaksen—Rodwell—Woo—Marcus (1992).

13. For discussion of the forms used in English for “situation reference” see Frake (1992).

14. Another evasive device is to use the plural, if gender is not distinguished there. Thus Alambik (a Sepik Hill language of Papua New Guinea) distinguishes masculine and feminine in the singular, but not in the dual or plural. If “the speaker is either unable or unwilling to indicate the gender of an object” (Bruce 1984:98) then the third plural is used:

(i) yen-m heurahum inom yamtn
    child-3PL she.will.bear.them another month.in
    ‘She will bear a child in another month.’

15. We are very grateful to Dee Ann Holisky for making her notes available to us.

16. There is, however, an interesting complication with conjunct plurals, which would take the same target gender form (either j or b). There are a couple of cases of an informant offering this form (the plural which each conjunct would take individually). For discussion of the same phenomenon in Chichewa and in Serbo-Croat see Corbett and Mitenje (1987: 20-26). It is argued that in these languages gender resolution is triggered only if there is a clash of genders or if another resolution rule (number or person) operates - since if one type of resolution rule operates, all available resolution rules must operate.

References

Alper, Barry
1987

Aronoff, Mark
1992

1994

Austin, Peter
1981
*A grammar of Dyirbal, South Australia*. Cambridge: Cambridge University Press.

Bani, Ephraim
1987

Bosch, Peter
1987

Briscoe, Ted—Valeria de Paiva—Ann Copestake (eds.)
1993
*Inheritance, defaults and the lexicon*. Cambridge: Cambridge University Press.

Brown, Dunstan P.—Andrew R. Hippisley
1994

Bruce, Les
1984
*The Alambik language of Papua New Guinea (East Sepik)* (Pacific Linguistics, Series C, No. 81). Canberra: Department of
Linguistics, Research School of Pacific Studies, Australian National University.

Calder, Jo

Clahsen, Harald—Monika Rothweiler—Andreas Woert—Gary F. Marcus

Comrie, Bernard

Corbett, Greville G.
1979  Predicate agreement in Russian (Birmingham Slavonic Monographs 7). Birmingham: University of Birmingham.

Corbett, Greville G.—Norman M. Fraser

Corbett, Greville G.—R. J. Hayward

Corbett, Greville G.—Alfred D. Mtenje

Daelemans, Walter—Koennard De Smedt—Gerald Gazdar

Dolezel, Ursula

Dziwirck, Katarzyna

Evans, Roger—Gerald Gazdar

Faaehund, Jan Terje

Fortune, Roco F.
[1977]  [Reprinted, New York: AMS Press.]

Fraser, Norman M.—Greville G. Corbett

Fraser, Norman M.—Greville G. Corbett

Fraser, Norman M.—Richard A. Hudson

Fraunfeld, Kari

Gazdar, Gerald
CSRP 139, School of Cognitive and Computing Sciences, University of Sussex. 1-14.

1992


Gazdar, Gerald—Ewan Klein—Geoffrey K. Pullum—Ivan A. Sag
1985

Gibbon, Dafydd
1992

Goteri, Nigel
1984

Greenberg, Joseph H.
1966

1988

Hayward, Richard J.
1989
"The notion of 'default gender': a key to interpreting the evolution of certain verb paradigms in East Ometo, and its implications for Ometo", Afrika und Ubersee, 72: 17-32.

Hayward, Richard J.—Greville G. Corbett
1988

Hedlund, Cecilia
1992

Herbert, R. K.—Nykiel-Herbert, B.
1986

Holisky, Dee Ann (with Rusudan Cagvari)

Ilola, Evu—Arto Mustajoki
1989
Report on Russian Morphology as it appears in Zalijnyak's Grammatical Dictionary (=Slavica Helsingiensia 7). Helsinki: Department of Slavonic Languages, University of Helsinki.

Jackendoff, Ray
1975

Jespersen, Otto
1924

Källström, Roger
1993

Kibrik, A. E.
1972

1979

Klajn, Ivan
1984-85

Lazova, M. V. (ed.)
1974
Obratnyi slovar' russkogo jazyka: okolo 125 000 slov. [Reverse dictionary of Russian: about 125,000 words.] Moskva: Sovetskaja Entsiklopedija.

Marcus, Gary F.—Ursula Brinkmann—Harald Clasen—Richard Wiese—Andreas Woest—Steven Pinker
1993

Mačnik, I. P.
1971
Gramatičeskie kategorii glagola i imeni v sovremennom russkom literaturnom jazyke. [The grammatical categories of the verb and the noun in contemporary standard Russian.] Moskva: Nauka.

Pulman, Stephen G.
1983
Word Meaning and Belief. London: Croom Helm.

Reinhard, Sabine—Dafydd Gibbon
1991
Rosch, Eleanor  

Schane, Sanford A.  

Schupbach, Richard D.  
1984  Lexical Specialization in Russian (UCLA Slavic Studies 8). Columbus, Ohio: Slavica.

Schwartz, Linda J.  

Shieber, Stuart M.  

Samarin, G. A.  

Suprun, A. E.  

Tasinowski-De Ryck, Liliane—S. Paul Verheyen  


Timberlake, Alan  

Tucker, A. N.—J. Tompo ole Mpayei  

Van Dam, J.  

Weiss, Daniel  

Worth, Dean S.  

Zaliznjak, Andrej Anatol'evich  

Zwicky, Arnold  