

Gender, Animacy, and Declensional Class Assignment: A Unified Account for Russian

NORMAN M. FRASER and GREVILLE G. CORBETT

1. INTRODUCTION¹

In an earlier paper we introduced Network Morphology, an approach to inflectional morphology which relies heavily on the notion of default inheritance (Corbett & Fraser 1993), and showed how a complex set of data pertaining to the inflectional morphology of Russian nominals could be handled in this framework. We extend that analysis here to include a range of phenomena which, we believe, has not previously been described in a single formally explicit framework. The key concept underlying our analysis is that of default inheritance. We introduce this below, together with a lexical knowledge representation language called DATR, which we use to make our analysis fully explicit.² In Section 3 we describe our approach to gender assignment in Russian; animacy assignment is dealt with in Section 4; and declensional class assignment in Section 5.

The issues tackled here – the interrelation of meaning, gender, declensional class and phonology – have been considered by Aronoff (1992), in his discussion of Arapesh based on Fortune (1942), and by Corbett (1991: 7–69). Aronoff and Corbett both highlighted the complex interactions of factors in the assignment of gender, and Aronoff made extensive use of defaults. However, neither had a formal apparatus to show clearly whether their claims were valid. In this paper we go further in that we give a formal account, which allows us to demonstrate that our model does indeed make the correct predictions. Our demonstration is based on Russian, which provides a good range of overlapping factors involved in assignment. Russian has the advantage that, following work by Iloa & Mustajoki (1989), based on Zaliznjak (1977), we have a clear picture of the numbers of nouns of each type in the lexicon. We believe that Aronoff's account of Arapesh is similarly robust, and that it yields the right results if subjected to a comparable formal analysis (Fraser & Corbett 1993). We offer some brief conclusions in Section 6. The DATR encoding of our theory and the output listing which demonstrates the validity of our claims are given as appendices.

2. DEFAULT INHERITANCE

Default inheritance is usually introduced by means of non-linguistic examples, and we make no exception here. Consider the taxonomic hierarchy in Figure 1. The lines in the taxonomy indicate instantiation rather than sub-classifica-

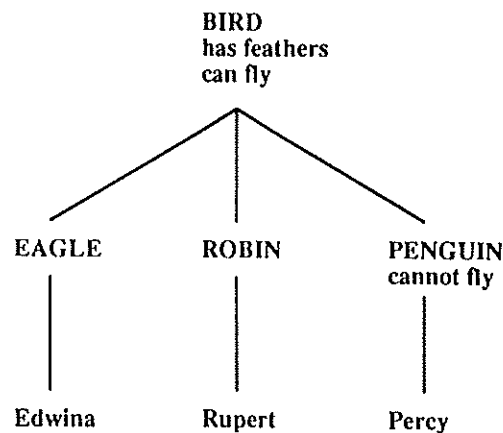


Fig. 1. A simple instantiation hierarchy.

tion. So an eagle is a bird, as is a robin and a penguin; Edwina is an eagle, Rupert is a robin, and Percy is a penguin.

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 some attribute and thus overrides the default (i.e. inheritable) value for that attribute. 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, thereby, 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.

This approach to information organization allows generalizations to be expressed once at a high level, and then automatically to apply to everything which inherits from there. In this way regularities, subregularities and exceptions can be encoded with considerable ease and parsimony. This approach has the added advantage of marking exceptions as such, as in the case of PENGUIN in Figure 1. If Percy were a truly exceptional penguin who had obtained a pilot's license, this extreme degree of markedness would be visible in the inheritance structure by virtue of an exceptional fact (PERCY can fly) overriding an exceptional fact (PENGUINS cannot fly), overriding a default (BIRDS can fly).

The history of linguistics offers many examples of linguistic theories which conflate substance and form. It is now widely acknowledged that the ability to separate these offers at least two benefits: first, it focuses attention on the substantive content of theories rather than on their notational sugar; second,

it facilitates cross-theoretical comparisons (Shieber 1987). Network Morphology is concerned with substantial matters of linguistic theory, which could, in principle, be formalized in a number of different ways. For this purpose, we use an existing formalism called DATR.³ The DATR knowledge representation language was developed by Roger Evans and Gerald Gazdar as a logic-based formalism for describing inheritance networks (Evans & Gazdar 1989a, b). Since computer implementations of the DATR language are available, it is possible to check that an inheritance network expressed in DATR captures the intended generalizations.⁴ All of the analyses in this paper have been checked in this fashion.

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' (also called 'attributes'); 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 clear how this kind of formalism might be useful in expressing linguistic generalizations. Figure 2 shows the inheritance network for Russian nominals which we presented in our earlier work (Corbett & Fraser 1993).

N_I to N_IV are declensional classes. (Notice that we posited the existence of a node N_O from which N_I and N_IV inherit, thereby finding a principled reason for sitting on the fence in the debate over whether Russian has three or four noun declensional classes). The following (incomplete) fragment is taken from our earlier analysis.⁵

```

(2)  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'.
  
```

The first sentence at NOUN should be read as saying that the locative singular consists of the stem followed by an *-e* ending.⁶ A path enclosed in double

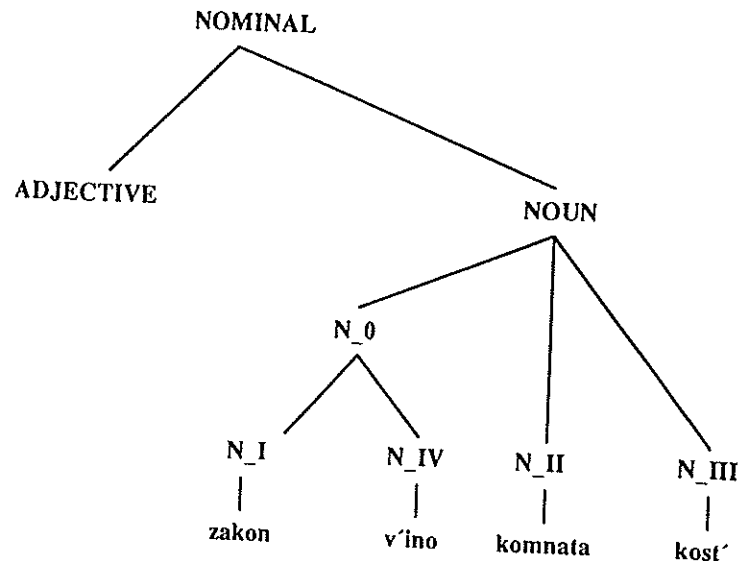


Fig. 2. An inheritance structure for Russian nominals.

quotes in a DATR sentence is used to retrieve the specified value for the thing inheriting from that sentence. If we wanted to find the nominative plural of *Kost'*, we would inherit the sentence $\langle \text{mor nom pl} \rangle == \langle \text{stem} \rangle$ $_i$. Before going any further we would have to find out what the $\langle \text{stem} \rangle$ 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. The definition of locative singular at N_III establishes an asymmetric identity between the locative singular form of an N_III noun and its dative singular.

It is often 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. Consider the following fragment.⁷

- (3) N_II: $\langle \rangle == \text{NOUN}$
 $\langle \text{mor gen sg} \rangle == \langle \text{stem} \rangle _i$
 N_III: $\langle \rangle == \text{NOUN}$
 $\langle \text{mor gen sg} \rangle == \text{N_II}$.

This 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:

- (4) N_III: $\langle \text{mor gen sg} \rangle == \text{N_II}:\langle \text{mor gen sg} \rangle$.

The last form of DATR construction we shall consider here is the so-called 'evaluable path'. Consider the following (slightly simplified) fragment:

- (5) NOUN: $\langle \text{mor gen pl} \rangle == \langle \langle \text{mor stem hardness} \rangle \text{ mor gen pl} \rangle$
 $\langle \text{soft mor gen pl} \rangle == \langle \text{stem} \rangle _ej$
 N_III: $\langle \rangle == \text{NOUN}$
 Kost': $\langle \rangle == \text{N_III}$
 $\langle \text{stem} \rangle == \text{kost'}$
 $\langle \text{mor stem hardness} \rangle == \text{soft}$.

Nouns in Russian can be divided into those having soft (palatalized) stems and those having hard (non-palatalized) stems. The difference is not usually of morphological significance, except in the case of genitive plurals, where it is crucial. The first equation shown here can be interpreted as follows: to obtain the genitive plural form, find out the hardness of the stem and prepend the value to the path $\langle \text{mor gen pl} \rangle$, then evaluate the resulting path.⁸ Thus, to find the genitive plural of *Kost'*, inference would proceed as follows (expressed syllogistically in terms of premises and conclusions):⁹

- (6) a. NOUN: $\langle \text{mor gen pl} \rangle == \langle \langle \text{mor stem hardness} \rangle \text{ mor gen pl} \rangle$
 N_III: $\langle \rangle == \text{NOUN}$
 Kost': $\langle \rangle == \text{N_III}$

 Kost': $\langle \text{mor gen pl} \rangle == \langle \langle \text{mor stem hardness} \rangle \text{ mor gen pl} \rangle$
 b. Kost': $\langle \text{mor gen pl} \rangle == \langle \langle \text{mor stem hardness} \rangle \text{ mor gen pl} \rangle$
 Kost': $\langle \text{mor stem hardness} \rangle == \text{soft}$

 Kost': $\langle \text{mor gen pl} \rangle == \langle \text{soft mor gen pl} \rangle$
 c. NOUN: $\langle \text{soft mor gen pl} \rangle == \langle \text{stem} \rangle _ej$
 N_III: $\langle \rangle == \text{NOUN}$
 Kost': $\langle \rangle == \text{N_III}$

 Kost': $\langle \text{soft mor gen pl} \rangle == \langle \text{stem} \rangle _ej$
 d. Kost': $\langle \text{soft mor gen pl} \rangle == \langle \text{stem} \rangle _ej$
 Kost': $\langle \text{stem} \rangle == \text{kost'}$

 Kost': $\langle \text{soft mor gen pl} \rangle == \text{kost' } _ej$
 e. Kost': $\langle \text{mor gen pl} \rangle == \langle \text{soft mor gen pl} \rangle$
 Kost': $\langle \text{soft mor gen pl} \rangle == \text{kost' } _ej$

 Kost': $\langle \text{mor gen pl} \rangle == \text{kost' } _ej$

This concludes our brief introduction to default inheritance and to the DATR formalism. More detailed introductions to default inheritance can be found in Gazdar (1987) and Daelemans, de Smedt and Gazdar (1992), and to DATR in Gazdar (1990; forthcoming). In the following sections we show how Network Morphology uses default inheritance to produce elegant solutions for some well known problems in the analysis of Russian nouns.

3. GENDER ASSIGNMENT

Russian has three genders, masculine, feminine and neuter.¹⁰ 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). Specifically 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* '(female) teacher'

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

These rules leave very few exceptions, but there are many nouns which are not covered by them. These remaining nouns 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 class 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 *mužčina* '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.¹¹

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.¹²

In our Network Morphology account, 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.¹³

```
[1] 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 referent is retrieved) and then a path consisting only of the sex is evaluated at the node GENDER (shown below):

```
[2] 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 referent of the noun is male then the gender is masculine; if the referent is female then the gender is feminine.¹⁴ 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 five main declensional classes. Thus, for example, the following equation is located at the node for declensional class II nouns, N_II:

```
[3] 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). 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 denoting a non-sex-differentiated referent (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.

- [4] N_V: <formal_gender> == < "<sem animacy>" >
 <animate> == masc
 <inanimate> == neut.

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 referent and neuter gender if the referent is 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 correct predictions are indeed made (see Appendix B).

4. ANIMACY ASSIGNMENT

Animacy in Russian is of particular interest. It is reflected in the accusative case forms of certain nouns, and in agreement. There are no exclusively animate accusative forms in Russian; instead, animacy shows itself by different patterns of syncretism. With an animate noun of the first declensional class we find the accusative identical to the genitive.

- (7) pervogo (acc = gen) studenta (acc = gen)
 first student
 'the first student'

With an inanimate, by contrast, the accusative is as the nominative:

- (8) pervyj (acc = nom) zakon (acc = nom)
 first law
 'the first law'

In both these examples, animacy is reflected not only in the morphological form of the noun, but also in the agreement of the adjective. This agreement can also be found with those nouns which decline according to pattern N_II, and so have a unique accusative form, but which are masculine animate, as is the case with *mužčina* 'man':

- (9) pervogo (acc = gen) mužčinu (acc)
 first man
 'the first man'

Animacy is a sub-gender (Corbett 1991: 165–167). It is narrower in range than the three main genders, since it affects only the accusative case. It is a more recent addition to the system than the main genders, and assignment to the animate sub-gender is (still) largely based on semantics: animate nouns are those denoting humans, animals and insects. Returning to the morphology, the crucial point is that we get the same regularity but in different paradigms:

the accusative matches the genitive for animate nouns under certain circumstances, and the nominative for inanimate nouns, though the phonological form of the nominative or genitive varies from paradigm to paradigm.

In Corbett & Fraser (1993) we specified animacy in lexical entries as a semantic feature; then by default, syntactic animacy was the same as semantic animacy. However, there are further simplifications available. First we can say that all sex-differentiables will be animate, hence semantic animacy need not be specified in their lexical entries. Second, by default, nouns are semantically inanimate. This second default makes sense in the light of the statistical evidence provided by Ilola & Mustajoki (1989: 15–17), based on Zaliznjak (1977): of the 45,779 nouns in Zaliznjak's dictionary, 75% (34,178) denote inanimates. Together, these two generalizations mean that in most lexical entries there is no need to specify animacy.

The default identity between syntactic animacy and semantic animacy can be expressed trivially in DATR:

- [5] NOUN: <syn animacy> == "<sem animacy>"

This simply says that the syntactic animacy is identical to the semantic animacy. Since this default generalization is stated at the level of the NOUN node, it may be overridden by equations in the lexical entries of words with conflicting semantic and syntactic animacy. *Kon'* 'knight (chess piece)' is such a word: it is semantically inanimate but syntactically animate.

So much for syntactic animacy. For nouns which are sex-differentiable, semantic animacy is obtained as follows:

- [6] NOUN: <sem animacy>:
 == ANIMACY:< "<sem sex>" >.
- [7] ANIMACY: <> == animate
 <undifferentiated> == inanimate.

The path shown at [6] indicates that the semantic animacy may be found by evaluating the sex of the word's referent at the ANIMACY node. If the sex is undifferentiated, the value inanimate is found at the ANIMACY node, as shown in [7]. For all other values of <sem sex> (i.e. male or female) the value animate is found. Thus the ANIMACY node can be thought of as a look-up table specifying default sex/animacy associations.

Nouns are inanimate by default because their sex is undifferentiated by default and the final equation in [7] establishes the connection between the two. The undifferentiated sex default is encoded by means of the following equation, which is located at the NOUN node:

- [8] NOUN: <sem sex> == undifferentiated

Our analysis now accounts for the syntactic animacy as well as the gender of the vast majority of Russian nouns in a formally explicit way. While animacy is less complex than gender, it is perhaps surprising to see just how simple

it becomes: with appropriate use of defaults, animacy need hardly ever be specified in lexical entries. Again, we make the right predictions as Appendix B shows.

5. DECLENSIONAL CLASS ASSIGNMENT

So far we have specified the declensional class of each noun 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.¹⁵ 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 N_I: for example, *student* '(male) student';
2. Sex-differentiable nouns denoting females are of declensional class N_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' (as in example (3)), which denote males but which decline according to declensional class N_II. (Lazova 1974: 942-943 puts the figure at 273, but the number involved is actually larger because there are many hypocoristics of this type, like *Saša* '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 N_III.

Formal assignment rule for declensional class

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

The effect of this rule is to make nouns whose stem ends in a vowel indeclinable (class N_V are the indeclinables). Our rule follows Worth (1966), though as he points out the idea goes back much further. A noun like *taksi* '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:

```
[9] NOUN: <mor> == "<declensional_class>"
      <declensional_class> == DECLENSION:
      <"<infl_root final>" "<sem sex>" >
```

The first equation indicates that in order to find a value (or values) for the <mor> path (or paths), 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 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.¹⁶

```
[10] NOUN: <infl_root final> == 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 referent. The DECLENSION node is given below:

```
[11] DECLENSION: <consonant male> == N_I:<>
                  <consonant female> == N_II:<>
                  <vowel $sex> == 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 referents. Such nouns must typically specify declensional class in their lexical entries. However, as we shall see shortly (in the discussion leading to [11']), a substantial proportion of them can be assigned to a declensional class by the setting of a default.

Let us first look at the working of [11]. Take a noun such as *student* 'student', the lexical entry for which is as follows:

```
[12] Student: <> == NOUN
              <gloss> == student
              <infl_root all> == student
              <sem sex> == male.
```

To which declensional class should this word be assigned? Interesting though this question is, it is likely to be of less importance than specific questions such as 'what is the dative plural form of this word?', or 'what is its nominative singular?' Let us take the second question. In terms of our theory, it maps onto the question 'what is the value of the <mor nom sg> path?' Since the Student node inherits directly from the NOUN node, the two equations shown in [9] 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 [9]) requires the evaluation at the DECLENSION node of a path consisting of the value of <infl_root final> for *student*, followed by the value of <sem sex> for the same word. The value of <infl_root final> is consonant according to the unoverridden equation [10], and the value of <sem sex> for *student* 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. However, this is not the end of the look-up process. We were interested in the value of the path <mor nom sg>, not just <mor>.

It is necessary at this point to introduce another kind of DATR equation which was not mentioned earlier. An expression of the form NODE:<> may appear on the right hand side of an equation, for example, NODE1:Path1 == NODE2:<>. In this example, an attempt to evaluate NODE1:SomePath will succeed and return a value if SomePath can be formed by concatenating Path1 with a path defined at NODE2. In effect, the NODE:<> notation allows the description of a path to be distributed over more than one node.

The entry in the DATR theory for N_I is given 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 Student node will lead via NOUN to be following inference:¹⁷

```
(10) Student: <mor> == N_I:<>
```

This allows the <mor> path from NOUN to be concatenated with the <nom sg> path from N_I to produce the following inference:

```
(11) Student: <mor nom sg> == "<stem sg>"
```

In other words, the nominative singular of *student* has a zero ending:

```
(12) Student: <mor nom sg> = student
```

It is worth contrasting declensional class assignment with gender assignment. A clear difference is the outcome when semantic and formal rules make dif-

ferent 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,¹⁸ the formal factor dominates. Thus *attaše* 'attaché' denotes a male and 'should' be in declensional class N_I, but it ends in a vowel and so 'should' be in declensional class N_V. It is actually in declensional class N_V. Conversely it denotes a male and so 'should' be masculine but it is in class N_V and so 'should' be neuter; as already noted, semantic factors take precedence in gender assignment and so the noun 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 N_I, N_II, N_III or N_IV.¹⁹ 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 1.

Table 1. Number of nouns in the different declensional classes.

Class	N_I	N_II	N_III	N_IV	N_V
Example	<i>zakon</i> 'law'	<i>komnata</i> 'room'	<i>kość</i> 'bone'	<i>vino</i> 'wine'	<i>taksi</i> 'taxi'
No of nouns	20,850	16,050	5,150	11,050	450 ²⁰

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 N_III have the suffix *-ost'* which forms abstract nouns from adjectives (*star-yj* 'old', *star-ost'* 'old age'). If the suffix is labelled as belonging to class N_III, then the number of distinct members of the class is substantially reduced. Similarly nominalizations in *-anie/-enie* (like *razrušenie* 'destruction', derived from *razrušit'* 'destroy') inflate the figure for N_IV (see Schupbach 1984 for discussion).

Table 1 shows that N_I and N_II have substantially more members than the other three. We can therefore set a default which will put nouns in the majority declensional class N_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 N_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 N_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 more debatable. If N_I is the default declensional class, 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 [11].

```
[11'] 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.

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 for nouns. Note that we are claiming only that the masculine is the default gender for *nouns*. In a fuller analysis of Russian, 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). Under these circumstances 'neutral' agreement forms occur (these are usually identical to neuter forms. Corbett 1991: 204-205, 216). 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 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 and in another it is the neuter.²¹

In this section we have shown how the declensional class can be predicted for a substantial proportion of Russian nouns. We have seen how declensional class assignment differs from gender assignment in interesting ways; further research is needed to show whether these differences are specific to Russian or, as we suspect, are found more generally.²²

6. CONCLUSION

In previous work we gave an account of gender assignment in Russian. Other analyses have given at least a partial account of declensional class assignment. To the best of our knowledge no-one has been able to combine them in a single analysis, as our present Network Morphology analysis does. It is difficult to be certain about this since many accounts are vague. DATR requires us to make our claims explicit. Having laid out our claims, we can demonstrate that they do indeed make the correct predictions, as shown in Appendix B.

The success of our gender assignment analysis of the Russian data lends further support to the typology of gender assignment of Corbett (1991: 7-69). Since the account of Russian gender assignment (one of the more complex systems) has been shown to be robust, it follows that the entire typology is more securely founded. The fact that we have been able to combine gender, animacy and declensional class assignment into a single account, with a resulting considerable simplification of lexical entries, suggests that this approach to morphology based, in large part, on default inheritance and implemented in DATR is an extremely fruitful one.

APPENDIX A: DATR THEORY

```
*****
% TITLE:    Russian nominal inflection
% AUTHORS:  Grev Corbett & Norman Fraser, University
%           of Surrey
% NOTE:     The labels for lexical entries are
%           simple transliterations; linguistic forms
%           are in phonological transcription.
*****
# Vars $number: sg pl.
# vars $sex: male female undifferentiated.
```

```

#####
% WORD CLASS HIERARCHY
#####
NOMINAL:    <stem> == "<infl_root all>"
            <phon stem hardness> == hard.

NOUN:       <> == NOMINAL
            <infl_root final> == consonant    [10]
            <mor> == "<declensional_class>"    [9]
            "<declensional_class>" ==
            DECLENSION: < "<infl_root
            final>" "<sem sex>" >
            <syn cat> == n
            <sem sex> == undifferentiated    [8]
            <syn gender> == GENDER:< "<sem sex>" > [1]
            <syn animacy> == "<sem animacy>" [5]
            <sem animacy> ==
            ANIMACY:< "<sem sex>" >. [6]

#####
% SOME CASE SWITCHES
#####
GENDER:     <male> == masc [2]
            <female> == fem
            <undifferentiated> ==
            "<mor formal_gender>".

ANIMACY:    <> == animate [7]
            <undifferentiated> == inanimate.

ACCUSATIVE: <> == "<mor nom>"
            <pl animate> == "<mor gen pl>"
            <sg animate masc> == "<mor gen sg>"

DECLENSION: <consonant male> == N_I:<> [11]
            <consonant female> == N_II:<>
            <vowel $sex> == N_V:<>

#####
% DECLENSIONAL CLASS HIERARCHY
#####
N_NOM:      <stem hardness> ==
            "<phon stem hardness>"
            <acc $number> == ACCUSATIVE: < $number
            "<syn animacy>" "<syn gender>" >

```

```

<dat pl> ==
  "<stem pl>" "<mor theme_vowel>" _m
<inst pl> ==
  "<stem pl>" "<mor theme_vowel>" _m'i
<loc pl> ==
  "<stem pl>" "<mor theme_vowel>" _x.

N_NOUN:     <> == N_NOM
            <loc sg> == "<stem sg>" _e
            <nom pl> == "<stem pl>" _i
            <gen pl> ==
            "< mor "<mor stem hardness>" gen pl>"
            <soft gen pl> == "<stem pl>" _ej
            <theme_vowel> == _a.

N_O:        <> == N_NOUN
            <gen sg> == "<stem sg>"_a
            <dat sg> == "<stem sg>"_u
            <inst sg> == "<stem sg>"_om.

N_I:        <> == N_O [13]
            <formal_gender> == masc
            <nom sg> == "<stem sg>"
            <hard gen pl> == "<stem pl>"_ov.

N_II:       <> == N_NOUN [3]
            <formal_gender> == fem
            <nom sg> == "<stem sg>"_a
            <acc sg> == "<stem sg>"_u
            <gen sg> == "<stem sg>"_i
            <dat sg> == "<stem sg>"_e
            <inst sg> == "<stem sg>"_oj
            <hard gen pl> == "<stem pl>"

N_III:      <> == N_NOUN
            <formal_gender> == fem
            <stem hardness> == soft
            <nom sg> == N_I
            <gen sg> == N_II
            <dat sg> == "<mor gen sg>"
            <inst sg> == "<stem sg>"_ju
            <loc sg> == "<mor dat sg>".

N_IV:       <> == N_O
            <formal_gender> == neut
            <nom sg> == "<stem sg>"_o
            <nom pl> == "<stem pl>"_a
            <hard gen pl> == N_II.

```

```

N_V:      <> == "<infl_root all>"
          <formal_gender> == < "<sem animacy>" >
          <animate> == masc
                    % IF animate THEN masc
          <inanimate> == neut.
                    % IF inanimate THEN neut
          %%%%%%%%%%%
          % LEXICAL ENTRIES
          %%%%%%%%%%%
NZakon:   <> == NOUN
          <declensional_class> == N_I:<>
          <gloss> == law
          <infl_root all> == zakon.

Nož:     <> == NOUN
          <declensional_class> == N_I:<>
          <gloss> == knife
          <infl_root all> == nož
          <mor stem hardness> == soft.

Student: <> == NOUN                                [12]
          <gloss> == student
          <infl_root all> == student
          <stem sex> == male.

Bolgarin: <> == NOUN
          <gloss> == bulgarian_man
          <infl_root all> == bolgar
          <stem sg> == "<infl_root all>" _'in
          <mor gen pl> == N_II:<hard gen pl>
          <sem sex> == male.

Komnata: <> == NOUN
          <declensional_class> == N_II:<>
          <gloss> == room
          <infl_root all> == komnat.

Učitel'nica: <> == NOUN
          <gloss> == female_teacher
          <infl_root all> == uč'itel'n'ic
          <sem sex> == female.

Mužčina: <> == NOUN
          <declensional_class> == N_II:<>
          <gloss> == man
          <infl_root all> == mužč'in
          <sem sex> == male.

```

```

Kost':   <> == NOUN
          <declensional_class> == N_III:<>
          <gloss> == bone
          <infl_root all> == kost'.

Myš':   <> == NOUN
          <declensional_class> == N_III:<>
          <gloss> == mouse
          <infl_root all> == miš
          <sem animacy> == animate.

Put':   <> == NOUN
          <declensional_class> == N_III:<>
          <gloss> == way
          <mor inst sg> == N_I:<inst sg>
          <mor formal_gender> ==
            N_I:<formal_gender>
          <infl_root all> == put'.

Vino:   <> == NOUN
          <declensional_class> == N_IV:<>
          <gloss> == wine
          <infl_root all> == v'in.

More:   <> == NOUN
          <declensional_class> == N_IV:<>
          <gloss> == sea
          <infl_root all> == mor'
          <phon stem hardness> == soft.

Taksi:  <> == NOUN
          <gloss> == taxi
          <infl_root all> == taks'i
          <infl_root final> == vowel.

Gnu:    <> == NOUN
          <gloss> == gnu
          <infl_root all> == gnu
          <infl_root final> == vowel
          <sem animacy> == animate.

```

APPENDIX B: SOME THEOREMS OF THE DATR THEORY

```

Zakon:  <gloss> = law.
Zakon:  <mor nom sg> = zakon.
Zakon:  <mor acc sg> = zakon.
Zakon:  <mor gen sg> = zakon _a.
Zakon:  <mor dat sg> = zakon _u.

```

Zakon: <mor inst sg> = zakon _om.
 Zakon: <mor loc sg> = zakon _e.
 Zakon: <mor nom pl> = zakon _i.
 Zakon: <mor acc pl> = zakon _i.
 Zakon: <mor gen pl> = zakon _ov.
 Zakon: <mor dat pl> = zakon _a _m.
 Zakon: <mor inst pl> = zakon _a _m'i.
 Zakon: <mor loc pl> = zakon _a _x.
 Zakon: <syn gender> = masc.
 Zakon: <syn animacy> = inanimate.

Nož: <gloss> = knife.
 Nož: <mor nom sg> = nož.
 Nož: <mor acc sg> = nož.
 Nož: <mor gen sg> = nož _a.
 Nož: <mor dat sg> = nož _u.
 Nož: <mor inst sg> = nož _om.
 Nož: <mor loc sg> = nož _e.
 Nož: <mor nom pl> = nož _i.
 Nož: <mor acc pl> = nož _i.
 Nož: <mor gen pl> = nož _ej.
 Nož: <mor dat pl> = nož _a _m.
 Nož: <mor inst pl> = nož _a _m'i.
 Nož: <mor loc pl> = nož _a _x.
 Nož: <syn gender> = masc.
 Nož: <syn animacy> = inanimate.

Student: <gloss> = student.
 Student: <mor nom sg> = student.
 Student: <mor acc sg> = student _a.
 Student: <mor gen sg> = student _a.
 Student: <mor dat sg> = student _u.
 Student: <mor inst sg> = student _om.
 Student: <mor loc sg> = student _e.
 Student: <mor nom pl> = student _i.
 Student: <mor acc pl> = student _ov.
 Student: <mor gen pl> = student _ov.
 Student: <mor dat pl> = student _a _m.
 Student: <mor inst pl> = student _a _m'i.
 Student: <mor loc pl> = student _a _x.
 Student: <syn gender> = masc.
 Student: <syn animacy> = animate.

Bolgarin: <gloss> = bulgarian_man.
 Bolgarin: <mor nom sg> = bolgar _'in.
 Bolgarin: <mor acc sg> = bolgar _'in _a.
 Bolgarin: <mor gen sg> = bolgar _'in _a.

Bolgarin: <mor dat sg> = bolgar _'in _u.
 Bolgarin: <mor inst sg> = bolgar _'in _om.
 Bolgarin: <mor loc sg> = bolgar _'in _e.
 Bolgarin: <mor nom pl> = bolgar _i.
 Bolgarin: <mor acc pl> = bolgar.
 Bolgarin: <mor gen pl> = bolgar.
 Bolgarin: <mor dat pl> = bolgar _a _m.
 Bolgarin: <mor inst pl> = bolgar _a _m'i.
 Bolgarin: <mor loc pl> = bolgar _a _x.
 Bolgarin: <syn gender> = masc.
 Bolgarin: <syn animacy> = animate.

Komnata: <gloss> = room.
 Komnata: <mor nom sg> = komnat _a.
 Komnata: <mor acc sg> = komnat _u.
 Komnata: <mor gen sg> = komnat _i.
 Komnata: <mor dat sg> = komnat _e.
 Komnata: <mor inst sg> = komnat _oj.
 Komnata: <mor loc sg> = komnat _e.
 Komnata: <mor nom pl> = komnat _i.
 Komnata: <mor acc pl> = komnat _i.
 Komnata: <mor gen pl> = komnat.
 Komnata: <mor dat pl> = komnat _a _m.
 Komnata: <mor inst pl> = komnat _a _m'i.
 Komnata: <mor loc pl> = komnat _a _x.
 Komnata: <syn gender> = fem.
 Komnata: <syn animacy> = inanimate.

Učitel'nica: <gloss> = female_teacher.
 Učitel'nica: <mor nom sg> = uč'itel'n'ic _a.
 Učitel'nica: <mor acc sg> = uč'itel'n'ic _u.
 Učitel'nica: <mor gen sg> = uč'itel'n'ic _i.
 Učitel'nica: <mor dat sg> = uč'itel'n'ic _e.
 Učitel'nica: <mor inst sg> = uč'itel'n'ic _oj.
 Učitel'nica: <mor loc sg> = uč'itel'n'ic _e.
 Učitel'nica: <mor nom pl> = uč'itel'n'ic _i.
 Učitel'nica: <mor acc pl> = uč'itel'n'ic.
 Učitel'nica: <mor gen pl> = uč'itel'n'ic.
 Učitel'nica: <mor dat pl> = uč'itel'n'ic _a _m.
 Učitel'nica: <mor inst pl> = uč'itel'n'ic _a _m'i.
 Učitel'nica: <mor loc pl> = uč'itel'n'ic _a _x.
 Učitel'nica: <syn gender> = fem.
 Učitel'nica: <syn animacy> = animate.

Muščina: <gloss> = man.
 Muščina: <mor nom sg> = mušč'in _a.
 Muščina: <mor acc sg> = mušč'in _u.

Mužčina: <mor gen sg> = mužč'in _i.
 Mužčina: <mor dat sg> = mužč'in _e.
 Mužčina: <mor inst sg> = mužč'in _oj.
 Mužčina: <mor loc sg> = mužč'in _e.
 Mužčina: <mor nom pl> = mužč'in _i.
 Mužčina: <mor acc pl> = mužč'in.
 Mužčina: <mor gen pl> = mužč'in.
 Mužčina: <mor dat pl> = mužč'in __a _m.
 Mužčina: <mor inst pl> = mužč'in _a _m'i.
 Mužčina: <mor loc pl> = mužč'in _a _x.
 Mužčina: <syn gender> = masc.
 Mužčina: <syn animacy> = animate.

Kost': <gloss> = bone.
 Kost': <mor nom sg> = kost'.
 Kost': <mor acc sg> = kost'.
 Kost': <mor gen sg> = kost' _i.
 Kost': <mor dat sg> = kost' _i.
 Kost': <mor inst sg> = kost' _ju.
 Kost': <mor loc sg> = kost' _i.
 Kost': <mor nom pl> = kost' _i.
 Kost': <mor acc pl> = kost' _i.
 Kost': <mor gen pl> = kost' _ej.
 Kost': <mor dat pl> = kost' _a _m.
 Kost': <mor inst pl> = kost' _a _m'i.
 Kost': <mor loc pl> = kost' _a _x.
 Kost': <syn gender> = fem.
 Kost': <syn animacy> = inanimate.

Myš': <gloss> = mouse.
 Myš': <mor nom sg> = miš.
 Myš': <mor acc sg> = miš.
 Myš': <mor gen sg> = miš _i.
 Myš': <mor dat sg> = miš _i.
 Myš': <mor inst sg> = miš _ju.
 Myš': <mor loc sg> = miš _i.
 Myš': <mor nom pl> = miš _i.
 Myš': <mor acc pl> = miš _ej.
 Myš': <mor gen pl> = miš _ej.
 Myš': <mor dat pl> = miš _a _m.
 Myš': <mor inst pl> = miš _a _m'i.
 Myš': <mor loc pl> = miš _a _x.
 Myš': <syn gender> = fem.
 Myš': <syn animacy> = animate.

Put': <gloss> = way
 Put': <mor nom sg> = put'.

Put': <mor acc sg> = put'.
 Put': <mor gen sg> = put' _i.
 Put': <mor dat sg> = put' _i.
 Put': <mor inst sg> = put' _om.
 Put': <mor loc sg> = put' _i.
 Put': <mor nom sg> = put' _i.
 Put': <mor acc pl> = put' _i.
 Put': <mor gen pl> = put' _ej.
 Put': <mor dat pl> = put' _a _m.
 Put': <mor inst pl> = put' _a _m'i.
 Put': <mor loc pl> = put' _a _x.
 Put': <syn gender> = masc.
 Put': <syn animacy> = inanimate.

Vino: <gloss> = wine.
 Vino: <mor nom sg> = v'in _o.
 Vino: <mor acc sg> = v'in _o.
 Vino: <mor gen sg> = v'in _a.
 Vino: <mor dat sg> = v'in _u.
 Vino: <mor inst sg> = v'in _om.
 Vino: <mor loc sg> = v'in _e.
 Vino: <mor nom pl> = v'in _a.
 Vino: <mor acc pl> = v'in _a.
 Vino: <mor gen pl> = v'in.
 Vino: <mor dat pl> = v'in _a _m.
 Vino: <mor inst pl> = v'in _a _m'i.
 Vino: <mor loc pl> = v'in _a _x.
 Vino: <syn gender> = neut.
 Vino: <syn animacy> = inanimate.

More: <gloss> = sea.
 More: <mor nom sg> = mor' _o.
 More: <mor acc sg> = mor' _o.
 More: <mor gen sg> = mor' _a.
 More: <mor dat sg> = mor' _u.
 More: <mor inst sg> = mor' _om.
 More: <mor loc sg> = mor' _e.
 More: <mor nom pl> = mor' _a.
 More: <mor acc pl> = mor' _a.
 More: <mor gen pl> = mor' _ej.
 More: <mor dat pl> = mor' _a _m.
 More: <mor inst pl> = mor' _a _m'i.
 More: <mor loc pl> = mor' _a _x.
 More: <syn gender> = neut.
 More: <syn animacy> = inanimate.

Taksi: <gloss> = taxi

Taksi: <mor nom sg> = taks'i.
 Taksi: <mor acc sg> = taks'i.
 Taksi: <mor gen sg> = taks'i.
 Taksi: <mor dat sg> = taks'i.
 Taksi: <mor inst sg> = taks'i.
 Taksi: <mor loc sg> = taks'i.
 Taksi: <mor nom pl> = taks'i.
 Taksi: <mor acc pl> = taks'i.
 Taksi: <mor gen pl> = taks'i.
 Taksi: <mor dat pl> = taks'i.
 Taksi: <mor inst pl> = taks'i.
 Taksi: <mor loc pl> = taks'i.
 Taksi: <syn gender> = neut.
 Taksi: <syn animacy> = inanimate.

Gnu: <gloss> = gnu.
 Gnu: <mor nom sg> = gnu.
 Gnu: <mor acc sg> = gnu.
 Gnu: <mor gen sg> = gnu.
 Gnu: <mor dat sg> = gnu.
 Gnu: <mor inst sg> = gnu.
 Gnu: <mor loc sg> = gnu.
 Gnu: <mor nom pl> = gnu.
 Gnu: <mor acc pl> = gnu.
 Gnu: <mor gen pl> = gnu.
 Gnu: <mor dat pl> = gnu.
 Gnu: <mor inst pl> = gnu.
 Gnu: <mor loc pl> = gnu.
 Gnu: <syn gender> = masc.
 Gnu: <syn animacy> = animate.

NOTES

¹ Versions of this paper were read at the Spring Meeting of the Linguistics Association of Great Britain, University of Birmingham, March 1993 and at the Finnish Circle of Slavists, Helsinki, May 1994. We would like to thank those present for helpful comments, and particularly Alan Timberlake and Ursula Doleschal for discussion. We are also very grateful to Gerald Gazdar and Dafydd Gibbon for a number of useful suggestions, and to Petra Barg for comments on an earlier draft. Any errors are ours. This research was supported by the Economic and Social Research Council (grant R000233633) and the Leverhulme Trust (grant F.242M); the support of both funding bodies is gratefully acknowledged.

² Our approach is complementary to Gibbon's work on ILEX. The latter is an approach to the lexicon in computational linguistics, implemented in DATR. '... the ILEX concept may be thought of as a set of linguistic constraints on the form of possible DATR representations' (Gibbon 1992: 47). Network Morphology can be viewed as a set of linguistic constraints on possible DATR representations of morphology.

³ The choice 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 & Hudson 1992). DATR's special attractions are a conceptually simple syntax and semantics, and a range of existing computer interpreters for the language.

⁴ The implementations we know of are by A. Boltz (Konstanz), Roger Evans (Sussex), Norman Fraser (Surrey), Dafydd Gibbon (Bielefeld), James Kilbury (Düsseldorf), Hagen Langer (Bielefeld) and A. Sikorski (Poznań).

⁵ Morphological forms to be accounted for are given in phonemic transcription (while other forms, such as in examples (7)–(9) below are in the standard transliteration). The automatic phonological correspondences we assume are as follows:

1. /i/ is retracted to its allophone [i] after non-back hard (unpalatalized) consonants. Thus the nominative plural form /zakoni/ will be realized with [i] 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 /knig'i/, 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 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'ol' 'field' (orthographically *pole*) is realized as either [ə] 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: 828–832).

⁶ Some implementations of DATR require value sequences to be enclosed in round brackets; other implementations allow bracket enclosed sequences but do not require them. Here we follow the current practice of Gazdar and Evans and omit all round brackets.

⁷ Only facts relevant to the immediate discussion are shown at nodes N_II and N_III in (3). In our full analysis of Russian nominal morphology, each of these nodes records a much richer set of facts than is shown here. The facts for N_III are quite different from those for N_II, so an analysis of the form 'N_III:<> == N_II' is out of the question. See Appendix A for further details.

⁸ A full account of the complexities of the genitive plural is given in Brown and Hippiusley (1994).

⁹ This presentation is illustrative only, since it masks a formal distinction in DATR between the '=' and '== ' symbols. See Evans & Gazdar (1989a) for further discussion.

¹⁰ For a recent discussion of how gender can be defined for Russian see Hubenschmid (1993: 47–77).

¹¹ 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 Doleschal (1992, 1993: 40–46, 138–144).

¹² 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.

¹³ DATR fragments from our theory of Russian nominal morphology are identified by square-bracketed numbers. These numbers are also used to pick out the fragments in the full theory given in Appendix A.

¹⁴ 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*.

¹⁵ Thus Thelin (1975: 30) claims that 'the inflectional class membership of Russian nouns can to a great extent be predicted from information on gender and/or the nature of the final consonant of the stem'; see also Crockett (1976: 12). And Stankiewicz (1978: 666–667) 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. Corbett (1982) demonstrates that full prediction of declensional class from gender is not possible, but that prediction of gender from meaning and declensional class is possible, given four declensional types. Our earlier paper (Corbett & Fraser 1993) allowed us to reconcile the competing claims of the three- or four-declensional models (our objections to Stankiewicz's two-declensional model are given in Corbett 1982: 207). Here we are adding in the predictability of declensional class which, though weaker than that of gender (Corbett 1982: 224), still permits a reduction in the lexical entry of large numbers of nouns. Halle (1992: 37–38, 45) postulates the prediction of declensional class from gender for Latvian, Latin and Russian, but does not argue for this position rather than for the converse. Aronoff (1994: 73–74) allows predictions in both directions.

Some claim to predict declensional class from gender but use the nominative singular as the lexical entry rather than the stem; for many nouns the nominative singular effectively provides the stem and the declensional class, hence there is actually very little prediction going on.

¹⁶ There is no reason why such a phonological component could not also be formalized using DATR, as the work of Reinhard & Gibbon (1991) and Gibbon (1992) demonstrates.

¹⁷ The inferences shown in (10)–(12) may be drawn from the DATR theory but do not belong to it. Hence they cannot be indexed with any part of the theory by means of square bracketed identifiers.

¹⁸ 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.

¹⁹ If the stem ends in a hard consonant, then type N_III can be eliminated. We have no rule to predict membership of class N_III; stems are by default 'hard' (they end in a hard consonant). The N_III class is quite small (see discussion after Table 1).

²⁰ This figure includes the small number of nouns (32 in Lazova's table) which are indeclinable 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 indeclinable. These require an exceptional marker just to indicate that they are indeclinable: the semantic gender assignment rule will then correctly assign them to the feminine gender. Rare borrowings which are indeclinable, 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.

²¹ 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). In Arapesh, Aronoff (1992) discovered remarkable consistency in the way the grammatical defaults lined up. In other cases, as in Russian, the picture is less straightforward. For an unusual type of gender default, see Corbett & Mtenje (1987: 15) on Chichewa.

²² The declensional class assignments are not given in Appendix B; this gives all the available morphological forms (which require the declensional class assignment to be correct) and the gender and animacy. In other words, we provide all the information which might be required by the syntax (for agreement purposes), and the morphological forms supplied according to syntactic specifications. This seems to be a realistic set of outputs. The declensional class assignments are correct, and could be shown; the point is simply that there is nothing outside the morphological component which requires information about them.

REFERENCES

- Aronoff, Mark. 1992. "Noun Classes in Arapesh". In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 21–32.
- Aronoff, Mark. 1994. *Morphology By Itself: Stems and Inflectional Classes* (= Linguistic Inquiry Monograph 22). Cambridge MA: MIT Press.
- Brown, Dunstan and Andrew Hippisley. 1994. "Conflict in Russian Genitive Plural Assignment: A Solution Represented in DATR". *Journal of Slavic Linguistics* 2, 30–48.
- Corbett, Greville G. 1982. "Gender in Russian: an Account of Gender Specification and its Relationship to Declension". *Russian Linguistics* 6, 197–232.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. and Norman M. Fraser. 1993. "Network Morphology: A DATR Account of Russian Inflectional Morphology". *Journal of Linguistics* 29, 113–142.
- Corbett, Greville G. and Alfred D. Mtenje. 1987. "Gender Agreement in Chichewa". *Studies in African Linguistics* 18, 1–38.
- Crockett, Dina B. 1976. *Agreement in Contemporary Standard Russian*. Cambridge, MA: Slavica.
- Daelemans, Walter, Koehnraad De Smedt and Gerald Gazdar. 1992. "Inheritance in Natural Language Processing". *Computational Linguistics* 18, 205–218.
- Doleschal, Ursula. 1992. "Genus und (Ko-)Referenz". In Tilmann Reuther (ed.), *Slavistische Linguistik 1991: Referate des XVII. Konstanzer Slavistischen Arbeitstreffens, Klagenfurt-St. Georg/Längsee 10.–14.9.1991* (= Slavistische Beiträge 292). Munich: Otto Sagner, 123–135.
- Doleschal, Ursula. 1993. *Genus als grammatische und textlinguistische Kategorie: Eine kognitiv-funktionalistische Untersuchung des Russischen*. Unpublished doctoral dissertation, 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, England, 66–71.
- Evans, Roger and Gerald Gazdar. 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.
- 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. 1993. "It's not my Default: a Network Morphology Analysis of Noun Classes in Arapesh". Paper read at the Autumn Meeting of the Linguistics Association of Great Britain, Bangor, 13–15 September.
- Fraser, Norman M. and Richard A. Hudson. 1992. "Inheritance in Word Grammar". *Computational Linguistics* 18, 133–158.
- Gazdar, Gerald. 1987. "Linguistic Applications of Default Inheritance Mechanisms". In P. Whitelock, M. M. Wood, H. L. Somers, R. Johnson and P. Bennett (eds.), *Linguistic Theory and Computer Applications*. London: Academic Press, 37–68.
- Gazdar, Gerald. 1990. "An Introduction to DATR". In Roger Evans and Gerald Gazdar (eds.), *The DATR Papers* (= Cognitive Science Research Paper CSRP 139), School of Cognitive and Computing Sciences, University of Sussex, 1–14.
- Gazdar, Gerald. Forthcoming. "Ceteris paribus". To appear in J. A. W. Kamp and C. Rohrer (eds.), *Aspects of Computational Linguistics*. Berlin: Springer.
- Gibbon, Dafydd. 1992. "ILEX: A Linguistic Approach to Computational Lexica". In Ursula 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.
- Halle, Morris. 1992. "The Latvian Declension". In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 33–47.
- Hubenschmid, Markus. 1993. *Genus und Kasus der russischen Substantive: Zur Definition und Identifikation grammatischer Kategorien* (= Vorträge und Abhandlungen zur Slavistik 21). Munich: Otto Sagner.

- Iloa, Eeva and Arto Mustajoki. 1989. *Report on Russian Morphology as it appears in Zaliznyak's Grammatical Dictionary* (= Slavica Helsingiensia 7). Helsinki: Department of Slavonic Languages, University of Helsinki.
- Lazova, M. V. (ed.). 1974. *Obratnyj slovar' russkogo jazyka: okolo 125 000 slov*. Moscow: Sovetskaja Ėnciklopedija.
- Mučnik, I. P. 1971. *Grammatičeskie kategorii glagola i imeni v sovremennom ruskom literaturnom jazyke*. Moscow: Nauka.
- 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*, 131–136.
- Schupback, R. D. 1984. *Lexical Specialization in Russian* (= UCLA Slavic Studies 8). Columbus, Ohio: Slavica.
- Shieber, Stuart M. 1987. "Separating Linguistic Analyses from Linguistic Theories". In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod L. Johnson and Paul Bennett (eds.), *Linguistic Theory and Computer Applications*. London: Academic Press, 1–36.
- Smirnova, G. A. 1979. "Kategorija roda nesklonjaemyx suščestvitel'nyx". In V. P. Grigor'ev (ed.), *Lingvistika i poëtika*. 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: Slavica, 666–681.
- Thelin, Nils B. 1975. *Notes on General and Russian Morphology* (= Studia Slavica Upsaliensia 15). Stockholm: Almqvist & Wiksell.
- Timberlake, Alan. 1993. "Russian". In Bernard Comrie and Greville G. Corbett (eds.), *The Slavonic Languages*. London: Routledge, 827–886.
- Weiss, Daniel. 1991. "Sexus Distinctions in Polish and Russian". In Maciej Grochowski and Daniel Weiss (eds.), *Words are physicians for an ailing mind* (= Sagner 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.
- Zaliznjak, A. A. 1977. *Grammatičeskij slovar' russkogo jazyka: slovoizmenenie*. Moscow: Russkij jazyk.

(Corbett)

Dept. of Linguistic and International Studies
University of Surrey
Guildford, Surrey GU2 5XH
England
e-mail: g.corbett@surrey.ac.uk.

(Fraser)

Vocalis Ltd.
Chaston House
Mill Court
Great Shelford
Cambridge CB2 5LD
England
e-mail: norman@vocalis.demon.co.uk

Evaluative Affixes in Italian

DONNA JO NAPOLI and BILL REYNOLDS

I. INTRODUCTION

Italian nouns and adjectives make use of almost two dozen evaluative affixes (to borrow Scalise's 1984 term), most of which can also be found on verbs. While a few of these affixes are strongly productive on N and A and weakly productive on V, today most of those that occur on V are lexicalized (though not all; see Cortelazzo & Cardinale 1989). Nevertheless, their evaluative sense is obvious, and their occurrence suggests a period in the history of Italian (from around 1300 to 1600, judging by Cortelazzo & Zolli 1979) when all these affixes were productive. Since prepositions in Italian form an inert class with respect to morphology (with the exception of portmanteau preposition-articles, as in Napoli & Nevis 1987) and always have in the history of Italian, it is natural that these affixes did not occur on P. Thus, evaluative suffixes did occur on all morphologically relevant categories. This means that most evaluative affixes went through a period when they did not select for category.

Aronoff claims that word formation rules (WFRs), such as affixation, operate on a base that is "always specified syntactically" (1976: 47). The general syntactic and semantic conditions that a WFR might be sensitive to are "category, subcategory, selection, and lexically governed entailment and presupposition" (1976: 48). The specification is, furthermore, always unique. Aronoff dubs this claim the Unitary Base Hypothesis and argues that if any WFR can be found to apply to both nouns and verbs, for example, the UBH would be refuted, given that N and V cannot be stated as a single syntactic category. In support of the UBH, Aronoff looks at a potential counterexample, the English affix *-able*, which attaches to both N (*sizable*) and V (*readable*), and argues that the identity here is only apparent – that we have, in fact, two distinct affixes with different properties. Aronoff does not admit the possibility of an affix which would put no syntactico-semantic conditions on the base to which it attaches.

The evaluative suffixes studied here, which definitely attach to N as well as V, provide a clear counterexample to Aronoff's UBH, even as modified in Scalise (1984), as we discuss below (see Section 8).

Still, as Zwicky & Pullum (1987) have pointed out, "plain" morphology contrasts with "expressive" morphology in a number of ways, including precisely the fact that expressive morphology is promiscuous with respect to the base it operates on. By *plain morphology* they mean morphological rules that belong properly to the rules of grammar of natural languages. By *expressive morphology* they mean extragrammatical phenomena in word formation,