

*Language, culture and cognition*

Editor Stephen C. Levinson, Max Planck Institute for  
Psycholinguistics, Nijmegen

This new series looks at the role of language in human cognition – language in both its universal, psychological aspects and its variable, cultural aspects. Studies will focus on the relation between semantic and conceptual categories and processes, especially as these are illuminated by cross-linguistic and cross-cultural studies, the study of language acquisition and conceptual development, and the study of the relation of speech production and comprehension to other kinds of behaviour in cultural context. Books come principally, though not exclusively, from research associated with the Max Planck Institute for Psycholinguistics in Nijmegen, and in particular the Cognitive Anthropology Research Group.

- 1 Jan Nuyts and Eric Pederson (eds.) *Language and Conceptualization*
- 2 David McNeill (ed.) *Language and Gesture*
- 3 Melissa Bowerman and Stephen Levinson (eds.) *Language Acquisition and Conceptual Development*
- 4 Gunter Senft (ed.) *Systems of Nominal Classification*

SYSTEMS OF NOMINAL  
CLASSIFICATION

EDITED BY

GUNTER SENFT

*Max Planck Institute for Psycholinguistics, Nijmegen*

 CAMBRIDGE  
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk  
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org  
10 Stamford Road, Oakleigh, Melbourne 3166, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain

© Cambridge University Press, 2000

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Set in 11/12.5pt Baskerville No. 2 [G G]

*A catalogue record for this book is available from the British Library*

*Library of Congress Cataloguing in Publication data*

Systems of nominal classification / edited by Gunter Senft.  
p. cm. – (Language, culture and cognition; 4)  
Includes bibliographical references.

ISBN 0-521-77075-0

1. Categorization (Linguistics) 2. Classifiers (Linguistics)  
3. Grammar, Comparative and general – Nominals.

I. Senft, Gunter, 1952– II. Series.

P128.C37 S97 2000

415-dc21 99-051374

ISBN 0 521 770750 hardback

## Contents

|  | <i>page</i> |
|--|-------------|
| <i>Acknowledgements</i>  | ix          |
| Introduction<br>GUNTER SENFT   | I           |
| 1 What do we really know about nominal<br>classification systems?<br>GUNTER SENFT  | II          |
| 2 A morphosyntactic typology of classifiers<br>COLETTE GRINEVALD   | 50          |
| 3 Unusual classifiers in Tariana<br>ALEXANDRA Y. AIKHENVALD  | 93          |
| 4 Multiple classifier systems in Akatek (Mayan)<br>ROBERTO ZAVALA  | 114         |
| 5 Ants, ancestors and medicine: a semantic and<br>pragmatic account of classifier constructions in<br>Arrernte (Central Australia)<br>DAVID P. WILKINS       | 147         |
| 6 Visualizing ability and nominal classification:<br>evidence of cultural operation in the agreement rules<br>of Japanese numeral classifiers<br>KYOKO INOUE | 217         |
| 7 Isolation of units and unification of isolates:<br>the gestalt-functions of classifiers<br>JÜRGEN BROSCHE  | 239         |

- Paroz, R. A. 1974. *Southern Sotho-English Dictionary*. Morija: Morija Sesuto Book Depot. (Revision of A. Mabile and H. Dieterlen, *Sesuto-English Dictionary*).
- Richardson, I. 1967. *Linguistic evolution and Bantu noun class systems*. Paris: Centre National de la Recherche Scientifique. 373-89.
- Spitulnik, D. 1987. *Semantic superstructuring and infrastructuring: nominal class struggle in ChiBemba*. Bloomington, IN: Indiana University Linguistics Club.
- Stucky, S. U. 1978. How a noun class system may be lost: evidence from Kituba. *Studies in Linguistic Sciences* 8: 216-33.
- Suzman, S. 1980. Acquisition of the noun class system in Zulu. *Papers and Reports on Child Language Development*. Stanford University. 19: 45-52.
1991. Language acquisition in Zulu. PhD dissertation, University of the Witwatersrand.
- Tsonope, J. 1987. The acquisition of Setswana noun class and agreement morphology, with special reference to demonstratives and possessives. PhD dissertation, SUNY Buffalo.
- Voorhoeve, J., and P. de Wolf (eds.), 1969. *Benue-Congo noun class systems*. Afrika Studiecentrum, University of Leiden.
- Watters, J. 1980. The Ejagam noun class system: Ekoid Bantu revisited. In L. Hyman (ed.), *Noun classes in the Grassfields Bantu Borderland*. SCOPIL vol. 8. University of Southern California. 99-137.
- Welmrs, W. 1973. *African language structures*. Berkeley: University of California Press.

## CHAPTER 9

*Gender assignment: a typology and a model*

Greville G. Corbett and Norman M. Fraser

*Department of Linguistics and International Studies, University of Surrey*

## I INTRODUCTION

Gender systems have agreement as their defining characteristic. This is argued by Dixon (1986) and by Corbett (1991) and is the generally accepted approach (there are major problems with alternative views). Given this defining feature, the term 'gender' covers what in some traditions is called 'noun class': for instance, in Dravidian linguistics it is normal to talk of 'genders', while those working on Caucasian languages, even those with systems of classification very similar to those of Dravidian, usually talk of 'noun classes'. Gender is not restricted to sex-based classifications ('male/female'): other semantic possibilities include 'animate', 'small', 'insect', 'non-flesh food' and so on.

We shall further develop our earlier claim that the assignment of nouns to a particular gender is always possible for the vast majority of nouns, from information which is in any case required in the lexical entry. The type of information accessed varies and, as we shall see, forms the basis for a typology of gender assignment (section 2). This much is no longer new. However, while in some languages the implicational relationship between the types of information is clear (and so our claim is easily confirmed), in others there is a complex overlapping of sub-regularities. In such cases there may be competing analyses which are difficult to evaluate, particularly when they are formulated in vague terms. We shall focus on two such less straightforward cases. We shall adopt the same principled approach to the two rather different languages and show how a formal approach can elucidate and indeed substantiate our claim. Thus besides offering a typology of a particular phenomenon we hope to demonstrate the value of a genuinely formal approach for typology. We shall concentrate here on principles, since the technical detail is provided elsewhere (Fraser and Corbett 1995; 1997).

## 2 A TYPOLOGY OF GENDER ASSIGNMENT

We are concerned with the native speaker's ability to allot nouns to genders. This is done on the basis of information stored as part of lexical entries. The type of information used leads us to a typology of assignment systems. The major distinction is between semantic systems (where only semantic information is required) and formal systems (where semantic information is supplemented by morphological and/or phonological information). In a sense, all assignment systems are semantic systems: purely formal systems are excluded by our typology. These would be systems in which formal properties of nouns were sufficient to predict gender (as demonstrated by agreement evidence), irrespective of meaning. We find instances of formal information allowing correct gender assignment for many nouns, but nevertheless gender systems always have a semantic core. While there are no systems in which formal information is alone sufficient, we do find systems in which semantic information is sufficient for gender assignment, and it is these which we call semantic systems.

2.1 *Semantic assignment*

A clear example of this type of gender assignment is found in Godoberi. This is 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 Daghestan. The data are from Kibrik (1996); a sketch of the language can be found in Gudava (1967). There are three genders, which we may label 'masculine', 'feminine' and 'neuter'. Caucasianists often prefer the term 'noun class' precisely because of the semantic transparency of these classes of nouns (when compared with Indo-European genders). Caucasianists also typically use numbers (I, II, III) instead of names. The distinction, as noted earlier, is one of tradition. The assignment of nouns to the three genders/noun classes is shown in table 9.1.

Table 9.1. *Gender assignment in Godoberi*

| criteria        | gender        | example | gloss  |
|-----------------|---------------|---------|--------|
| male rational   | masculine (I) | ima     | father |
| female rational | feminine (II) | ila     | mother |
| other           | neuter (III)  | hamaXi  | donkey |

This assignment system is simple and operates consistently. Gender can be predicted from the meaning of a noun without reference to its form. 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'. Note that, apart from terminology, there is little dispute about such systems. There is no obvious alternative analysis. It is worth noting too that such systems are found in various parts of the world; the Dravidian family provides several examples.

2.2 *Predominantly semantic assignment*

There are several languages with clear-cut semantic systems like that of Godoberi. But many languages have semantic assignment rules which appear to allow sets of exceptions. The numbers of exceptions may not be significant as a proportion of the nouns in the languages, but they cannot be merely dismissed. A fine example is Zande, a member of the Zande-subgroup of the Ubangian branch of Adamawa-Ubangian. (Adamawa-Ubangian is in turn a branch of Niger-Congo, which is the major part of the Niger-Kordofanian family.) Zande has over 700,000 speakers; about half a million living in Zaire, most of the rest in the Sudan and 25-30,000 in the Central African Republic (data are from Claudi 1985; see that work for details of original sources).

While gender is reflected primarily in the personal pronoun, agreement in gender is spreading to other sentence elements. Assignment to the four genders is as in table 9.2.

The first two genders are straightforward, comparable in their simplicity to those of Godoberi. Nouns denoting male humans are of masculine gender and, equally, nouns of masculine gender denote male humans. Feminines are similar. The one minor complication is that for small children the pronoun for the third 'animal' gender is used. (This

Table 9.2. *Gender assignment in Zande*

| criteria      | gender    | example | gloss |
|---------------|-----------|---------|-------|
| male human    | masculine | gude    | boy   |
| female human  | feminine  | dia     | wife  |
| other animate | animal    | nya     | beast |
| residue       | neuter    | mvua    | grass |

is a common pattern: small children are often treated grammatically as not being quite human.)

Besides the familiar human/non-human distinction, Zande also has an animate/inanimate distinction, and the two genders based on this distinction are more interesting for our purposes. Nouns are almost always assigned to them according to the rule given, but there are about eighty exceptions. These are all nouns denoting inanimates (which we would expect to be neuter) which are actually in the animal gender.

The exceptions include:

1. nouns denoting heavenly objects: *diwi* 'moon', *ake* 'star', *wangu* 'rainbow';
2. metal objects (many of which are round): *de* 'chisel', *funde* 'needle', *tongo* 'ring (for finger)', *ngbagido* 'wheel';
3. edible plants (including round ones): *abangbe* 'sweet potato', *baundu* 'earth or ground pea';
4. non-metallic objects (mainly round): *mbasa* 'whistle', *badupo* 'ball';
5. other: *agboro* 'foam', *ze* 'scar'.

Some exceptions might be explained by reference to the mythology of the Zande but in most cases we can say at present only that they are exceptions.

There seems again to be little question about the principle according to which such languages should be analysed: they have a semantic system, with some exceptions to the semantic rules. We should set the assignment rules as defaults, with various overrides, many at the level of individual lexical items. A more fine-grained approach, however, reveals that there is great variety and interest in such systems, which are found in many different parts of the world (for some examples, see Corbett 1991: 15–30). There is often considerable debate on the detail of the analysis of such systems: the assignment criteria may be formulated in slightly different ways, in an attempt to reflect the worldview of the speakers and at the same time to minimize the number of exceptions. Perhaps the best-known example of this type is Dyirbal, which after Dixon's classic account (1972: 44–7, 60–2, 306–12; 1982: 178–83) continues to attract attention and reanalysis, for instance by Mylne (1995).

In some cases it is possible to cover almost all of the cases which at first sight seem exceptions to semantic assignment rules. In rather different languages, however, increasing ingenuity in the formulation of such rules may lead us to miss the point of the system. Consider a

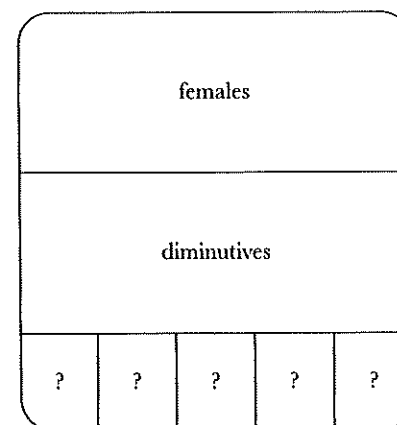


Figure 9.1. Hypothetical feminine gender

language, one of whose genders might be presented schematically as in figure 9.1.

To this gender, which we will call the feminine, are assigned nouns denoting females and those which have the affective meaning diminutive. If that were the complete story, and were typical of all the genders of the language, we would have a semantic system. However, in our imagined language there are many nouns which are not covered by these rules. We may construct ever more ingenious semantic rules, covering smaller and smaller groups of nouns. And in a given language this may be the right approach. In many languages, however, the 'exceptions' to the semantic rules are in fact subject to very general rules of a different type, based on a criterion depending on form (taking form in a broad sense here).

### 2.3 Formal systems A: morphological assignment

We now come therefore to languages in which large numbers of nouns fall outside the semantic assignment rules (there will always be semantic assignment rules, of course, since no language has a purely formal assignment system). Those nouns whose meaning does not allow gender assignment can be handled instead by rules which depend on the form of the nouns. These rules are in turn of two types, morphological (section 2.3) and phonological (section 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 clear-cut. 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 access to word structure, typically to inflectional class information. (Since these two types of information are often interrelated, establishing the type of rule involved in a particular language may require detailed analysis.)

For a clear example of a morphological assignment system we will consider Russian. A member of East Slavonic, Russian has three genders. Two of them, masculine and feminine, have a semantic core, as can be seen from the semantic assignment rules:

*Semantic assignment rules*

1. Sex-differentiable nouns denoting males (humans and higher animals) are masculine: *syn* 'son', *djadja* 'uncle', *lev* 'lion';
2. Sex-differentiable nouns denoting females are feminine: *doč'* 'daughter', *teja* 'aunt', *l'vica* 'lioness'.

Nouns which are sex-differentiable in Russian are those where the sex of the denotatum matters to humans; this is clearly the case with nouns denoting humans and domesticated animals. It also includes instances where the difference is visually striking (as in the case of lions).

These semantic assignment rules allow extremely few exceptions. The problem is that the majority of nouns remain unaccounted for. It is the case that further semantic regularities can be established, some quite exotic.<sup>2</sup> However, more and more rules are required, covering fewer and fewer nouns and with substantial numbers of exceptions (the situation envisaged in figure 9.1). The difficulty may be seen if we consider the triples in table 9.3, chosen for their semantic comparability.

It is most unlikely that we could find semantic factors to account for the gender of these nouns. Nevertheless, gender in Russian is highly predictable; for many nouns it is determined by formal rather than semantic factors, namely by the declensional type of the noun, as we shall see. The gender of the nouns in the table is predictable, given their declensional class (which is of course information which the native speaker must have access to). From some of the examples given already, it might appear that simple phonological rules would be sufficient: for example, nouns ending in *-o* are neuter. Unfortunately, there are examples for which no such rule works, pairs such as *portfel'* (masculine)

Table 9.3. *Triples of semantically similar nouns in Russian*

| masculine          | feminine                  | neuter              |
|--------------------|---------------------------|---------------------|
| <b>zakon</b> 'law' | norma 'standard'          | pravilo 'rule'      |
| žurnal 'magazine'  | <b>gazeta</b> 'newspaper' | pis'mo 'letter'     |
| nerv 'nerve'       | <b>kost'</b> 'bone'       | serdce 'heart'      |
| čaj 'tea'          | voda 'water'              | <b>vin</b> o 'wine' |
| avtomobil' 'car'   | mašina 'car'              | taksi 'taxi'        |
| večer 'evening'    | noč' 'night'              | utro 'morning'      |
| dom 'house'        | izba 'hut'                | zdanie 'building'   |
| glaz 'eye'         | ruka 'hand'               | uxo 'ear'           |
| lokot' 'elbow'     | lodyžka 'ankle'           | koleno 'knee'       |
| nož 'knife'        | vilka 'fork'              | bljudo 'dish'       |

*Note:* Forms here are transliterated: ' transliterates the Russian soft sign, which normally indicates palatalization of the preceding consonant. Items in bold reappear (in phonemic transcription) in table 9.4.

'briefcase' versus *pyl'* (feminine) 'dust', *lokot'* (masculine) 'elbow' versus *kost'* (feminine) 'bone'.

The assignment rules require access to information on the noun's inflectional behaviour. Since Russian will be considered in detail below, we will not justify this claim here. We should note that such systems are common in Indo-European languages and are also widespread in Bantu languages.

Why are such systems sufficiently problematic to require the detailed attention we propose? Analytical problems can arise when the number of declensional classes and genders is similar: it is not self-evident which should be predicted from which. The argument depends in part on the number of declensional classes and here there is not the clear-cut criterion which is available for the number of genders (namely agreement). Certainly in accounts of Russian, the number of declensional classes posited depends more on tradition than on argument. As hinted at above, the declensional class of a noun is often related to its phonology and so there is an additional set of possible interactions. The different models can and should be established by argument as is normal in linguistics. We have taken the additional step of developing a computer implementation of the account we propose, as will be described below, showing that it does indeed give the correct predictions. Other proposed accounts ought to be similarly validated in order to be admitted as viable alternatives.

## 2.4 Formal systems B: phonological assignment

Let us now consider phonological systems, where a single form is sufficient to establish the gender of a noun. These too are found in different families around the world; a striking example is Qafar (Afar), an East Cushitic language with about 250,000 speakers in northeastern Ethiopia and in Djibouti. The data are from Parker and Hayward (1985, especially p. 225), and Dick Hayward (p.c.). Like other Cushitic languages, Qafar has two genders. Nouns denoting male humans and the males of sexually differentiable animals are masculine, as in the case of *bàqla* 'husband'. Nouns denoting females (human and animal) are feminine: *barrà* 'woman, wife'. So much for the semantic assignment rules. The phonological rules are of greater interest:

*Phonological assignment*

1. Nouns whose citation form ends in an accented vowel are feminine: *catò* 'help', *kammà* 'autumn'. The accent position is indicated with ` . This is the potential site for high tone.
2. Others are masculine:
  - (a) those ending in a consonant: *cedèr* 'supper time', *gilàl* 'winter';
  - (b) those with a citation form ending in a vowel but with non-final accent: *tàmu* 'taste', *baànta* 'trumpet'.

Exceptions to these phonological rules are very few; one is *doònik* 'sail-boat', which 'ought' to be masculine but is exceptionally feminine.

When the two types of rules conflict then the semantic rule dominates. Thus *abbà* 'father' 'should' be masculine because of its meaning, and 'should' be feminine because it ends in an accented vowel; it is masculine. And *gabbixeèra* 'slender-waisted female' is feminine because of its meaning, even though the accent is not on the final vowel, and so shows a masculine pattern. Qafar demonstrates clearly why the prediction must be from phonological form to gender and not vice versa. The point is that any of the five vowels of Qafar (*a, e, i, o, u*) may carry the accent in word-final position on nouns. We can predict the gender from the final segment, but we could not predict the final segment from the gender, nor fully predict the position of the accent.<sup>3</sup>

## 3 NETWORK MORPHOLOGY

We now turn to a formally explicit framework for expressing these generalizations: Network Morphology. In this framework, generalizations

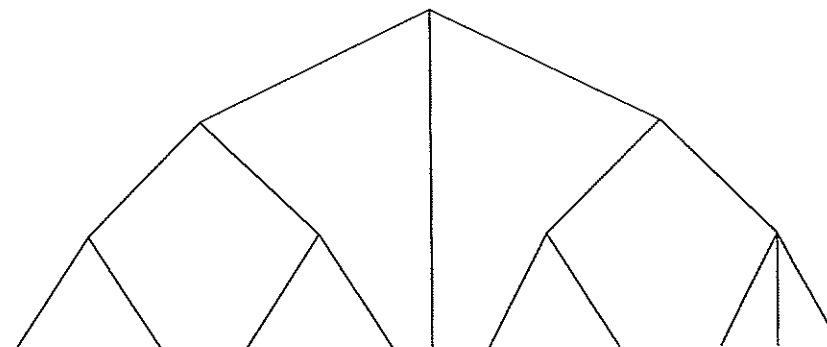


Figure 9.2. A water course diagram

are layered in such a way as to produce extremely parsimonious accounts of complex data. A particular strength of Network Morphology is that it provides a practical tool for stating explicit accounts of different grammatical phenomena (such as the categories of case, number, gender and animacy) and for describing the interactions between the different systems.

We begin with an analogy. Linguists are familiar with tree diagrams such as the one shown in figure 9.2. For present purposes, the analogy of a tree is inappropriate; it is more helpful to think in terms of a water course. The top of the diagram is the origin of the water course, rather like a spring in the mountains. Water flows down the mountain sides and, as it goes, it divides into a number of different branches, all sharing the same origin.

In our model, it is information which flows downwards from top to bottom of the hierarchy. If some linguistic generalization becomes available at the top of the hierarchy, its effects are felt in the branches of the hierarchy below it. Generalizations can be introduced at any point in the structure, but they will only become available in the portion of the structure located by direct lines of descent beneath the point of introduction. Just as water cannot flow upwards from a spring, so generalizations cannot flow upwards in an information structure of this type.<sup>4</sup>

It is common to find encyclopedic knowledge organized in structures of this kind. This is particularly clear in the case of formal taxonomies of the variety used by natural scientists. The Linnean classification arranges living organisms in a hierarchy whose levels are shown in figure 9.3. Generalizations stated at a higher level trickle down to the

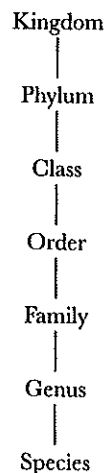


Figure 9.3. Linnean classificatory levels

lower levels. Thus, a fact which is true for a given class is likely to be true of most species within the class. Of course, there are exceptions: whereas the members of the class of mammals in general produce embryos which are connected to their mother by a placenta, those of the order of marsupials do not. In a taxonomic account of real-world data, the exceptional fact about marsupials overrides the generalization about mammals, and blocks it from flowing any further down the structure. This exceptional fact about marsupials itself becomes a sub-generalization which trickles down to the levels below, so members of the order (including kangaroos, koalas, tasmanian devils and wombats) all share this same feature.

Linguistic entities can be organized in a similar hierarchical structure, with high-level generalizations trickling downwards through the structure by default, unless blocked by exceptional information. Consider, for instance, the data on Russian nouns presented in table 9.4. Certain facts hold true of all the nouns presented (and each represents a large number of similar nouns); for instance, the dative plural consists of stem plus *-am*. There are facts which are true of classes I and IV but not of the other two. There are facts which hold for three classes (locative singular consists of stem plus *-e*), which could be stated as a default of nouns and overridden for just one type. And then there are higher generalizations which hold of adjectives as well as nouns. For details of these claims see Corbett and Fraser (1993). On the basis of the

Table 9.4. Major noun paradigms of Russian

|                 | zakon<br>'law' | gazeta<br>'newspaper' | kost'<br>'bone' | v'ino<br>'wine' |
|-----------------|----------------|-----------------------|-----------------|-----------------|
| <i>Singular</i> |                |                       |                 |                 |
| nom(inative)    | zakon          | gazeta                | kost'           | v'ino           |
| acc(usative)    | zakon          | gazetu                | kost'           | v'ino           |
| gen(itive)      | zakona         | gazeti                | kost'i          | v'ina           |
| dat(ive)        | zakonu         | gazete                | kost'i          | v'inu           |
| inst(rumental)  | zakonom        | gazetoj               | kost'ju         | v'inom          |
| loc(ative)      | zakone         | gazete                | kost'i          | v'ine           |
| <i>Plural</i>   |                |                       |                 |                 |
| nom             | zakoni         | gazeti                | kost'i          | v'ina           |
| acc             | zakoni         | gazeti                | kost'i          | v'ina           |
| gen             | zakonov        | gazet                 | kost'ej         | v'in            |
| dat             | zakonam        | gazetam               | kost'am         | v'inam          |
| inst            | zakonam'i      | gazetam'i             | kost'am'i       | v'inam'i        |
| loc             | zakonax        | gazetax               | kost'ax         | v'inax          |
|                 | I              | II                    | III             | IV              |

*Notes:*

- (i) forms are given here in phonemic transcription.<sup>5</sup> Palatalization (or 'softening') is indicated by '.
- (ii) there is no overt ending in the nominative/accusative singular in types I and III, nor in the genitive plural of types II and IV.

evidence discussed there, we suggest a hierarchical organization to describe Russian nominals in Network Morphology (figure 9.4).

As noted above, there are useful generalizations to be made at the level of all nominals. For example, the dative, instrumental and locative plural forms of nouns and adjectives share some properties. Generalizations which apply to all or most nouns but not adjectives are made at the NOUN node. These include generalizations about the locative singular and nominative and genitive plural forms. N\_I to N\_IV are nodes at which information specific to declensional classes is located. N\_O is a node which we have postulated based on the observation that there are enough generalizations to be made about classes I and IV (but not II and III) to wish to treat them together, but not enough to argue that the classes are indistinguishable.

Approaches which organize information in terms of hierarchical structures allowing information to flow downwards unless blocked by more specific facts instantiate what is generally known as *default inheritance*. Such approaches to knowledge representation have been studied

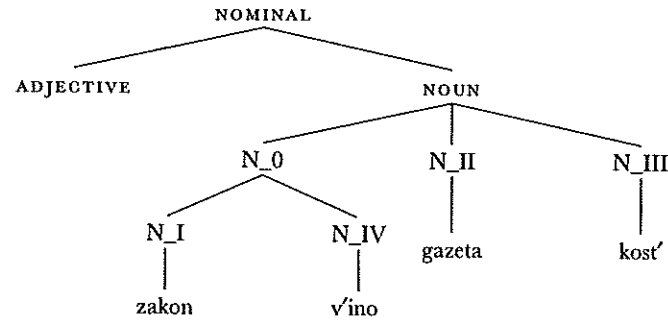


Figure 9.4. An inheritance structure for Russian nominals

in the field of artificial intelligence for many years, so there is a valuable body of knowledge for linguists to draw on. There are also some practical tools which can greatly simplify the task of developing a linguistic theory based on default inheritance. We chose to use an existing default inheritance formalism called DATR to encode Network Morphology theories. The DATR knowledge representation language was developed by Roger Evans and Gerald Gazdar as a logic-based formalism for describing inheritance networks (Evans and Gazdar 1989a,b; 1996). Since computer interpreters for the DATR language exist, it is possible to check that a Network Morphology analysis expressed in DATR captures the intended generalizations.

A fragment of the theory illustrated in figure 9.4 is shown in (1).

- (1) NOUN:  
 <mor loc sg> == "<stem>"\_e  
 <mor nom pl> == "<stem>"\_i.  
 N\_III:  
 <> == NOUN  
 <mor loc sg> == "<mor dat sg>".  
 Kost':  
 <> == N\_III  
 <stem> == kost'.

In this formalism, the elements to the left of the colon are the names of nodes in the inheritance hierarchy at which generalizations (which appear after the colons) are stored. In DATR sentences, anything enclosed between paired angle brackets is a *path*. Paths have an internal ordering, such that any fact which is stated relative to a short path such as <mor> (indicating that the path describes morphological information)

may also be taken to apply to any extension of it, such as <mor gen> or <mor gen sg>. The logical conclusion of this is that the shortest possible path (the empty path <>) is a descriptor which can be taken to describe any path in the theory. The first sentence at NOUN should be read as saying that the locative singular consists of the stem followed by an *-e* ending. A path enclosed in double quotes (a *quoted path*) in a DATR sentence is used to refer to the value described by that path when globally evaluated, that is, when evaluated as from the original query node.

If we wanted to find the nominative plural of *Kost'*, we would inherit the sentence <mor nom pl> == "<stem>"\_i. Before going any further we would have to find out what the <stem> of *Kost'* is. Since the answer is *kost'*, the nominative plural is *kost'i*. If we wanted to know the locative singular of *Kost'*, we would never inherit the definition of locative singular at NOUN because it is overridden at N\_III, from which *Kost'* inherits more immediately. The definition of locative singular at N\_III establishes an identity between the locative singular form of an N\_III noun and its dative singular.

It may be desirable to inherit most information from one source, but to have access to certain kinds of information stored elsewhere. DATR supports this kind of *multiple inheritance*. One consequence of this possibility is that, in practice, inheritance structures are seldom pure hierarchies; rather, they are complex networks of interconnecting generalizations. (In the language of our water course metaphor, it is not just that rivers may branch to send parts of the flow off to different destinations, but tributaries may also converge to bring separate flows together.) Consider the following extract:

- (2) N\_II:  
 <> == NOUN  
 <mor gen sg> == "<stem>"\_i.  
 N\_III:  
 <> == NOUN  
 <mor gen sg> == N\_II.

This DATR fragment says that N\_III may inherit its schema for forming the genitive singular from N\_II, even though N\_III (like N\_II) inherits primarily from NOUN. This may be expressed more explicitly as follows:

- (3) N\_III:  
 <mor gen sg> == N\_II:<mor gen sg>.

Network Morphology theories organize information in a number of distinct inheritance structures, each of which has its own root. These include among others a lexemic hierarchy (or word-class hierarchy), an inflectional hierarchy (for declensional information) and a stress hierarchy (Brown, Corbett, Fraser, Hippisley and Timberlake 1996; Brown 1998). Information flow between these structures is highly constrained. In deriving a particular lexical form, the appropriate information must be accessed from each of the relevant structures by means of the permitted network relations, and combined to derive the correct result.

#### 4 GENDER ASSIGNMENT IN RUSSIAN

The interactions of semantic and formal information in gender assignment can be complex, and have often been characterized in vague terms. We will therefore outline an explicit Network Morphology account of the interactions of semantics, gender, declensional class and phonology in Russian, again using the lexical knowledge representation language DATR.

We have established already that Russian has semantic assignment rules of a rather standard type, that these fail to cover many nouns, and that these remaining nouns cannot be accounted for by further, more elaborate semantic rules. The remaining nouns are distributed over the three genders and we claim that their distribution is accounted for by 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.

The declensional classes are as in table 9.4. Thus *zakon* 'law' belongs to declensional class I and is masculine, *gazeta* 'newspaper' (class II) and *kost'* 'bone' (class III) are feminine, and *v'ino* 'wine' (class IV) is neuter.

Two questions arise immediately: the interaction of these rules with the semantic assignment rules, and the justification for postulating four major declensional classes (on which the success of the morphological assignment rules depends).

Consider first the interaction with the semantic assignment rules. It is the case that many nouns which denote males are members of declensional class I. We might therefore consider the semantic assignment

rules superfluous, and assign such nouns to the masculine gender by virtue of their morphology. However, the two sets of rules can make conflicting assignments, and then gender is assigned by the semantic rules. This is clear with nouns like *mužčina* 'man', which 'ought' to be masculine according to the semantics (male), but feminine according to the morphology (declensional class II). In fact, such nouns are of masculine gender.<sup>6</sup> Thus both sets of rules are required, and the semantic rules dominate. The fact that there is overlap between them is a point to which we shall return. Let us at this point formalize the analysis so far, before returning to the question of the number of declensional classes.

Every noun inherits from the NOUN node in our Network Morphology account. Since the following path equation is specified at NOUN, every noun inherits it unless it is overridden by more specific information given in the lexical entry:

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

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

(5) GENDER:  
 <male> == masc  
 <female> == fem  
 <undifferentiated> == "<mor formal\_gender>".

The interpretation of this fragment is straightforward in the case of sex-differentiated nouns. If the referent of the noun is male then the gender is masculine; if the referent is female then the gender is feminine.<sup>7</sup> This is simply the reflection of the semantic assignment rules. However, if the sex is undifferentiated, it is necessary to consider additional (formal) criteria, and this is done by evaluating the path <mor formal\_gender>. Formal gender is defined for each of the main declensional classes. Thus, for example, the following equation is located at the node for declensional class II nouns, N\_II:

(6) N\_II:  
 <formal\_gender> == fem

This equation is used in the evaluation of the path <mor formal\_gender>. Thus, 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 with a non-sex-differentiated denotatum (such as *gazeta* 'newspaper') will have feminine gender, the default gender for class II nouns. This is the way in which the morphological assignment rules are reflected in our account. Note that (5) above also takes care of the relationship between the two types of assignment.

There are a few nouns which belong to smaller inflectional classes (such as *znamja* 'banner'); these do not raise new issues of principle. But there are also nouns which do not decline, and so are outside the scope of the morphological assignment rules above. However, their lack of declension is itself a matter of morphology; 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, there is an interaction with the semantic feature of animacy: they are masculine if animate and neuter if not.<sup>8</sup> Some equations from the node N\_V, from which all class V nouns inherit, are given below.

- (7) N\_V:  
 <formal\_gender> == <"<sem animacy>">  
 <animate> == masc  
 <inanimate> == neut.

To assign gender for a class V noun not covered by the usual semantic assignment rules (and few are), 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 referent and neuter gender if the referent is inanimate.

The account so far is, perhaps surprisingly, sufficient to account 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 demonstrate that the correct predictions are indeed made.<sup>9</sup> Russian also has a subgender of animacy. This involves nouns of all genders; it is not restricted to indeclinables, as is the case for assignment to gender based on animacy. It is reflected in agreement, but involves the accusative case only. There are no exclusively animate accusative forms in Russian; instead, animacy shows itself by different patterns of syncretism. Animacy is less complex than gender, and raises no new questions of principle, so we shall not discuss it further here. With appropriate use of defaults, animacy need hardly ever be specified in lexical entries and our implemented version demonstrates that we make the right predictions (Fraser and Corbett 1995: 130–2).

We now return to the question of the number of declensional classes. We have stated that there are four main classes, so far without argument. There are two sets of interwoven problems here: first there are conflicting views as to the number of classes, and second there are attempts to run the predictions in the other direction, that is, to predict declensional class given the gender.<sup>10</sup> Traditional accounts of Russian typically treat our classes I and IV as a single class, for historical reasons; to predict the appropriate inflections within these classes the gender must then be known. Unfortunately there is little discussion of the criterion for treating instances such as the paradigms we have labelled I and IV as one class or two. It is clear, however, that if they are treated as one, then it is not possible to assign nouns to gender on the basis of their inflectional class. But equally it is not possible to assign nouns to declensional class on the basis of their gender since, given a stem in a soft consonant of feminine gender, we could not predict whether it was of declensional class II or III (and there are thousands of such nouns). These points are discussed in detail in Corbett (1982). Since it is possible to assign gender on the basis of declensional class, provided I and IV are recognized as distinct, we take this as an argument that they should indeed be recognized as distinct." Furthermore, since this allows languages like Russian to fit into the general typology, according to which gender is always predictable for the vast majority of nouns, this is also an argument in favour of our approach.

Nevertheless, as we noted earlier, there are correspondences between meaning and declensional class; we shall see that as a result we can simplify a substantial number of lexical entries. For many nouns, 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, *monax* 'monk', *byk* 'bull';
2. Sex-differentiable nouns denoting females are of declensional class II: for example, *žena* 'wife', *l'vica* 'lionness'.

There is a significant number of nouns whose declensional class must be specified to override rule 1. These are nouns like *mužčīna* 'man', which denote males but which decline according to declensional class II. (Lazova 1974: 942–3 puts the figure at 273, but the actual number is larger because there are many hypocoristics of this type, like *Saša*

'Sasha', which her dictionary naturally does not include.) 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 (instead of declensional class II). Attempting to assign nouns to declensional class according to phonology is more difficult.<sup>12</sup> Inflecting nouns have a stem ending in a consonant, which may be hard or soft. There is a restriction in that if the stem ends in a hard consonant, then the noun cannot be of declensional class III. But since class III is the smallest of the major classes this restriction is of little predictive value. However, the requirement that the stem end in a consonant is useful as the basis of the following rule:

*Formal assignment rule for declensional class*

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

This rule follows Worth (1966); a noun like *taksi* 'taxi' has the stem *taksi* entered in the lexicon and this guarantees its indeclinability.

These declensional class assignment rules can now be formalized. Consider the following DATR fragment (part of a more sophisticated account than that suggested in the examples used for illustration in (1)–(3) above), which is located at the NOUN node:

```
(8) NOUN:
    <mor> == "<declensional_class>"
    <declensional_class> ==
    DECLENSION:<"<infl_root final>" "<sem sex>">
```

To find a value (or values) for the <mor> path (or paths), it is necessary to evaluate the <declensional\_class> path. The second equation, which defines this path, is quite 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 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. In a more complete account, this information would be supplied by a phonological component (which could also be formalized using DATR, as the work of Reinhard and Gibbon (1991) and Gibbon (1992) demonstrates), though for convenience here we simulate the phonological component crudely by means of a simple equation:

```
(9) NOUN:
    <infl_root final> == consonant
```

This holds for all native Russian nouns; it is overridden in the lexical entries of some nouns, typically of relatively recent foreign origin.

Values for the <sem sex> path are supplied either in the lexical entries or by means of a default assignment of undifferentiated sex at the NOUN node, as shown in (10).

```
(10) NOUN:
    <sem sex> == undifferentiated
```

Thus, where declensional class is predictable, it is found by evaluating a path at the DECLENSION node, which 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 shown below:

```
(11) DECLENSION:
    <consonant male> == N_I:<>
    <consonant female> == N_II:<>
    <vowel $sex> == N_V:<>.
```

Each equation in fragment (11) includes a type of right-hand side which we have not yet described. Suppose a DATR theory includes the following equation:  $\text{NODE1:Path1} == \text{NODE2:<>}$ . An attempt to evaluate  $\text{NODE1:SomePath}$  can succeed and return a value if and only if  $\text{SomePath}$  can be formed by concatenating  $\text{Path1}$  with a path defined at  $\text{NODE2}$ . In effect, the  $\text{NODE:<>}$  notation allows the description of a path to be distributed over more than one node.

In fragment (11), the first two paths encode the semantic assignment rules for declensional class. The first attribute in each path is consonant so these paths potentially apply to the bulk of Russian nouns. The second attributes in the paths narrow down on two subsets of these. The first path assigns core nouns (typical native nouns) denoting males to class I. It does this by evaluating at  $\text{N\_I}$  the remainder of the original query path (as left after the first equation of (8) has accounted for the initial *mor*), using the  $\text{N\_I:<>}$  notation introduced above.

The second equation in (11) assigns core nouns denoting females to class II in similar fashion. The third equation picks out those nouns whose inflectional root ends in a vowel and assigns them to class 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 referents. Such nouns must typically specify declensional class in their lexical entries. Let us check the working of (8)–(11). The noun *monax* ‘monk’ has this lexical entry:

```
(12) Monax:
      <> == NOUN
      <gloss> == monk
      <infl_root all> == monax
      <sem sex> == male.
```

Since the *Monax* node inherits directly from the *NOUN* node, the two equations shown in (8) are available to it. The first of these can be paraphrased as saying ‘for paths beginning <mor . . .> evaluate the path <declensional\_class>’. The definition of that path (also given in (8)) requires the evaluation at the *DECLENSION* node of a path consisting of the value of <infl\_root final> for *monax*, followed by the value of <sem sex> for the same word. The value of <infl\_root final> is *consonant* according to the unoverridden equation (9), and the value of <sem sex> for *Monax* is *male* according to the last equation in the lexical entry (shown in (12)). Thus the path <consonant male> must be evaluated at the *DECLENSION* node. According to the first equation in (11) this returns a value of *N\_I : <>* for the path <mor>, where our search originated.

It is important to note that this analysis has no way of assigning nouns to declensional classes in the abstract. It can only do so in the course of solving some more specific query, such as retrieving some particular form of a noun. This reflects the fact that, unlike gender (which is required in syntax for agreement), information about declensional class is not required outside the morphological component. Suppose, for example, that the theory is queried concerning the nominative singular form of *monax*. This translates into a query to find the value of *Monax : <mor nom sg>*; we have just discussed paths beginning <mor . . .> and the route to the *N\_I* node. Moving on now, the relevant parts of the *DATR* theory for *N\_I* are shown in (13).

```
(13) N_I:
      <> == N_O
      <formal_gender> == masc
      <nom sg> == "<stem sg>"
      <hard gen pl> == "<stem pl>_ov."
```

Evaluating the path <mor nom sg> at the *Monax* node leads via *NOUN* to the following inference (in the sequence of steps described above):

```
(14) Monax:
      <mor> == N_I : <>
```

This leads the <mor> path from *NOUN* to be concatenated with the <nom sg> path from *N\_I* in (13) to produce the following inference:

```
(15) Monax:
      <mor nom sg> == "<stem sg>"
```

In short, the nominative singular of *monax* consists of the bare stem:

```
(16) Monax:
      <mor nom sg> = monax
```

Declensional class assignment can be compared with gender assignment to reveal significant differences. One of these is the outcome when semantic and formal rules make different predictions. For gender assignment, quite generally, it is the semantic rule which dominates. As we saw 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 and we predict more generally, the formal factor dominates. Thus *attaše* ‘attaché’ when it denotes a male ‘should’ be in declensional class I, but it ends in a vowel and so ‘should’ be in declensional class V. It is actually in declensional class V (the formal rule dominates). Conversely when it denotes a male it ‘should’ be masculine but it is in class V and so ‘should’ be neuter. It is actually masculine (the semantic rule dominates). These points are summarized in table 9.5.

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.

Thus, in our account, declensional class can be predicted for a substantial proportion of Russian nouns, and declensional class assignment differs from gender assignment in interesting ways. More importantly, we have justified an account assigning gender to nouns in part on the basis of their morphology. We have referred to earlier argument for this position based on language-internal reasons. We also argue that it allows languages of this type to fit into a coherent typology. And

Table 9.5. *Differing assignments for attaše '(male) attaché'*

|                               | Semantic prediction         | Formal prediction          | Actual    | Conclusion                   |
|-------------------------------|-----------------------------|----------------------------|-----------|------------------------------|
| Declensional class assignment | denotes male<br>→ class I   | stem in vowel<br>→ class V | class V   | formal criterion dominates   |
| Gender assignment             | denotes male<br>→ masculine | class V →<br>neuter        | masculine | semantic criterion dominates |

finally we have implemented our analysis, in a way which can be seen to make the right predictions as to the gender of Russian nouns (and also as to their declensional class where possible).

#### 5 GENDER ASSIGNMENT IN ARAPESH

Our second case study is Arapesh, a language of the Toricelli family, spoken on the north coast of Papua New Guinea, between Dagur and Matapau. The data are drawn from Fortune's (1942) grammar and recent work based on it by Aronoff (1992, 1994: 89–114). The type of the gender assignment system is not obvious. Our purpose is to show how Arapesh fits into the proposed typology, and to indicate that our analysis, which follows that of Aronoff to a considerable extent, again gives the correct predictions, as our implementation demonstrates (see Fraser and Corbett 1997).<sup>13</sup>

At first sight it might appear that Arapesh is a language with a phonological assignment system; in many instances the gender of a noun could be predicted from the phonology. However, on closer examination Arapesh turns out to be more similar to Russian: we need to refer to the notion of morphological class and it is from this that the gender of most nouns can be predicted.<sup>14</sup> Besides the initially attractive possibility of direct assignment from phonology, a second confusing factor is the fact that the morphological classes are minimal, having two cells only: singular and plural. (For this reason it might seem strange to call them 'declensional classes', maintaining the parallelism with Russian, though in principle the smaller number of cells is not significant.) Nevertheless, as we shall see, we find grounds for assignment of morphological class on the basis of phonology (with some lexical exceptions) and for the assignment of gender on the basis of morphological class (with some other lexical exceptions).

Table 9.6. *Arapesh morphological classes for nouns*

| Class | Alternation | Singular | Plural     | Gloss          |
|-------|-------------|----------|------------|----------------|
| 1     | by/bys      | agaby    | agabys     | back           |
| 2     | bør/ryb     | ñibør    | ñiryb      | belly          |
| 3     | ag/as       | ajag     | ajas       | leg            |
| 4     | g/gas       | aweg     | awegas     | seed           |
| 5     | kü/meb      | ilokü    | ilameb     | a bird         |
| 6     | kü/rib      | yahakü   | yaharib    | a fruit tree   |
| 7     | kü/ib       | unukü    | unib       | teeth mother   |
| 8     | kü/guhijer  | anikü    | aniguhijer | rattan species |
| 9     | kü/ijer     | barahokü | barahijer  | grand-daughter |
| 10    | kü/u        | amagokü  | amagou     | fly            |
| 11    | m/ipj       | irum     | iripj      | breadfruit     |
| 12    | n/b         | narun    | narøb      | wave           |
| 13    | n/m         | araman   | aramum     | man            |
| 14    | in/ij       | kobif    | kobij      | ditch          |
| 15    | V/Vhas      | bode     | bodehas    | stone axe      |
| 16    | pü/gwis     | barupü   | barugwis   | mountain track |
| 17    | pü/s        | apapü    | apas       | banana         |
| 18    | r/guh       | jur      | jguh       | snake          |
| 19    | t/togu      | alit     | alitogu    | shelf          |
| 20    | t/gü        | nybat    | nybagü     | dog            |
| 21    | uh/ruh      | nauh     | naruh      | tooth          |
| 22    | ah/eh       | atah     | ateh       | ear            |

It is possible to identify at least twenty-two morphological classes on the basis of noun stem phonology. Table 9.6 shows the singular/plural phonological alternations characteristic for each of the morphological classes.<sup>15</sup> We follow Aronoff (1992: 31–2) in analysing the singular forms as bearing no morphological marker, and the plural forms as morphologically complex. Aronoff argues (assuming a process framework) that the plurals result from the action of realization rules which vary according to the morphological class. We are not concerned here with explicating the rules for relating plural forms to stems.

While most nouns can be assigned to one of the morphological classes on the basis of the phonology of their stem alone, a small number of nouns fall outside the system outlined. For instance, there are a few noun stems ending in *-b*, *-k* and *-s*, though no morphological assignments are defined for these. According to Fortune, only two nouns in the language end in *-b* (*kwagesab* 'croton' and *mib* 'thigh'), and likewise only two end in *-k* (*bokok* 'cannibalistic ogre' and *ñibiok* 'sacred flute'). In such cases, the nouns fall into the default morphological class. Nouns

belonging to the default class all bear the plural marker *-ehas* (which is also the plural marker of regular class 15 nouns). There are also some genuine exceptions to the morphological class assignment rules – nouns whose final segment ought to lead them to be assigned to a class by one of the standard assignment rules but which are actually in a different class. These exceptional cases fail to be assigned in the normal fashion, instead being assigned to the default morphological class and receiving the *-ehas* plural marker. An example of this kind is *lim* ‘roller for launching canoe’, which should be assigned to morphological class 11 and take the plural marker *-ipj*. Instead, it bears the default class plural marker, and is realized as *limehas*.

A very few items are yet more exceptional, not fitting into either the normal or the default assignment systems. Fortune (1942: 14) lists five nouns whose stems end in *-n* and which ought therefore to have their plurals in *-b* (class 12). Instead, these mark plurality with class 1 *-bys*. So, for example, *awhon/awhobys* ‘eel’. Nevertheless it remains the case that most nouns can be assigned to their morphological class on the basis of their phonology (and certainly to a much greater degree than appears to be the case with Russian). The formalization of the assignment of morphological class is conceptually relatively straightforward, though our implementation finesses the phonological detail (see Fraser and Corbett 1997: 45ff. for the DATR implementation and output). It is gender assignment which we shall concentrate on here.

There are thirteen genders in Arapesh, as demonstrated by the agreement of adjectives, verbs and prowords (Aronoff 1992: 22–6). Some of the genders are coextensive with morphological classes (1, 2, 11, 12, 13, 18, 21). In the case of four genders, two morphological classes correspond to a single gender (3 and 4, 14 and 15, 16 and 17, 19 and 20). One gender has no less than six corresponding morphological classes (5, 6, 7, 8, 9, 10). Correspondences between morphological class and gender are shown in table 9.7.

For the vast majority of nouns in Arapesh, there is an implicational relationship between stem phonology and morphological class; similarly there is an implicational relationship between morphological class and gender, as shown below:<sup>16</sup>

Stem Phonology => Morphological Class => Gender

Once again some nouns fall outside the normal assignment system and end up being assigned to the default, and the default gender is gender VIII. Notice that there appears to be a non-arbitrary connection

Table 9.7. Correspondences between morphological classes and genders in Arapesh

| Morphological class | Gender |
|---------------------|--------|
| 1                   | I      |
| 2                   | II     |
| 3                   | III    |
| 4                   |        |
| 5                   |        |
| 6                   |        |
| 7                   | IV     |
| 8                   |        |
| 9                   |        |
| 10                  |        |
| 11                  | V      |
| 12                  | VI     |
| 13                  | VII    |
| 14                  | VIII   |
| 15                  |        |
| 16                  | IX     |
| 17                  |        |
| 18                  | X      |
| 19                  | XI     |
| 20                  |        |
| 21                  | XII    |
| 22                  | XIII   |

between the default status of gender VIII and the fact that nouns in the default morphological class take their plurals in *-ehas*, as do normally assigned nouns in morphological class 15.

Nouns which have been assigned to the default morphological class are automatically assigned the default gender. Thus nouns with stems ending in rare segments like *-b*, *-k* and *-s* are assigned to the default morphological class and, as a result, they are then assigned default gender. In similar fashion, nouns which are lexically marked to take a normal singular but an exceptional plural (such as *lim* ‘roller for launching canoe’) are assigned to the default morphological class and receive default gender. However, membership of the default morphological class is not a prerequisite for assignment of default gender. Some nouns belonging to a normal class morphologically are lexically marked to receive the

default gender. For example, the noun *diliat* 'side post which supports eaves of house', plural *diliatogy*, is a normal member of morphological class 19, but instead of being assigned to gender XI, it receives default gender. Such examples of this kind demonstrate that the direction of implication runs from phonology to morphological class to gender. Membership of the default morphological class implies default gender, but default gender does not necessarily imply membership of the default morphological class.

We have concentrated on the phonology and morphology as being of greatest interest, but we should not forget our earlier claim that there are no purely formal assignment systems. This is borne out by Arapesh, which has semantic assignment rules, which dominate the formal assignment rules. Most nouns denoting female persons are assigned to gender IV, while all and only the nouns denoting male persons and the exclusively male roles of warfare and men's initiation ceremonies are assigned to gender VII. Nouns which denote persons without differentiation of sex are assigned to the default gender.

Let us consider how this analysis can be formalized in a Network Morphology framework. Most of the function of gender assignment is served by three nodes, which we label GENDER\_ASSIGN, GENDER\_FROM\_SEX, and GENDER\_FROM\_MOR. GENDER\_ASSIGN is used to check whether a lexical item denotes a person or not. If it does, gender assignment is based purely on semantics, specifically, sex. If it does not, gender assignment is based purely on formal criteria, namely morphological class. These nodes are shown below.<sup>17</sup>

```
(17) GENDER_ASSIGN:
      <> == VIII
      <person> == GENDER_FROM_SEX:<"<sem sex>">
      <non_person> == GENDER_FROM_MOR:<"<mor class>">.
```

This node shows the relationship between the two types of assignment rule in Arapesh. Personhood (a semantic category) is checked before anything else, thus reflecting the general situation in which semantic assignment of gender dominates morphological assignment. (Personhood is not mentioned in many lexical entries, since all nouns inherit from the NOUN node, and there it is stated that the default personhood for a noun is 'non-person'.) In the event that some value other than 'person' or 'non-person' is checked, there is a default assignment to gender VIII, the default gender.

Node (18) formalizes the semantic assignment rules.

```
(18) GENDER_FROM_SEX:
      <> == GENDER_ASSIGN
      <female> == IV
      <male> == VII.
```

Any sex values looked up at this node other than female or male (for example, undefined) must be inherited from GENDER\_ASSIGN, where the catch-all default will ensure assignment to gender VIII.

The following node represents the rules for morphological assignment to gender:

```
(19) GENDER_FROM_MOR:
      <> == GENDER_ASSIGN
      <1> == I
      <2> == II
      <3> == III
      <4> == III
      <5> == IV
      <6> == IV
      <7> == IV
      <8> == IV
      <9> == IV
      <10> == IV
      <11> == V
      <12> == VI
      <13> == VII
      <16> == IX
      <17> == IX
      <18> == X
      <19> == XI
      <20> == XI
      <21> == XII
      <22> == XIII
```

In the event that any value other than a recognized morphological class label is checked at this node, default inheritance ensures that the default assignment to class VIII is inherited from GENDER\_ASSIGN.

Thus for *aijag* 'leg' personhood is 'non-person'. In order to find a value for the path <non\_person> at node GENDER\_ASSIGN, it is necessary to evaluate <mor class> at the node GENDER\_FROM\_MOR. The morphological class of *aijag* is 3, as the morphological class assignment rules will establish on the basis of the phonology. Evaluating <3> at the node GENDER\_FROM\_MOR yields the result III. We have established that *aijag* is assigned to gender III on the basis of reasoning

from the phonology of the stem, to the morphological class, to the gender.

Now consider, a lexical entry denoting a person, such as *barahoku/ barahoguhijer* 'grand-daughter' (20).

(20) BarahokU:<sup>18</sup>  
 <> == NOUN  
 <gloss> == grand\_daughter  
 <sem personhood> == person  
 <sem sex> == female  
 <mor class> == 9  
 <phon stem> == barahokU  
 <phon pl stem> == barah.

The lexical entry specifies that <sem personhood> == person, and so we are faced with a different option at GENDER\_ASSIGN, namely the requirement to evaluate <sem sex> at GENDER\_FROM\_SEX. The lexical entry specifies that <sem sex> == female, so this information permits retrieval of the gender IV from the node GENDER\_FROM\_SEX. Now consider a noun which does not refer to a person, but belongs to the default class, as in the case of *bokok/bokokehas* 'cannibalistic ogre'. Here the important information ought to be found at the GENDER\_FROM\_MOR node (19). This node is a simple table of mappings from morphological classes to genders, very similar to that shown in table 9.7. However, there is no explicit mapping from 'default\_class' to anything. Instead, inheritance from GENDER\_ASSIGN yields a mapping from the maximally underspecified path '<>' to gender VIII, the default gender. In this way, any word which falls outside the normal assignment system ends up being assigned to the default gender.

In cases where the morphology is regular, but the gender is not, it is necessary to mark the exceptionality lexically. *Diliat/diliatogy* 'eave support' is such a case (21). Here, normal gender assignment is blocked by the presence of an overriding fact in the lexical entry <syn gender> == GENDER\_ASSIGN:<undefined>. This captures the intuition that the normal inputs to gender assignment cannot be used. Looking up undefined at GENDER\_ASSIGN fails to produce an exact match, so the default value is retrieved, namely VIII, the default gender.

(21) Diliat:  
 <> == NOUN  
 <gloss> == eave\_support  
 <syn gender> == GENDER\_ASSIGN:<undefined>  
 <mor class> == 19  
 <phon stem> == diliat.

No special problems for gender assignment are raised by words with extremely irregular morphology such as *gun/gunabys* 'sago pounder', since by definition they belong to the default morphological class which allows appropriate gender assignment.

We have omitted a good deal of detail here, in order to concentrate on the main points, namely that in Arapesh the morphological class of a noun is typically assignable from the phonology, and the gender in turn from the morphological class. As is universally the case, the formal gender assignment rules (morphological in this case) are dominated by the semantic gender assignment rules. Our analysis of Arapesh is, at a high level of generalization, rather similar to our analysis of Russian. And in both cases the analyses have been implemented to demonstrate that they make the correct predictions.

## 6 CONCLUSION

First we presented a typology of gender assignment systems. Then we homed in on two examples in which there are complex interacting factors in assignment. In both cases we outlined analyses according to which gender was seen to be predictable for the vast majority of nouns. We did this in an explicit way, and referred to implementations showing that our analyses do make the correct predictions. This in turn suggests that our confidence in the typology is well placed. It also serves as an illustration of the potential value of formal approaches in typology.

## NOTES

Research for this chapter was supported in part by the Economic and Social Research Council (grants R000233633 and R000236063); this support is gratefully acknowledged. A version of this chapter was read at the workshop 'Back to Basic Issues in Nominal Classification' held at the Max Planck Institute for Psycholinguistics, Nijmegen, 25-27 May, 1993. We are grateful to those present for helpful comments.

- 1 The latter reference is to a revised version of an important article first published in 1968.
- 2 According to Superanskaja (1965: 58), before the Revolution the Russian names of towns on the left bank of the Volga were feminine, and those of towns on the right bank were masculine.
- 3 There are further regularities (Dick Hayward, p.c.); *e* and *o* are not found finally unstressed to give masculine nouns (comparison with the related Saho suggests that final unstressed *e* and *o* were raised). Final *i* and *u* are found more with masculines than with feminines.

- 4 Though standard syntactic tree diagrams have a completely different logic, similar notions to ours have been overlaid on them through notions such as dominance.
- 5 The following automatic phonological correspondences are assumed:
1. /i/ is retracted to its allophone [i̠] after non-back hard (unpalatalized) consonants. Thus the nominative plural form /zakoni/ will be realized with [i̠] but /kost'i/ retains [i] since [t'] is soft.
  2. All consonants which can be palatalized are automatically palatalized before /e/. Thus the locative singular of /zakon/, namely /zakone/, will be realised with a palatalized [n']. If the consonant is already palatalized as in genitive plural /kost'-ej/, it simply remains palatalized. Some consonants are always hard (/š, ž, c/), and remain so before /e/. On the other hand, /č/ and /šč/ are always soft (palatalized), and naturally remain so before /e/. We have chosen to mark softening redundantly for greater clarity in this instance. In addition, the gutturals /k, g, x/ are palatalized before /i/, so that the genitive form /knigi/, from /kniga/ 'book', will be realized with palatalized [g'] (which then demands the front allophone [i]).
  3. There are complex patterns of reduction of vowels in unstressed position, which are not reflected in our transcription since our focus is on morphology. In particular, the unstressed /o/ ending of nouns with soft stems such as /pol'o/ 'field' (orthographically *polo*) is realized as either [ʰ] or [i]. Although this /o/ is never realized as anything approximating to a mid-rounded back vowel, positing /o/ is justified by the stressed [ó] which occurs in [v'inó] 'wine' and [p'it'jó] 'drink(ing)'.

For an informative sketch of Russian phonology, see Timberlake (1993: 828–32).

- 6 We do not examine the complexities of hybrid nouns here, for which see Corbett (1991: 39, 231–2), and for interesting data see Weiss (1991) and Doleschal (1992; 1993: 40–6, 138–44).
- 7 A more extended account to include cases where reference may be to either sex would require us to state the first equation as: GENDER:<> == masc. 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 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. For a Network Morphology account of Russian noun derivation see Hippiisley (1997).
- 9 The full version of the original DATR account and sample output are available in Fraser and Corbett (1995).
- 10 Aronoff (1994: 73–4) allows predictions in both directions; we believe that our simpler proposal should be favoured.
- 11 Our earlier paper (Corbett and Fraser 1993) allowed us to reconcile the competing claims of the three- or four-declensional models by postulating

- the N\_O node. Stankiewicz (1978: 666–7) postulates just two declensions for all the Slavonic languages, and then predicts the differences within the paradigms from gender and from the morphological structure of the stem. Our objections to this analysis are given in Corbett (1982: 207).
- 12 The literature includes rules based on the nominative singular, but since this may include an inflectional ending, which indicates the declensional class, this is hardly predictive. The basis for substantial predictions would be the stem.
- 13 That work is particularly concerned with the nature of defaults, a topic which is not of special relevance here.
- 14 This is what we suggested for the Papuan language Yimas (Corbett 1991: 57).
- 15 Table 9.6 is adapted from Aronoff (1992: 23); note in particular that the morphological class labels used here differ from those used by Aronoff.
- 16 Unlike Aronoff (1992: 22) we do not consider this 'the reverse of the normal pattern'; we consider it fully normal, fitting into the typology proposed.
- 17 In the DATR implementation, upper-case symbols such as the Roman gender labels are normally interpreted as the names of nodes. Since that is not what is intended here the labels must either be quoted thus 'XIII' or preceded by a non-upper character thus\_XIII. Since this has absolutely no linguistic significance, neither device is shown in our discussion here.
- 18 Most DATR interpreters restrict the characters available; here we use U for u.

## REFERENCES

- Aronoff, Mark. 1992. Noun classes in Arapesh. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer. 21–32.
1994. *Morphology by itself: stems and inflectional classes* (Linguistic Inquiry Monograph 22). Cambridge, MA: MIT Press.
- Brown, Dunstan. 1998. From the general to the exceptional: a Network Morphology account of Russian nominal inflection. PhD thesis, University of Surrey.
- Brown, Dunstan, Greville Corbett, Norman Fraser, Andrew Hippiisley and Alan Timberlake. 1996. Russian noun stress and network morphology. *Linguistics* 34: 54–107.
- Claudi, Ulrike. 1985. *Zur Entstehung von Genussystemen: Überlegungen zu einigen theoretischen Aspekten, verbunden mit einer Fallstudie des Zande: mit einer Bibliographie und einer Karte*. Hamburg: Buske.
- Corbett, Greville G. 1982. Gender in Russian: an account of gender specification and its relationship to declension. *Russian Linguistics* 6 (2): 197–232.
1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G., and M. Norman Fraser. 1993. Network morphology: a DATR account of Russian inflectional morphology. *Journal of Linguistics* 29: 113–142.
- Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.

1982. Nominal classification. In R. M. W. Dixon, *Where have all the adjectives gone? and other essays in semantics and syntax*. Berlin: Mouton. 157–233.
1986. Noun classes and noun classification in typological perspective. In Colette Craig (ed.), *Noun classes and categorization*. Amsterdam: Benjamins. 105–12.
- Doleschal, Ursula. 1992. Genus und (Ko-)Referenz. In Tilmann Reuther (ed.), *Slavistische Linguistik 1991: Referate des XVII. Konstanzer Slavistischen Arbeitstreffens, Klagenfurt-St. Georgen/Längsee 10.-14.9.1991* (=Slavistische Beiträge 292). Munich: Otto Sagner. 123–35.
1993. Genus als grammatische und textlinguistische Kategorie: eine kognitiv-funktionalistische Untersuchung des Russischen. PhD, University of Vienna.
- Evans, Roger, and Gerald Gazdar. 1989a. Inference in DATR. *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*. Manchester, UK. 66–71.
- 1989b. The semantics of DATR. In A. G. Cohn (ed.), *Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*. London: Pitman/Morgan Kaufmann. 79–87.
1996. DATR: a language for lexical knowledge representation. *Computational Linguistics* 22: 167–216.
- Fortune, Reo F. 1942. *Arapesh*. (Publications of the American Ethnological Society 19). New York: J. J. Augustin. (Reprinted 1977, New York: AMS Press.)
- Fraser, Norman M., and Greville G. Corbett. 1995. Gender, animacy and declensional class assignment: a unified account for Russian. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1994*. Dordrecht: Kluwer. 123–50.
1997. Defaults in Arapesh. *Lingua* 103: 25–57.
- Gibbon, Dafydd 1992. ILEX: a linguistic approach to computational lexica. In U. Klenk (ed.), *Computatio Linguae: Aufsätze zur algorithmischen und quantitativen Analyse der Sprache* (=Zeitschrift für Dialektologie und Linguistik, Beiheft 73). Stuttgart: Franz Steiner Verlag. 32–53.
- Gudava, T. E. 1967. Godoberinskij jazyk. In E. A. Bokarev and K. V. Lomtadze (eds.), *Ĵazyki narodov SSSR: IV: Iberijsko-kavkazskie jazyki*. Moscow: Nauka. 307–21.
- Hippisley, Andrew. 1997. Declarative derivation: a Network Morphology account of Russian word formation with reference to nouns denoting 'person'. PhD thesis, University of Surrey.
- Kibrik, Alexandr E. 1996. Agreement in conjoined NPs. In Alexandr E. Kibrik, Sergej G. Tatevosov and Alexander Eulenberg (eds.) *Godoberi* (LINCOM Studies in Caucasian Linguistics 2). Munich: Lincom Europa. 156–9.
- Lazova, M. V. (ed.). 1974. *Obratnyj slovar' russkog jazyka: okolo 125 000 slov*. Moscow: Sovetskaja Ėnciklopedija.
- Mylne, Tom. 1995. Grammatical category and world view: Western colonization of the Dyrbal language. *Cognitive Linguistics* 6: 379–404.

- Parker, E. M., and Hayward, R. J. 1985. *An Afar-English-French dictionary (with grammatical notes in English)*. London: School of Oriental and African Studies, University of London.
- Reinhard, Sabine and Dafydd Gibbon. 1991. Prosodic inheritance and morphological generalisations. *Proceedings, Fifth Conference of the European Chapter of the Association for Computational Linguistics*. Berlin 1991. 131–6.
- Smirnova, G. A. 1979. Kategorija roda nesklonjaemyx suščestvitel'nyx. In V. P. Grigor'ev (ed.), *Lingvistika i poetika*. Moscow: Nauka. 86–105.
- Stankiewicz, Edward. 1978. The inflection of Serbo-Croatian substantives and their genitive plural endings. In Henrik Birnbaum (ed.), *American Contributions to the 8th International Congress of Slavists, I: Linguistics and Poetics*. Columbus: Ohio. 666–81.
- Superanskaja, A. V. 1965. Rod zaimstvovannyx suščestvitel'nyx v sovremennom ruskom jazyke. *Voprosy kul'tury rec'i* 6: 44–58.
- Timberlake, Alan. 1993. Russian. In Bernard Comrie and Greville G. Corbett (eds.), *The Slavonic languages*. London: Routledge. 827–86.
- Weiss, Daniel 1991. Sexus distinctions in Polish and Russian. In Maciej Grochowski and Daniel Weiss (eds.), *Words are physicians for an ailing mind* (=Sagner's Slavistische Sammlung 17). Munich: Otto Sagner. 449–466.
- Worth, Dean S. 1966. On the stem/ending boundary in Slavic indeclinables. *Zbornik za filologiju i lingvistiku* 9: 11–16.