DOCUMENT DELIVERY ARTICLE

Journal Title: Journal of Linguistics

Article Author: Corbett, Greville G and Norman M. Fraser

Article Title: Corbett, Greville G. and Norman M. Fraser (1993) Network morphology: A DATR account of Russian nominal inflection.

Volume:

Issue: 29

Month/Year: 1993

Pages: 113-142

Imprint:

Trans. #: 309065

CUSTOMER INFORMATION:

HEIDI B HARLEY
PO BOX 210028
TUCSON, AZ 85721-0028
hharley@email.arizona.edu

STATUS: Faculty
DEPT: Linguistics

Network Morphology: a DATR account of Russian nominal inflection

GREVILLE G. CORBETT AND NORMAN M. FRASER

University of Surrey[1]

1. INTRODUCTION

In this paper we introduce a declarative approach to inflectional morphology, which we call Network Morphology, using the lexical representation language DATR. We show that we can account for a range of (Russian) data, for which previously various rule types were required, and can provide a more satisfying analysis than was previously available. First we outline the essential data (section 2), highlighting the problems they present. Section 3 introduces the basic tenets of Network Morphology. This draws heavily on DATR, which we present in outline in section 4. Next we reconsider the Russian declensional classes from this new perspective (section 5). We show how the approach described overcomes long-standing problems in an elegant fashion; the complexity of the data suggests that the approach adopted has implications well beyond Russian. We then tackle the complex problem of animacy in Russian, which exemplifies interesting regularities extending across declensional classes (section 6).

In the body of the paper, we draw out the parts of the analysis which are of greatest linguistic interest, and background the formalism. The specific parts discussed are identified so as to allow the reader to find them in their places in the full account (given as Appendix I). Since the analysis is expressed in the formally explicit DATR language, for which compilers (computer interpreters) are available, we are also able to supply the output (Appendix II), which demonstrates that the predictions made are indeed correct.

2. THE DATA

The inflectional morphology of Russian is complex: in nominal morphology six cases and two numbers are distinguished. The complexity is compounded by the fact that, instead of each possible combination of features being represented by a single form, there are various patterns of neutralization, some of which, as we shall see, extend across declensional classes. Consider first the basic data on the noun declensional classes, given in phonemic transcription[2] in Table 1.

We have presented four declensional classes. This is not the traditional account; most descriptions recognize only three, treating *zakon* and *v'ino* as variants of a single declensional class (as in, for instance, Vinogradov, Istrina & Barxudarov 1952, Unbegaun 1957 and Stankiewicz 1968). But there is usually no argumentation as to why precisely three declensional classes should be recognized for Modern Russian. Isačenko (1962: 87) is a little more forthcoming; he suggests that the main criterion for recognizing a group of nouns as forming a separate declensional class should be productivity. Yet he too treats *zakon* and *v'ino* as belonging to the same declensional class, even though both are members of productive groups. Indeed the four types listed are all productive (though it should be said that the productivity of the *v'ino* type is largely restricted to its soft variant which gains new verbal nouns in *-an'ij-o* and *-en'ij-o*, while the productivity of the *kost'* type depends on the suffix *-ost'* used to derive abstract nouns from adjectives). Zaliznjak (1967: 205–207) on the other hand proposes just two declensional classes (he adds the *komnata* type to *zakon* and *v'ino*, and derives the differences from gender differentiation). The number who have suggested four declensional classes is rather small (for instance, Karcevskij 1932:

---

[2] The following automatic phonological correspondences are assumed:

1 /i/ is retracted to its allophone [ɨ] after non-back hard (unpalatalized) consonants. Thus the nominative plural form /zakoni/ will be realized with [ɨ] 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 realized 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 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'o/ '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).

	zakon 'law'	komnata 'room'	kost' 'bone'	v'ino 'wine'
SINGULAR				
nom(inative)	zakon	komnata	kost'	v'ino
acc(usative)	zakon	komnatu	kost'	v'ino
gen(itive)	zakona	komnati	kost'i	v'ina
dat(ive)	zakonu	komnate	kost'i	v'inu
inst(rumental)	zakonom	komnatoj	kost'ju	v'inom
loc(ative)	zakone	komnate	kost'i	v'ine
PLURAL				
nom	zakoni	komnati	kost'i	v'ina
acc	zakoni	komnati	kost'i	v'ina
gen	zakonov	komnat	kost'ej	v'in
dat	zakonam	komnatam	kost'am	v'inam
inst	zakonam'i	komnatam'i	kost'am'i	v'inam'i
loc	zakonax	komnatax	kost'ax	v'inax
	I	II	III	IV

Table 1

Major noun declensional classes of Russian

Notes:

(a) Forms are given in phonemic transcription. Palatalization (or 'softening') is indicated by '.

(b) There is no overt ending in the nominative/accusative singular in declensional classes I and III, nor in the genitive plural of declensional classes II and IV.

(c) Complications induced by animacy are discussed in section 6.

65–66, and Corbett 1982). The main argument of Corbett (1982) is that if four declensional classes are postulated, the gender of Russian nouns can generally be derived from other lexical information (it need not be specified as an additional feature in the lexicon). If on the other hand, three declensional classes are postulated, it is not possible to derive gender (and equally, if gender is specified in the lexicon, then it is not possible to derive the declensional class from it). Given the evident differences between the nouns we treat as belonging to declensional classes I and IV, and the facts of gender assignment just noted, there is a strong case for recognizing four declensional classes. As we shall see in section 5, a more subtle approach is now available.

Given data of this type, it is natural to consider approaches to inflection which place special importance on the declensional or conjugational class – most notably the Word and Paradigm framework (Robins 1959; Matthews 1972) and Extended Word and Paradigm framework (Thomas-Flinders 1981; Anderson 1982). In these the relationship between a word and its

paradigm is fairly clear, at least for words which inflect regularly. Thus, in Russian, *zakon* belongs to declensional class I, and takes the inflectional endings appropriate to that declensional class. What is less clear is what kind of relationship, if any, holds between different declensional classes and, indeed, how the declensional class relates to other categories in the grammar. Zwicky's (1985) **rules of referral** represent an important step in this investigation. (See also Stump's (1991, forthcoming) formalization of the notion of referral in the framework of Paradigm Function Morphology.)

3. NETWORK MORPHOLOGY

In this section we introduce Network Morphology, a framework for describing inflection which offers a formally explicit account of lexical entries, declensional classes and word classes, and the relationships between and among these categories. In particular, it offers an explicit unified account of inflectional regularities, sub-regularities and exceptions. Few, if any, of the insights of Network Morphology presented here are new. The main inspiration of Network Morphology is clearly DATR, but Network Morphology abstracts away from the fine detail of Evans and Gazdar's DATR formalism (described in section 4).[3] Our motivation for this is twofold. First, we wish to focus on the linguistic insights, which are primary, rather than the formalism, which is secondary. There is no reason in principle why the same insights could not be encoded in some superficially different formalism (such as the ELU formalism of Russell *et al.* (1992) or the Word Grammar formalism of Fraser & Hudson (1992)). Second, one of the criticisms levelled against some approaches to morphology (such as Word and Paradigm) is that they are too powerful (Spencer 1991: 52). A key aim of Network Morphology is to identify a set of universal constraining principles of morphology. This is why we wish to keep it conceptually separate from the DATR formalism which, while elegant and suggestive, is also extremely powerful.[4]

Network Morphology rests on the following assumption (compare Hudson 1984: 1):

Networks
Lexical information is organized as a network whose basic elements are nodes and facts, and whose structure consists of relationships between basic elements.

---

[3] This strategy is analogous to Gibbon's work on ILEX, which 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).

[4] Moser has shown (1992) that DATR is equivalent in expressive power to a Turing machine; it places no constraints on the kind of theory which may be expressed in it.

A node is a named location in a network. At its simplest, a fact is an association between an attribute and a value.[5] All facts must be located at some node.

Nodes

A node is a named location at which one or more facts may be stored.

Figure 1 presents a graphic representation of the arrangement of a fragment of the Russian nominal data in this framework. Here, 'ZAKON' and 'CLASS I' are nodes. Two facts are spelled out at the ZAKON node, though more are possible and, indeed, likely given the Russian data. The value of the ⟨stem⟩ attribute is 'zakon'; that of the ⟨gloss⟩ attribute is 'law' (the latter may be thought of as no more than a place-holder for the lexical semantics). By convention, attributes are enclosed in angle brackets. Thus, facts about the stem and gloss of this word (and probably other facts, too) are stored in the network at the ZAKON node. In the diagram, the ZAKON node is connected by a link to the CLASS I node. Intuitively, this link may be taken to signify the fact that *zakon* belongs to Declensional Class I, though we shall explore this further below.

CLASS I	
<nom sg>	<stem>
<gen sg>	<stem> + a
...	...

ZAKON	
<stem>	zakon
<gloss>	law
...	...
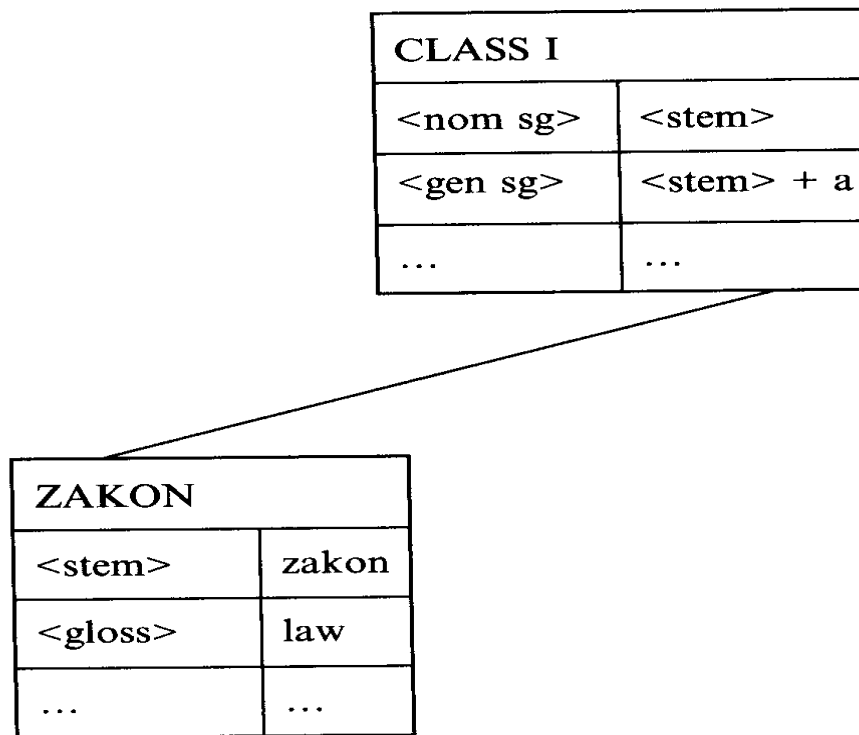
Figure 1
A fragment of the Russian nominal network

[5] Attribute-value representations have been widely used in computational linguistics since their introduction by Yngve (1958). Information is encoded by associating an attribute, such as number, with a value, such as singular, in attribute:value pairs. A more linguistically oriented presentation can be found in Pollard & Sag (1987).

Two facts are also shown at the CLASS I node, though more are possible and, indeed, probable. The first fact relates an attribute with another attribute. This appears to violate our earlier assertion that facts are defined to consist of attribute:value pairs. The following definition shows how the apparent contradiction can be resolved.

Facts
A fact consists of an attribute:value pair. A value may be stated directly or referenced indirectly by means of another attribute having that value. Chains of reference may be arbitrarily long, though a single attribute may appear only once in any chain. If, at the end of a chain of reference, no value can be found for an attribute, the fact in which that attribute appears is undefined.
Attributes
An attribute may be atomic or it may consist of a list of atoms. List attributes are descriptions which increase in specificity from left to right.
Values
Values may be atomic or list-structured, where a list consists of a sequence of atoms.

Thus, the first fact given for CLASS I in Figure I has ⟨nom sg⟩ as its attribute and, where a value ought to be, the name of another attribute – ⟨stem⟩ – appears. Thus, it is necessary to look for a fact consisting of the attribute ⟨stem⟩ and some value, and to take that value as the value of ⟨nom sg⟩ here. But where may we look?

An obvious place to start is at the same node. As we shall see below, it is possible to refer explicitly to facts stored at other nodes. However, unless such references to other nodes are supplied explicitly, only attributes available at a given node may be used for indirectly referencing values. Thus, the ⟨stem⟩ attribute in the first fact given at CLASS I cannot be used to reference indirectly the value 'zakon' in the first fact stored at the ZAKON node.

Further explanation of list-structured attributes is in order. Intuitively, what is intended here is that a short path (such as ⟨nom⟩) may be used to state a broad generalization (for example, 'all nominatives have some value'). A longer path increases the specificity of the generalization (for example, ⟨nom sg⟩ has the interpretation 'just those nominatives which are also singular have this value'). It is because of this list-structuring that attributes may also be referred to as **paths**.

So much for nodes and facts. But what is the benefit of using a *network* representation, and what does the link between ZAKON and CLASS I in Figure I signify? Improbable though it may seem, the answer to this is best understood by means of a digression into the world of pachyderms and, particularly, of Clyde the elephant.

Suppose person A tells person B that Clyde is an elephant. Person B ought

immediately to be able to infer the following facts (and probably a lot more besides).

(1) Clyde has a trunk
     Clyde is grey
     Clyde has thick skin

What has allowed multiple inferences to be drawn like this is a rule of inference which allows anything which is true of elephants also to be true of Clyde, since Clyde is an elephant. In fact, this description of the rule is too simple as it stands. Suppose person A further tells person B that Clyde is a pink elephant; this conflicts with what is known about the usual colour of elephants. The rule of inference must be revised to state that everything which is true of elephants is also true of Clyde, except for those facts about elephants which are blocked (that is, contradicted) by known facts about Clyde.

The first version of the rule is commonly referred to as **inheritance**; Clyde is said to inherit properties from elephant. The second version of the rule is called **default inheritance**; Clyde inherits properties from elephant only if those properties are not already specified for Clyde.

A graphic representation of a partial knowledge structure – or network – is shown in Figure 2. ELEPHANT inherits from PACHYDERM, therefore the facts stated at ELEPHANT may be augmented by the inherited fact that elephants have thick skin. Since CLYDE inherits from ELEPHANT, the following facts may

PACHYDERM	
<skin thickness>	thick

ELEPHANT	
<proboscis>	trunk
<colour>	grey
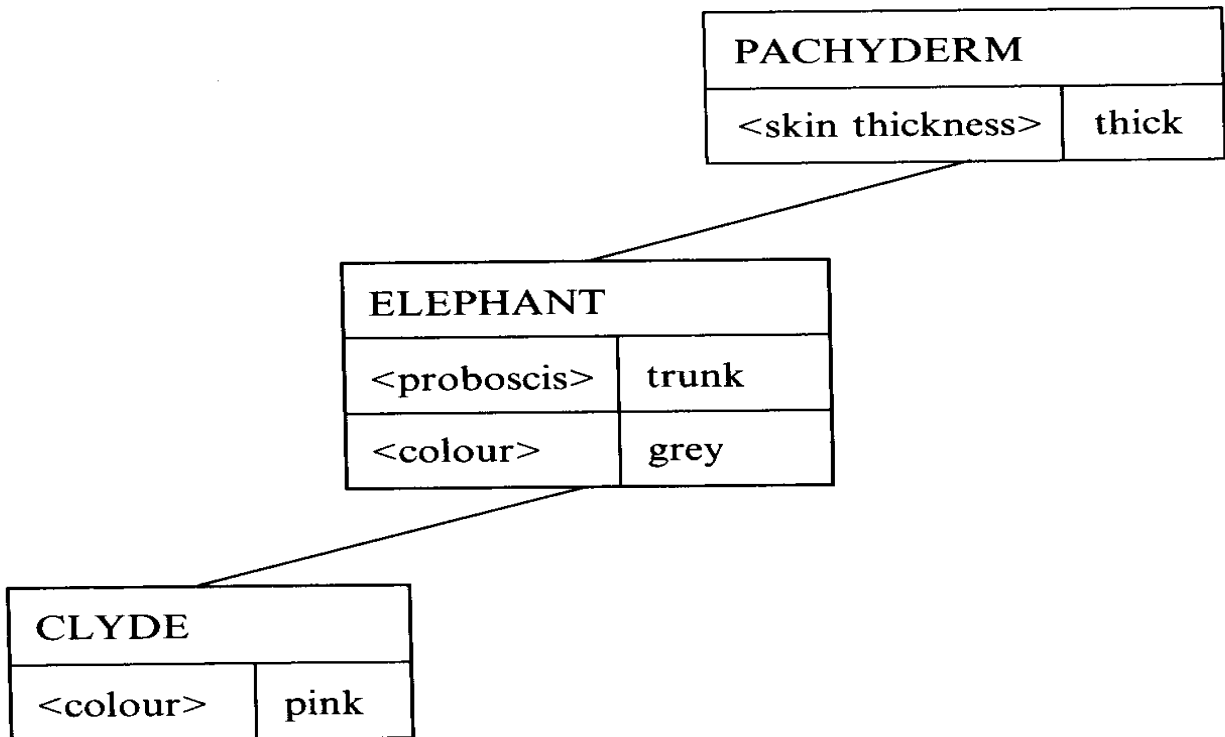
CLYDE	
<colour>	pink

Figure 2
The pachyderm network

be inferred: Clyde has a trunk, Clyde has thick skin. Note that the latter of these facts was inherited on the basis of a fact which was, itself, inherited. No fact 'Clyde is grey' is inherited, since the more specific fact 'Clyde is pink' blocks inheritance.

Default inheritance may be defined thus:

Default inheritance
If X and Y are nodes, X may inherit from Y if a fact identifying Y as an inheritance source is included at X. All attribute:value pairs at Y become available at X, except those having an attribute which is already present in an attribute:value pair at X.

Returning to our discussion of Figure 1, consider the relationship between ZAKON and CLASS 1. ZAKON inherits from CLASS 1. Thus, the fact that the nominative singular of *zakon* consists of the bare stem is inherited. Since the stem is also defined at this node, a value is defined for the ⟨nom sg⟩ attribute. Thus ⟨nom sg⟩ is defined at ZAKON, after inheritance, even though it was undefined at CLASS 1, where it was first stated. The information is useless at CLASS 1 because there it is no more than a schema. Only when it is inherited
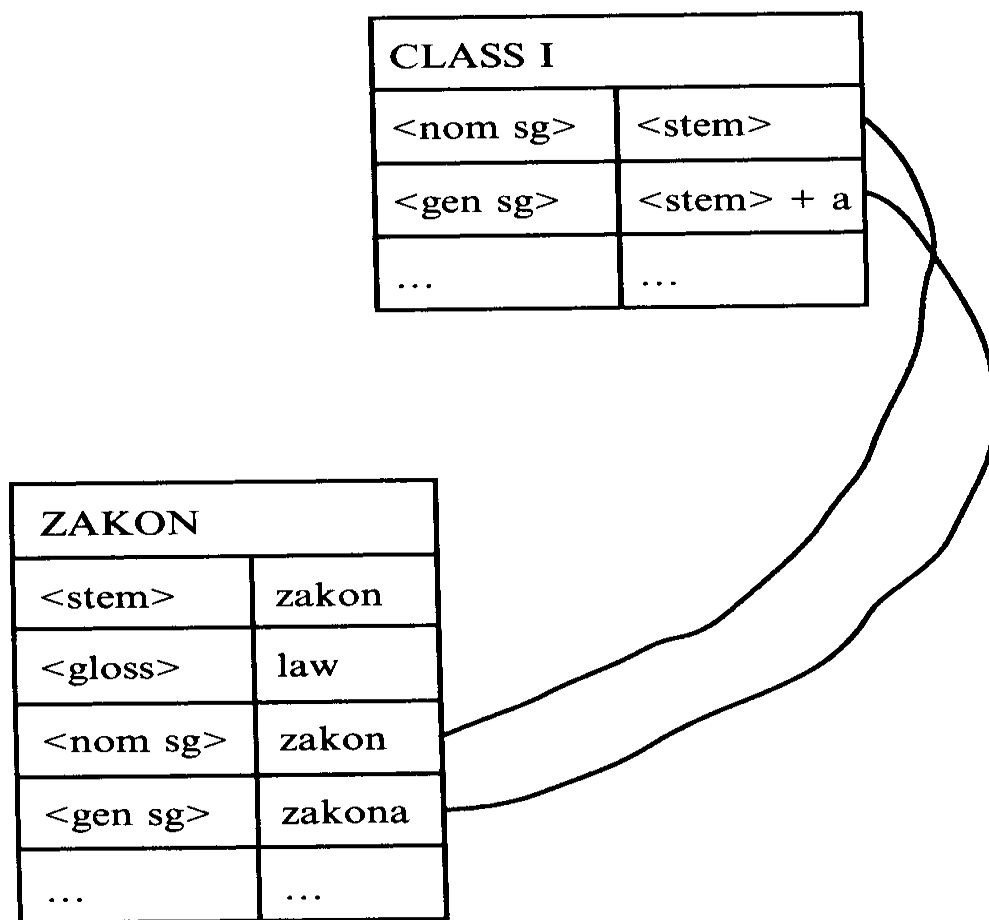
CLASS I	
<nom sg>	<stem>
<gen sg>	<stem> + a
...	...

ZAKON	
<stem>	zakon
<gloss>	law
<nom sg>	zakon
<gen sg>	zakona
...	...

Figure 3
ZAKON inheriting from CLASS I

by a lexical node which has the information to instantiate ⟨stem⟩ does it become useful. The point of expressing the information at the CLASS I node rather than at the ZAKON node is that any number of other lexical nodes may inherit the information from CLASS I, even though the generalization is stated only once. Figure 3 shows graphically the relationship between ZAKON and CLASS I.

Default inheritance underlies much work on knowledge representation (especially work on semantic nets) in the field of artificial intelligence (Fahlman 1979; Brachman 1985; Touretzky 1986).[6] It also underpins the 'object-oriented' family of computer programming languages (Stefik & Bobrow 1985). Default inheritance has found its way into theoretical linguistics through its central role in the theory of Word Grammar (Hudson 1990; Fraser & Hudson 1992). The idea has been around in computational linguistics for over a decade (Bobrow & Webber 1980), though interest in it has greatly increased recently (Flickinger, Pollard & Wasow 1985; Calder 1989; Daelemans & Gazdar 1992; Briscoe, Copestake & de Paiva forthcoming). Accessible introductions to default inheritance in linguistics and natural-language processing can be found in Gazdar (1987) and Daelemans, De Smedt & Gazdar (1992).

Default inheritance may seem intuitively simple: in fact, once the basic notion is accepted, a host of non-trivial issues needs to be addressed.[7] Let us consider another time-honoured example from the knowledge representation literature. Its concern is to find an inheritance network analysis which does justice to the following two propositions.

(2)   Nixon is a Republican
       Nixon is a Quaker

So far, we have only illustrated inheritance from a single node. However, a strong case can be made for allowing Nixon to inherit from both Republican and Quaker. The fact that Nixon is a Republican is sufficient to allow a picture of his broad political orientation to be built up; the fact that he is a Quaker is enough to allow the construction of a reasonably detailed picture of his socio-religious views. So long as there are no conflicts, this kind of multiple inheritance is unproblematic. In the language of this example, multiple inheritance is acceptable so long as political concerns do not overlap with socio-religious ones. However, formal definitions of multiple inheritance cannot make any such assumptions. Figure 4 shows what has come to be

---

[6] Clyde is something of a celebrity in the knowledge representation literature. For an introduction to the general field of knowledge representation in artificial intelligence see Brachman & Levesque (1985).

[7] Some of these are surveyed, and the relevant literature is cited, in Touretzky, Horty & Thomason (1987).
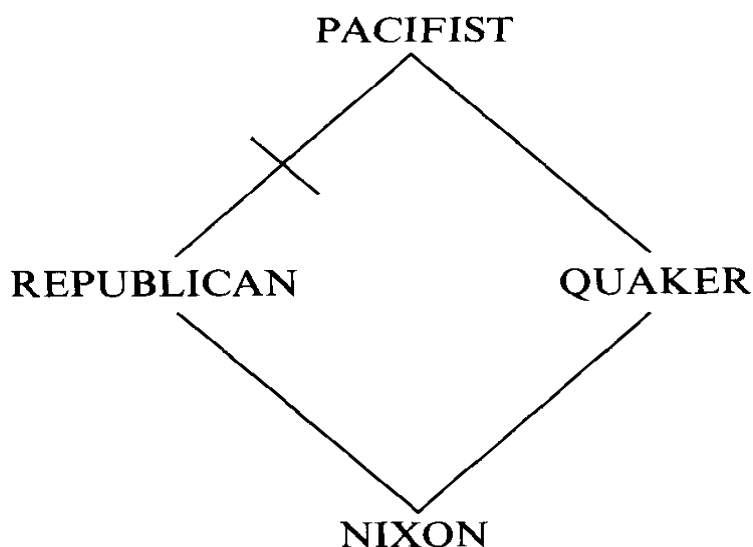
PACIFIST

REPUBLICAN   QUAKER

NIXON

Figure 4
The Nixon diamond

known as the Nixon Diamond, which illustrates the kinds of conflicts which may arise.

An inheritance link with a line through it – like the one linking REPUBLICAN with PACIFIST in Figure 4 – is a simple graphical shorthand indicating that the subordinate node may inherit NONE of the facts of the superior node. Thus a Republican is non-Pacifist, while a Quaker is solidly Pacifist. There is no actual conflict until a node is introduced which is subordinate to both REPUBLICAN and PACIFIST.

The solution adopted here is to insist that 'the most specific path always wins'. If the conflicting facts shown in (3) could potentially be inherited from different nodes, only (3b) will actually be inherited, since it is a specialization of (3a).

(3) (a)   NODE1 : $\langle abc \rangle$ == Value1
     (b)   NODE2 : $\langle abcd \rangle$ == Value2

If the facts to be inherited from different nodes have identical paths then NEITHER of them will be inherited. (Thus just the information represented in Figure 4 does not allow us to establish Nixon's views on war.) This is a version of what Touretzky calls 'orthogonal multiple inheritance' (1986: 73).

Orthogonal multiple inheritance
If, at a given node, some number of facts may potentially be inherited whose paths differ only in specificity, then if one path is more specific than any other, only it is inherited. If no single path is more specific than the others then none is inherited.

Network Morphology is a **declarative** system. It rests on a small inventory of basic operations (principally default inheritance) which operate on a body of static representations to produce an analysis. As we shall see below, there are

no **procedural** rules (such as rules of referral, or feature-change rules) to invoke when triggering patterns emerge. Network Morphology is thus part of the trend in recent years away from procedural accounts at all levels of grammatical theory. For example, approaches to phonology which stress declarative representations over procedural rules include Syntactic Phonology (Selkirk 1982, 1984), Government Phonology (Kaye, Lowenstamm & Vernaud 1985), Categorial Phonology (Wheeler 1988) and Event Phonology (Bird & Klein 1990). Most of the recent advances in declarative morphology have emerged from work in computational linguistics; witness, for example, Object-Oriented Morphology (Daelemans 1987), Paradigmatic Morphology (Calder 1989) and Lexeme-Based Morphology (Domenig 1989).

4. DATR

In this section we introduce a formal language within which Network Morphology theories may be expressed. DATR is a lexical knowledge representation language developed by Roger Evans and Gerald Gazdar (Evans & Gazdar 1989a, b; Gazdar 1990, forthcoming). The language provides a notation for expressing generalizations about lexical items, and for allowing these generalizations to apply to specific lexical items by means of default inheritance. Evans and Gazdar have also implemented a computer program capable of deriving relevant inferences from any hierarchically structured DATR lexicon (Evans 1990; Jenkins 1990).[8]

Knowledge is expressed in DATR in terms of path equations. The path equations we present in this paper take the forms shown in (4):

(4) (a) Node1 : $\langle\,\rangle$ == Node2
     (b) Node1 : $\langle$Path1$\rangle$ == Value1
     (c) Node1 : $\langle$Path1$\rangle$ == "$\langle$Path2$\rangle$"
     (d) Node1 : $\langle$Path1$\rangle$ == ("$\langle$Path2$\rangle$" Value1)
     (e) Node1 : $\langle$Path1$\rangle$ == Node2 : $\langle$Path2$\rangle$
     (f) Node1 : $\langle$Path1$\rangle$ == Node2

The form shown in (4a) is the special case in which a path at Node1 is empty. This allows Node1 to inherit all values available at Node2, except those which are overridden at Node1. The form shown in (4b) is used to assign values to paths, for example $\langle$infl_root$\rangle$ == komnat. Alternatively, a value may be indirectly referenced. (4c) is used to assign to Path1 whatever value is found for Path2 at the original query node. The double quotes are significant here because they indicate that Path2 must be evaluated globally. **Local evaluation** is restricted to a single node, while **global evaluation** has

---

[8] Other implementations exist. Those we know of are by A. Boltz (Konstanz), Norman Fraser (Surrey), Dafydd Gibbon (Bielefeld), James Kilbury (Düsseldorf), Hagen Langer (Bielefeld) and A. Sikorski (Poznań).

scope over the entire network, starting from the original query node. If the quotes were not present Path2 would be evaluated locally at Node1. A list of arbitrarily many values or paths may also be assigned to a path, as shown in (4d). For example, the following equation produces the instrumental singular form of a class II noun by suffixing 'oj' to the singular stem: ⟨mor inst sg⟩ == (" ⟨stem sg⟩ " _oj). The form shown in (4e) assigns to Node1: ⟨Path1⟩ whatever value is found at Node2: ⟨Path2⟩. A special case of this is (4f), which allows extensions of ⟨Path1⟩ to be specified at Node2.[9] It is path equations of forms (4e) and (4f) which are typically used in the Network Morphology correlate of rules of referral.

In our analysis of Russian nominal morphology, the proposed inheritance hierarchy is rooted in a NOMINAL node. Generalizations relating to nouns are collected at a NOUN node, beneath which nodes roughly corresponding to the noun declensional classes are located. We shall examine the shape of the inheritance network further below. By way of illustration, part of the entry for the NOUN node is expressed in DATR as shown in (5) below:

(5) NOUN: ⟨ ⟩ == NOMINAL
             ⟨mor loc sg⟩ == (" ⟨stem sg⟩ " _e)
             ⟨mor nom pl⟩ == (" ⟨stem pl⟩ " _i)
             etc.

The singular and plural stem of *komnata* 'room' is *komnat-*, therefore, by inheritance from NOUN, its locative singular form is *komnate* and its nominative plural form is *komnati* (orthographic *komnaty*). The DATR encoding for the third noun declension node (N_III) is shown in (6):

(6) N_III: ⟨ ⟩ == NOUN
             ⟨mor nom sg⟩ == N_I
             ⟨mor gen sg⟩ == N_II
             ⟨mor dat sg⟩ == " ⟨mor gen sg⟩ "
             ⟨mor inst sg⟩ == (" ⟨stem sg⟩ " _ju)
             ⟨mor loc sg⟩ == " ⟨mor dat sg⟩ ".

N_III default inherits from the NOUN node. The schema for forming nominative plural forms is common to three declensional classes and is inherited by N_III. The formation of the locative singular is common to most nouns, but not those in declensional class III. The final path equation in N_III blocks inheritance of the general rule from NOUN by specifying a local

---

[9] These path equations are all examples of DATR **definitional** sentences. These use the == symbol to associate with a node:path identifier (a left-hand side) either a value or a place in the network where a value may be found (a right-hand side). Thus the == symbol is used to describe an inheritance network. When we wish to express information extracted from a DATR network in accordance with its rules of inference we do so in **extensional** sentences using the = symbol. These always have a value as the right-hand side (Evans & Gazdar 1989a: 67). Numerous examples can be found in Appendix II.

neutralization. The instrumental singular is stated directly; all the other forms of nouns of this declensional class are either inherited from parent nodes (inherited via NOUN) or from elsewhere (other expressly stated nodes). The nominative singular is obtained from declensional class I, while the genitive singular rule is obtained from declensional class II (thus illustrating a use for multiple inheritance). The syncretism between genitive, dative and locative singulars is expressed by means of the fourth and sixth path equations.

In addition to the equation types we have seen so far, DATR offers the possibility of defining **evaluable paths.** These have the following form (where $N \geqslant 1$):

(7)   Node I: $\langle \text{PathX} \rangle == \langle$ "$\langle \text{Path I} \rangle$" ... "$\langle \text{PathN} \rangle$" $\rangle$

This equation can be used to create a path by evaluating separately the paths from "$\langle \text{Path I} \rangle$" to "$\langle \text{PathN} \rangle$" and concatenating the result values.[10] The following example shows this at work:

(8)   NOUN: $\langle \text{mor gen pl} \rangle ==$ "$\langle$ "$\langle \text{mor stem hardness} \rangle$" mor gen pl$\rangle$"
      $\langle \text{soft mor gen pl} \rangle ==$ ("$\langle \text{stem pl} \rangle$" _ej).

The genitive plural form of a noun depends on whether the stem is morphologically hard or soft (by default, stems which end in a phonologically soft consonant will be morphologically soft, but morphological softness must be specified in some instances). The first equation in (8) requires that the path $\langle \text{mor stem hardness} \rangle$ be globally evaluated. If the value of $\langle \text{mor stem hardness} \rangle$ is soft, the resulting path is $\langle \text{soft mor gen pl} \rangle$. The second path equation in (8) supplies the schema ("$\langle \text{stem pl} \rangle$" _ej) at this path. (The situation is actually more complex than this, involving other factors including stress; the additional complexities need not detain us here. It is for this reason that we do not give a lexical entry for *č'udov'išč'o* in our fragment in Appendix I, even though the neutralizations based on animacy are correctly predicted. For a full account of the genitive plural forms see Brown and Hippisley forthcoming.)

The Sussex implementation of DATR supplies a means of defining abbreviatory variables in the following fashion:

(9)   # vars $number: sg pl.

This defines a variable called $number which ranges over the values sg and pl. All and only variables in a DATR theory begin with the dollar character ($). Variables are purely a notational shorthand; they do not affect the expressive power of a DATR theory. We shall see an example of the use of a variable in section 6 below.

---

[10] Atoms may also appear together with paths on the right-hand side.

5. RUSSIAN DECLENSIONAL CLASSES: A NETWORK MORPHOLOGY PERSPECTIVE

Let us now consider again Russian nominal morphology from the perspective of Network Morphology, as encoded in DATR.[11] The traditional division of Russian nouns into declensional classes hides the fact that for all the nouns presented in Table 1 above, dative, instrumental and locative plural are identical. It is therefore misleading to claim, as in the traditional accounts, that, say, the dative plural of *zakon* is *zakonam* because *zakon* belongs to the first declensional class. Its dative plural is *zakonam* simply because it is a declinable noun. This fact is reflected in our account by treating these forms as a property of nouns rather than of individual declensional classes. Or rather, since adjectives have similar endings, though they are *-im*, *-im'i*, *-ix* and not *-am*, *-am'i*, *-ax*, the main part of the ending is lodged at a 'nominal' node, from which both nouns and adjectives inherit (see Appendix I: [2]; we present our DATR fragment in Appendix I, using numbered references to pick out salient points discussed in the text).

Let us go back to the disagreement as to the number of declensional classes to be recognized (as discussed in section 2). Some would treat *zakon* and *v'ino*
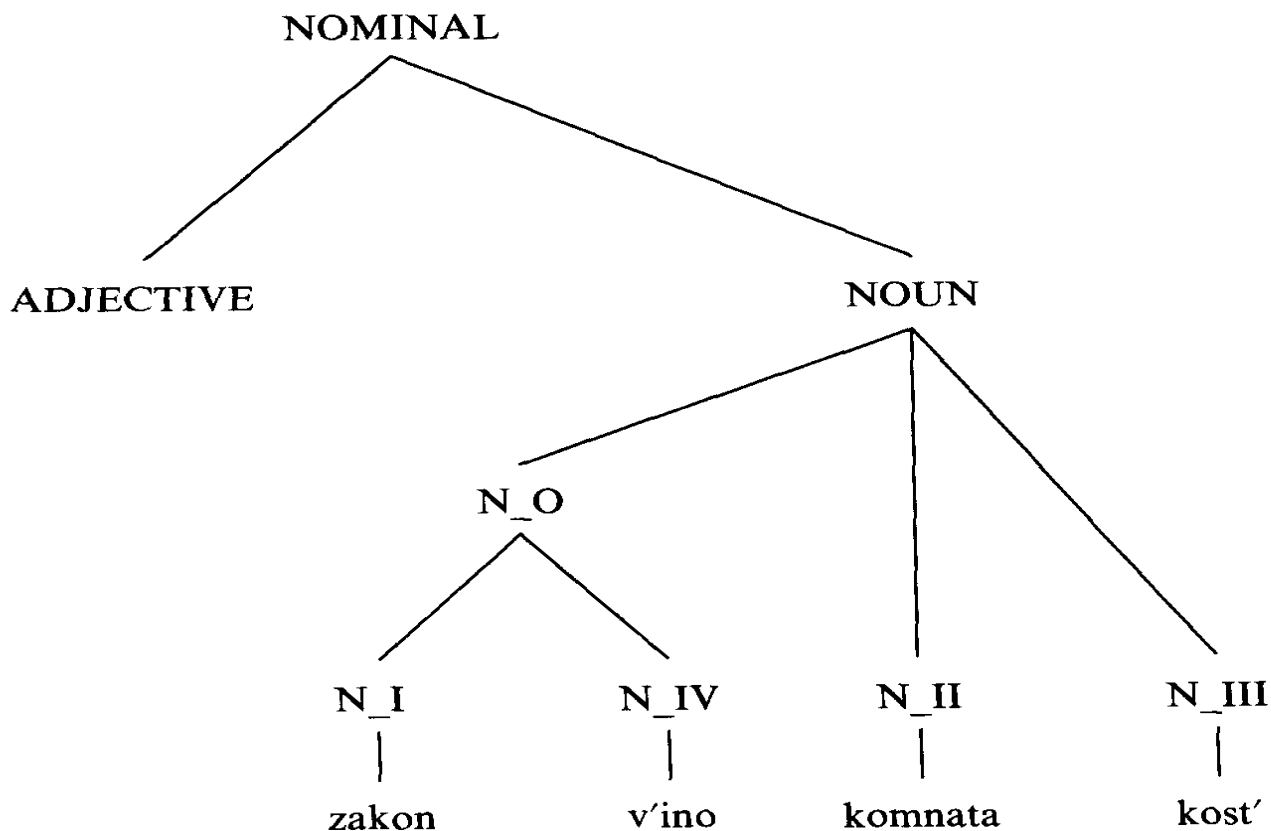

Figure 5
Inheritance hierarchy for Russian nominals

---

[11] For work on another Slavonic language, namely Slovene, using DATR see Erjavec (1992).

as belonging to the same declensional class, others do not. If we look back to the forms, there are indeed close similarities. On the other hand, they differ in nominative singular and nominative plural. The DATR account allows us to have it both ways. We set up a super-node, N_O, which looks after the singular oblique case forms which *zakon* and *v'ino* share. Both N_I and N_IV inherit from this node. This allows us to capture the fact that there are four main declensional classes, but that the differences between N_I and N_IV are not as great as those between either one of them and the other declensional classes.[12] This situation is shown graphically in Figure 5.

In this analysis, information about certain Russian nouns, those of the *zakon* (N_I) and *v'ino* (N_IV) types, is seen to belong to four different levels:

1 information shared with adjectives (for instance, the dative plural consisting of theme vowel[13] plus -*m*). This is lodged at the NOMINAL node.

2 information shared with most other nouns (for instance, the locative singular ending in -*e*). This is found at the NOUN node.

3 Information shared between the *zakon* type and the *v'ino* type (like the genitive singular ending -*a*). This is found at the node N_O.

4 information specific to the declensional class (for instance, that the nominative singular of nouns like *v'ino* is in -*o*). Such information is naturally recorded at the N_IV node. Information which is more idiosyncratic still belongs on yet lower nodes.

It is worth reflecting for a moment on what the node N_O represents. The obvious answer is that it represents a 'super-regularity', one shared by declensional classes N_I and N_IV. But this is a misleading description, if it suggests that the declensional class is supreme and that N_I and N_IV represent declensional classes in the old sense. Our analysis largely dissolves the traditional declensional classes, since declensional classes are parallel arrangements at a single level, while we have demonstrated an hierarchical arrangement across multiple levels. The information associated immediately with, say, node N_I is considerably less than in a traditional declensional class (only the nominative singular and genitive plural are specified at this node, the other case forms being inherited from parent nodes). The information at N_O does indeed represent a super-regularity in respect of

---

[12] This is to be compared with Carstairs' notion of 'macroparadigm' (1987: 67–70). Note, however, that we do not use gender as a determining factor; rather we claim that it is predictable from information available within the lexical entry, for the vast majority of nouns. We claim this generally, and with specific reference to Russian (Fraser & Corbett forthcoming).

[13] This is not a theme vowel in the traditional sense: it does not define declensional class membership and it does not extend throughout the declensional class. We are claiming only that certain plural endings can be segmented, and that when this is so we find -*a*- in noun endings and -*i*- in adjectival endings.

	student 'student' animate	zakon 'law' inanimate	uč'itel'n'ica 'teacher' animate	komnata 'room' inanimate	miš 'mouse' animate	kost' 'bone' inanimate	č'udov'išč'o 'monster' animate	v'ino[a] 'wine' inanimate
SINGULAR								
nom	student	zakon	uč'itel'n'ica	komnata	miš	kost'	č'udov'išč'o	v'ino
acc	studenta	zakon	uč'itel'n'icu	komnatu	miš	kost'	č'udov'išč'o	v'ino
gen	studenta	zakona	uč'itel'n'ici	komnati	miši	kost'i	č'udov'išč'a	v'ina
PLURAL								
nom	studenti	zakoni	uč'itel'n'ici	komnati	miši	kost'i	č'udov'išč'a	v'ina
acc	studentov	zakoni	uč'itel'n'ic	komnati	mišej	kost'i	č'udov'išč'	v'ina
gen	studentov	zakonov	uč'itel'n'ic	komnat	mišej	kost'ej	č'udov'išč'	v'in

Table 2

The morphological effect of animacy

[a] Orthographically these are: *student, zakon, učitel'nica, komnata, myš, kost', čudovišče, vino.*

N_I and N_IV, while equally representing a sub-regularity in respect of NOUN. Thus N_O is similar to NOUN and N_I, except that the latter two have (partly misleading) readily recognizable labels, while N_O has not. Our description allows us to see the old regularities, captured by the three-class approach (mentioned in section 2), and the new regularities too. Looking down from the top, Russian has three noun declensional classes (N_O, N_II and N_III); looking up from the bottom it has four: (N_I, N_II, N_III and N_IV).

6. ANIMACY

Perhaps the most interesting point about Russian nominal morphology is the role of animacy (see Comrie 1978, and for references to the extensive literature see Corbett 1988: n. 32). Table 2 shows that each of the declensional classes in fact exists in two variants (the remaining cases are identical between the two variants). The syncretic forms relevant to animacy are indicated by italic type.

It would be inadequate simply to list the forms. This would double the number of declensional classes and miss three major generalizations. The first is SEMANTIC: nouns with accusative–genitive syncretism are normally semantically animate (they denote animates: humans down to insects). The second is SYNTACTIC: animacy forms a subgender in Russian: agreeing modifiers differ for animate (10) and inanimate (11) nouns when in the accusative case:

(10)   pervovo (acc = gen)[14]   studenta (acc = gen)
       first   student
(11)   pervij (acc = nom)   zakon (acc = nom)
       first   law

However, consigning the problem entirely to the syntax, that is changing the case of the noun phrase to nominative or genitive, is not sufficient. This is demonstrated by those nouns which pattern according to declensional class II (the *komnata* type) but which are masculine animate, as is the case with *mušč'ina* 'man':

(12)   pervovo (acc = gen)   mušč'inu (acc)
       first   man

Returning to the morphology, we find that the third type of generalization is MORPHOLOGICAL. The crucial point is that we get the same regularity but in different declensional classes: the accusative matches the genitive for animate nouns under certain circumstances, and the nominative for

---

[14] Orthographically *pervogo*.

inanimate nouns, though the phonological form of the nominative or genitive varies from declensional class to declensional class.

An earlier analysis (Corbett 1981: 61) captured this informally by 'prediction rules' (following Perlmutter & Orešnik 1973), which 'borrow' the accusative form from elsewhere in the declensional class (rules of this type may be compared with Zwicky's 'rules of referral' 1985: 372).

Accusative prediction rules (ordered)
1 If there is an independent (that is, non-syncretic) accusative, it is selected.
2 For animates, the accusative is like the genitive.
3 For inanimates, the accusative is like the nominative.

For nouns like *student* 'student' and *zakon* 'law', the second and third rules give the desired effect. For nouns like *uč'itel'n'ica* and *komnata*, which have an independent (non-syncretic) accusative form, the first rule operates, irrespective of animacy. However, when we look at *miš* 'mouse' and *kost'* 'bone', and *č'udov'išč'o* 'monster' and *v'ino* 'wine', the situation is more complex: the accusative forms are all syncretic, and so we would expect them to come under the prediction rules. But *miš* 'mouse' and *č'udov'išč'o* 'monster' behave morphologically as though they are inanimate in the singular but animate in the plural, though clearly they are animate in semantic terms.

In attempting to give a more formal account, Corbett (1981) left the role of the prediction rules as given above (as we shall see, their formal equivalent was the feature-change rule). This is possible, if animacy is recognized as a semantic and as a syntactic feature, whose values need not match (just as sex and gender need not correspond). Copying rules copied features from the semantic to the syntactic characterization of lexical items, but with possible restrictions, including the following:

Copying restriction
Copy [+animate] only with [+masculine] or [+plural].

Of the nouns in Table 2, only *student* 'student' shows accusative–genitive syncretism in both singular and plural, and this is a masculine noun. Those which show accusative–genitive syncretism in the plural but not in the singular are feminine (like *miš* 'mouse') or neuter (like *č'udov'išč'o* 'monster'). Thus it can be seen that given the copying restriction, the prediction rules will indeed guarantee the correct accusative forms (and hence there is no need to double the number of declensional classes).

As mentioned earlier, these informal prediction rules can be expressed as 'feature-change' rules (following Dingwall 1969). Such rules, which change the value of certain feature combinations in certain environments, are extremely powerful, and most linguists would now wish to avoid them. The particular rules required are as follows:

Feature-change rules

1    $[+\text{animate}] \rightarrow [-\text{Rule 2}]$ / $\left[\left\{ \begin{array}{l} \left[\begin{array}{l} +\text{IInd declension} \\ -\text{plural} \end{array} \right] \\ \left[\begin{array}{l} +\text{adjective} \\ +\text{feminine} \end{array} \right] \end{array} \right\} \right]$

2    $\left[\begin{array}{l} +\text{accusative} \\ +\text{animate} \end{array} \right] \rightarrow [+\text{genitive}]$

The output of these rules was then the input to ordered inflectional rules (in the style of Bierwisch 1967), which rewrote the features as morphophonemic forms.[15]

In our account, the idea of the copying rule in its most general form is preserved (see Appendix I: [3]): syntactic animacy reflects semantic animacy:

(13)   NOUN: ...
        $\langle$syn animacy$\rangle$ == " $\langle$sem animacy$\rangle$ "

The main work done by the feature-change rules is given to default statements under NOMINAL (Appendix I: [1]):

(14)   NOMINAL: ...
        $\langle$acc$\rangle$ == " $\langle$mor nom$\rangle$ "
        $\langle$acc pl animate$\rangle$ == " $\langle$mor gen pl$\rangle$ "
        $\langle$acc sg animate masc$\rangle$ == " $\langle$mor gen sg$\rangle$ "
        $\langle$mor acc \$number$\rangle$ == $\langle$acc \$number
            " $\langle$syn animacy$\rangle$ " " $\langle$syn gender$\rangle$ "$\rangle$

These statements indicate that we expect the accusative to be as the nominative, unless the noun is animate and plural or masculine, when we expect it to be as the genitive. Note that the work of the restriction on animacy copying is done here in the new analysis. Note too that in order to account for the forms of the accusative we make reference to gender: the gender of individual nouns is itself accounted for in our analysis – unlike the earlier analysis with which we are comparing; this problem is dealt with in detail elsewhere (Fraser & Corbett forthcoming).

These defaults given in (14) are overridden in particular parts of individual declensional classes. For instance, under N_II there is the following (Appendix I: [4]):

(15)   N_II: ...
        $\langle$mor acc sg$\rangle$ == (" $\langle$stem sg$\rangle$ " _u)

This will override the defaults at (14) and ensure that nouns of this class have

---

[15] There was no need for any equivalent to the third prediction rule; the ordered inflectional rules assigned endings of decreasing markedness, assigning nominative (and accusative = nominative) endings last.

the accusative singular in *-u*, irrespective of animacy. Thus part of the work of the feature-change rule is rendered unnecessary in this way.

In working through this old problem again, we found a construction which could not be covered by the old account, but which we can now handle:

(16) I   vot   mat'   i   doč',   kotoryx vy uže   znaete[16]
      and here is mother and daughter whom you already know
      'And here are the mother and daughter whom you already know.'

In this construction, the relative pronoun must be marked as animate; it has the genitive = accusative form, and informants reject the accusative = nominative form **kotorye*. The problem for the earlier account is this: in the singular, *mat'* 'mother' and *doč'* 'daughter' are semantically animate but have accusative = nominative syncretism. This was handled by the copying restriction, which prevented semantic animacy being copied as a syntactic feature of animacy for feminine nouns. In the syntactic structure of (16), these nouns would be marked as inanimate, which normally gives the correct result for agreement purposes and for the morphology. The problem created by (16) is that by conjoining the noun phrases headed by these two nouns we can have plural modifiers, which are animate. In the old analysis, there is no animacy feature available to the agreement rule, and so there is no way to account for (16). In our present account, since features are not copied, changed or deleted, the animacy of these nouns is available to the syntax, specifically to an agreement rule, however this is formulated. Thus our account handles the morphological facts without leaving an impossible task for the syntax.

7. CONCLUSION

Russian inflectional morphology presents a set of particularly interesting problems; Network Morphology, as encoded in the formally explicit DATR language, sheds new light on these problems and more generally on problematic notions such as declensional class, as we have shown. DATR also offers the attractive possibility of computing the results of a given theory, so that it can be checked for accuracy (see the output in Appendix II). It is especially interesting to note that the forms involving animacy are correctly accounted for: our analysis allows us to cover the data discussed previously and it does so primarily by means of the simple device of default inheritance as compared to the much more complex mechanisms (copying rules with restrictions, feature-change rules and ordered inflectional rules) invoked earlier.

---

[16] Since this is a syntactic example we transliterate the standard orthography.

Authors' addresses: *Greville Corbett,*   *Norman Fraser,*
Department of Linguistic and International Studies, *Vocalis Ltd,*
University of Surrey,   *Chaston House,*
Guildford,   *Mill Court,*
Surrey GU2 5XH,   *Cambridge CB2 5LD,*
UK.   *UK.*

REFERENCES

Anderson, Stephen R. (1982). Where's morphology? *LIn* **13**. 571–612.

Bierwisch, Manfred (1967). Syntactic features in morphology: general problems of so-called pronominal inflection in German. In *To Honor Roman Jakobson I*. The Hague: Mouton. 239–270.

Bird, Steven & Klein, Ewan (1990). Phonological events. *JL* **26**. 33–56.

Bobrow, Robert J. & Webber, Bonnie Lynn (1980). Knowledge representation for syntactic/semantic processing. *Proceedings of the First Annual National Conference on Artificial Intelligence*. Stanford. 316–323.

Brachman, Ronald J. (1985). 'I lied about the trees.' Or, defaults and definitions in knowledge representation. *AI Magazine* **6**. 80–93.

Brachman, Ronald J. & Levesque, Hector (eds.) (1985). *Readings in knowledge representation*. San Mateo, CA: Morgan Kaufmann.

Briscoe, Ted, Copestake, Ann & de Paiva, Valeria (eds.) (1993). *Default inheritance within unification-based approaches to the lexicon*. Cambridge: Cambridge University Press.

Brown, Dunstan & Hippisley, Andrew (forthcoming). Conflict in Russian genitive plural assignment: a solution implemented in DATR. Paper read at the Spring Meeting of the Linguistics Association of Great Britain, University of Birmingham, 24 March 1993.

Calder, Jonathan (1989). Paradigmatic morphology. In *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*. Manchester, England. 58–65.

Carstairs, Andrew (1987). *Allomorphy in inflection*. London: Croom Helm.

Comrie, Bernard S. (1978). Genitive–accusatives in Slavic: the rules and their motivation. *International Review of Slavic Linguistics* **3**. 27–42.

Corbett, Greville G. (1981). Syntactic features. *JL* **17**. 55–76.

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. (1988). Gender in Slavonic from the standpoint of a general typology of gender systems. *Slavonic and East European Review* **66**. 1–20.

Daelemans, Walter (1987). Studies in language technology. An object-oriented computer model of morphophonological aspects of Dutch. PhD thesis, University of Leuven.

Daelemans, Walter & Gazdar, Gerald (eds.) (1992). Special issue on inheritance I & II. *Computational Linguistics* **18.2 & 18.3**.

Daelemans, Walter, De Smedt, Koenraad & Gazdar, Gerald (1992). Inheritance in natural language processing. *Computational Linguistics* **18**. 205–218.

Dingwall, W. O. (1969). Government, concord and feature-change rules. *Glossa* **3**. 200–240.

Domenig, M. (1989). Word Manager: a system for the specification, use, and maintenance of morphological knowledge. Habilitation thesis, University of Zürich.

Erjavec, Tomaž (1992). Treatments of Slovene verb morphology in inheritance models. Unpublished MSc thesis, University of Edinburgh.

Evans, Roger (1990). An introduction to the Sussex Prolog DATR system. In Roger Evans & Gerald Gazdar (eds.) *The DATR papers*. (Cognitive Science Research Report CSRP 139). University of Sussex. 63–71.

Evans, Roger & Gazdar, Gerald (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 & Gazdar, Gerald (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.

Fahlman, Scott E. (1979). Representing and using real-world knowledge. In Patrick H. Winston

& R. H. Brown (eds.) *Artificial intelligence: an MIT perspective*, vol. I. Cambridge, MA: MIT Press. 451–470.

Flickinger, Daniel, Pollard, Carl & Wasow, Thomas (1985). Structure-sharing in lexical representation. *Proceedings of the 23rd Meeting of the Association for Computational Linguistics*. Chicago. 262–267.

Fraser, Norman M. & Corbett, Greville G. (forthcoming). Russian gender assignment in DATR. Paper read at the Spring Meeting of the Linguistics Association of Great Britain, University of Birmingham, 24 March 1993.

Fraser, Norman M. & Hudson, Richard (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 & P. Bennett (eds.) *Linguistic theory and computer applications*. London: Academic Press. 37–68.

Gazdar, Gerald (1990). An introduction to DATR. In Roger Evans & 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 & 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*, 73). Stuttgart: Franz Steiner. 32–53.

Hudson, Richard (1984). *Word grammar*. Oxford: Blackwell.

Hudson, Richard (1990). *English word grammar*. Oxford: Blackwell.

Isačenko, A. V. (1962). *Die russische Sprache der Gegenwart*, vol. I: *Formenlehre*. Halle: Max Niemeyer.

Jenkins, Elizabeth A. (1990). Enhancements to the Sussex Prolog DATR system. In Roger Evans & Gerald Gazdar (eds.) *The DATR papers* (Cognitive Science Research Report CSRP 139). School of Computing and Cognitive Sciences, University of Sussex. 41–61.

Karcevskij, S. (1932). Sur la structure du substantif russe. In *Charisteria Guilelmo Mathesio Quinquagenario: a discipulis et circuli linguistici pragensis sodalibus oblata*. Prague: Cercle Linguistique de Prague. 65–73.

Kaye, J., Lowenstamm, J. & Vernaud, J.-R. (1985). The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook* **2**. 303–328.

Matthews, P. H. (1972). *Inflectional morphology*. Cambridge: Cambridge University Press.

Moser, Lionel (1992). Simulating Turing machines in DATR. Cognitive Science Research Paper CSRP 241. School of Cognitive and Computing Sciences, University of Sussex.

Perlmutter, David M. & Orešnik, Janez (1973). Language-particular rules and explanation in syntax. In Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*. New York: Holt Rinehart. 419–459.

Pollard, Carl & Sag, Ivan A. (1987). *Information-based syntax and semantics*, vol. I: *Fundamentals*. Stanford: CSLI.

Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society*. 116–144.

Russell, Graham, Ballim, Afzal, Carroll, John & Warwick-Armstrong, Susan (1992). A practical approach to multiple default inheritance for unification-based lexicons. *Computational Linguistics* **18**. 311–337.

Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Spencer, Andrew (1991). *Morphological theory*. Oxford: Blackwell.

Stankiewicz, Edward (1968). *Declension and gradation of Russian substantives in contemporary standard Russian*. The Hague: Mouton.

Stefik, Mark & Bobrow, Daniel G. (1985). Object-oriented programming: themes and variations. *AI Magazine* **6**. 40–62.

Stump, Gregory T. (1991). A paradigm-based theory of morphosemantic mismatches. *Language* **67**. 675–725.

Stump, Gregory T. (1993). On rules of referral. *Language* **69**.

Thomas-Flinders, Tracy (1981). Inflectional morphology: introduction to the extended word and paradigm theory. *Occasional Papers in Linguistics* **4**, UCLA Department of Linguistics.

Timberlake, Alan (1993). Russian. In Bernard Comrie & Greville G. Corbett (eds.) *The Slavonic languages*. London: Routledge. 827–886.

Touretzky, David S. (1986). *The mathematics of inheritance systems*. London: Pitman.

Touretzky, David S., Horty, John F. & Thomason, Richmond H. (1987). A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*. Milan. 476–482.

Unbegaun, Boris O. (1957). *Russian grammar*. Oxford: Oxford University Press.

Vinogradov, V. V., Istrina, E. S. & Barxudarov, S. G. (eds.) (1952). *Grammatika russkogo jazyka*, vol. I: *Fonetika i morfologija*. Moscow: AN SSSR.

Wheeler, Deirdre (1988). Consequences of some categorially-motivated phonological assumptions. In Richard T. Oherle, Emmon Bach & Deirdre Wheeler (eds.) *Categorial grammars and natural language structures*. Dordrecht: Reidel. 467–488.

Yngve, Victor H. (1958). A programming language for mechanical translation. *Mechanical Translation* 5. 25–41.

Zaliznjak, A. A. (1967). *Russkoe imennoe slovoizmenenie*. Moscow: Nauka.

Zwicky, Arnold (1985). How to describe inflection. In Mary Niepokuj, Mary Van Clay, Vassiliki Nikiforidou & Deborah Feder (eds.), *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: BLS, University of California. 372–386.

Appendix   I

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                      %
%    AUTHORS:      Greville Corbett and Norman Fraser                   %
%    DATE:         June 1993                                           %
%    DESCRIPTION:  A fragment of a Network Morphology account of        %
%                  Russian nominal morphology expressed in DATR         %
%                                                                      %
%    NOTE: ALL FORMS ARE IN PHONOLOGICAL TRANSCRIPTION                  %
%          AND NOT IN STANDARD ORTHOGRAPHY                              %
%                                                                      %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

#  vars $number: sg pl.


%  Material beginning with a '%' sign consists of comments.  These
%  are not read by the DATR compiler.

%  NOMINAL is the top node from which Nouns and Adjectives inherit.  By
%  default, both singular and plural stems are identical to the
%  inflectional root. Also by default, stems are phonologically hard and
%  morphological hardness matches phonological hardness. Paths
%  that begin '<acc...' are used to find a value for mor:acc:$number.
%  mor:acc:$number locally evaluates a path consisting of 'acc' followed
%  by the number, syntactic animacy, and gender of the Nominal, all
%  of which may be significant in determining the accusative form.  Form
%  assignments for dative, instrumental, and locative plurals are
%  straightforward; their position here at the top node testifies to
%  their generality.
%
NOMINAL:      <stem> == "<infl_root>"
              <phon stem hardness> == hard
              <mor stem hardness> == "<phon stem hardness>"
              <acc> == "<mor nom>"                                    [1]
              <acc pl animate> == "<mor gen pl>"
              <acc sg animate masc> == "<mor gen sg>"
              <mor acc $number> == < acc $number "<syn animacy>"
                                                "<syn gender>" >
```

```
<mor dat pl> == ("<stem pl>" "<mor theme_vowel>" _m)           [2]
<mor inst pl> == ("<stem pl>" "<mor theme_vowel>" _m´i)
<mor loc pl> == ("<stem pl>" "<mor theme_vowel>" _x).
```

```
%   The GENDER node is used to express the default relationships between
%   sex and gender. If sex is undifferentiated, the formal gender is
%   assigned, where formal gender is typically expressed at the level of
%   the noun type.
%
GENDER:      <male> == masc
             <female> == fem
             <undifferentiated> == "<formal gender>".
```

```
%   The next node encodes the broadest generalizations about Nouns which
%   do not apply to other nominals. The genitive plural form depends on
%   the (morphological) hardness of the stem. The appropriate form when
%   the stem is soft is given here. No such generalization can be made
%   when the stem is hard, so the assignments in those cases have to be
%   stated lower down in the hierarchy. The theme vowel which appears in
%   Nouns is /a/. By default, syntactic animacy is the same as semantic
%   animacy. Syntactic gender is set by reference to sex which, by
%   default, is undifferentiated.
%
NOUN:        <> == NOMINAL
             <mor loc sg> == ("<stem sg>" _e)
             <mor nom pl> == ("<stem pl>" _i)
             <mor gen pl> == "< "<mor stem hardness>" mor gen pl>"
             <soft mor gen pl> == ("<stem pl>" _ej)
             <mor theme_vowel> == _a
             <syn cat> ==  n
             <syn animacy> == "<sem animacy>"                     [3]
             <syn gender> == GENDER:< "<sem sex>" >
             <sem sex> == undifferentiated.
```

```
%   Nouns are assigned to declensional type nodes; these cannot be
%   mapped directly onto conventional declensional classes, since
%   hierarchical relations exist between type nodes. Since most
%   regularities are expressed at higher nodes much less information is
%   encoded at declensional type nodes than in conventional declensional
%   classes. Notice that each type has a formal gender.
```

```
%   Noun type N_O expresses those generalizations which are common to
%   types I and IV, which inherit from it.
%
N_O:         <> == NOUN   % traditional o-stems
             <mor gen sg> == ("<stem sg>"_a)
             <mor dat sg> == ("<stem sg>"_u)
             <mor inst sg> == ("<stem sg>"_om).
```

```
N_I:         <> == N_O
             <formal gender> == masc
             <mor nom sg> == "<stem sg>"
             <hard mor gen pl> == ("<stem pl>"_ov).
```

```
N_II:        <> == NOUN
             <formal gender> == fem
             <mor nom sg> == ("<stem sg>"_a)
             <mor acc sg> == ("<stem sg>"_u)                      [4]
             <mor gen sg> == ("<stem sg>"_i)
             <mor dat sg> == ("<stem sg>"_e)
             <mor inst sg> == ("<stem sg>"_oj)
             <hard mor gen pl> == "<stem pl>".
```

```
N_III:          <> == NOUN
                <formal gender> == fem
                <mor stem hardness> == soft
                <mor nom sg> == N_I
                <mor gen sg> == N_II
                <mor dat sg> == "<mor gen sg>"
                <mor inst sg> == ("<stem sg>"_ju)
                <mor loc sg> == "<mor dat sg>".


N_IV:           <> == N_O
                <formal gender> == neut
                <mor nom sg> == ("<stem sg>"_o)
                <mor nom pl> == ("<stem pl>"_a)
                <hard mor gen pl> == N_II.


%
%    EXAMPLE NOUN LEXICAL ENTRIES
%    (Thousands more lexical entries could be added without any need
%    to add to the theory.)

Zakon:          <> == N_I
                <gloss> == law
                <infl_root> == zakon
                <sem animacy> == inanimate.


%    In the next example morphological hardness (soft) differs from
%    phonological hardness (hard) and so is specified.  Given a full
%    phonological account, the discrepancy can be predicted from the
%    type of the final consonant.
%
Nož:            <> == N_I
                <gloss> == knife
                <infl_root> == nož
                <mor stem hardness> == soft
                <sem animacy> == inanimate.


 Student:       <> == N_I
                <gloss> == student
                <infl_root> == student
                <sem animacy> == animate
                <sem sex> == male.


%    In the next example, where the noun denotes a male, sex overrides the
%    formal gender for nouns of type II.
%
Mušč´ina:       <> == N_II
                <gloss> == man
                <infl_root> == mušč´in
                <sem animacy> == animate
                <sem sex> == male.


 Komnata:       <> == N_II
                <gloss> == room
                <infl_root> == komnat
                <sem animacy> == inanimate.


 Uč´itel´n´ica:     <> == N_II
                    <gloss> == female_teacher
                    <infl_root> == uč´itel´n´ic
                    <sem animacy> == animate
                    <sem sex> == female.


 Miš:           <> == N_III
                <gloss> == mouse
```

```
                  <infl_root> == miš
                  <sem animacy> == animate.

Kost´:            <> == N_III
                  <gloss> == bone
                  <infl_root> == kost´
                  <sem animacy> == inanimate.


V´ino:            <> == N_IV
                  <gloss> == wine
                  <infl_root> == v´in
                  <sem animacy> == inanimate.


%   An example of the default phonological hardness being overridden.
%
Mor´o:            <> == N_IV
                  <gloss> == sea
                  <infl_root> == mor´
                  <phon stem hardness> == soft
                  <sem animacy> == inanimate.


%   The ADJ node encodes generalizations about Adjectives which
%   do not apply to other nominals. Genders are distinguished so as to
%   allow Adjectives to agree with nouns of all genders. The theme vowel
%   for Adjectives is /i/. Syncretism in the accusative case is sensitive
%   to animacy.
%
ADJ:              <> == NOMINAL
                  <syn cat> == adj
                  <mor nom sg fem> == ("<stem sg>" _aja)
                  <mor nom sg neut> == ("<stem sg>" _ojo)
                  <mor nom sg> == ("<stem sg>" _ij)
                  <mor acc sg fem> == ("<stem sg>" _uju)
                  <mor gen sg fem> == ("<stem sg>" _oj)
                  <mor gen sg> == ("<stem sg>" _ovo)     % orthographic _ogo
                  <mor dat sg fem> == "<mor gen sg fem>"
                  <mor dat sg> == ("<stem sg>" _omu)
                  <mor inst sg fem> == "<mor gen sg fem>"
                  <mor inst sg> == ("<stem sg>" _im)
                  <mor loc sg fem> == "<mor gen sg fem>"
                  <mor loc sg> == ("<stem sg>" _om)
                  <mor theme_vowel> == _i
                  <mor nom pl> == ("<stem pl>" "<mor theme_vowel>" _e)
                  <mor gen pl> == "<mor loc pl>"
                  <animate mor acc> == "<mor gen>"
                  <inanimate mor acc> == "<mor nom>".


%
%   EXAMPLE ADJECTIVE LEXICAL ENTRY
%
Pervij:           <> == ADJ
                  <gloss> == first
                  <infl_root> == perv.
```

Appendix II

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                        %
%    AUTHORS:     Greville Corbett and Norman Fraser                     %
%    DATE:        June 1993                                              %
%    DESCRIPTION: Output of a DATR encoding of a Network Morphology      %
%                 account of Russian nominal morphology                  %
%                                                                        %
%    NOTE: ALL FORMS ARE IN PHONOLOGICAL TRANSCRIPTION                   %
%          AND NOT IN STANDARD ORTHOGRAPHY                               %
%                                                                        %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%%%%%%%%%%%%%%
%   NOUNS    %
%%%%%%%%%%%%%%

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.

Mušč´ina: <gloss> = man.
Mušč´ina: <mor nom sg> = (mušč´in _a).
Mušč´ina: <mor acc sg> = (mušč´in _u).
Mušč´ina: <mor gen sg> = (mušč´in _i).
Mušč´ina: <mor dat sg> = (mušč´in _e).
Mušč´ina: <mor inst sg> = (mušč´in _oj).
Mušč´ina: <mor loc sg> = (mušč´in _e).
Mušč´ina: <mor nom pl> = (mušč´in _i).
Mušč´ina: <mor acc pl> = mušč´in.
Mušč´ina: <mor gen pl> = mušč´in.
Mušč´ina: <mor dat pl> = (mušč´in _a _m).
Mušč´ina: <mor inst pl> = (mušč´in _a _m´i).
Mušč´ina: <mor loc pl> = (mušč´in _a _x).
Mušč´ina: <syn gender> = masc.
Mušč´ina: <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.

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

Miš: <gloss> = mouse.
Miš: <mor nom sg> = miš.
Miš: <mor acc sg> = miš.
Miš: <mor gen sg> = (miš _i).
Miš: <mor dat sg> = (miš _i).
Miš: <mor inst sg> = (miš _ju).
Miš: <mor loc sg> = (miš _i).
Miš: <mor nom pl> = (miš _i).
Miš: <mor acc pl> = (miš _ej).
Miš: <mor gen pl> = (miš _ej).
Miš: <mor dat pl> = (miš _a _m).
```

```
Miš: <mor inst pl> = (miš _a _m´i).
Miš: <mor loc pl> = (miš _a _x).
Miš: <syn gender> = fem.
Miš: <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.


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


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


%%%%%%%%%%%%%%%%
%   ADJECTIVES  %
%%%%%%%%%%%%%%%%%%%

% There are some practical difficulties in getting DATR to display all
% possible forms for contrasting values simultaneously.  This is of no
% theoretical consequence since, for example, an adjective will be
% animate or inanimate, but not both, in a given use.  The output which
% follows shows the results of specifying individually the different
% possible combinations.
```

```
Pervij: <gloss> = first.
Pervij: <mor nom sg masc> = (perv _ij).
Pervij: <mor nom sg fem> = (perv _aja).
Pervij: <mor nom sg neut> = (perv _ojo).
Pervij: <inanimate mor acc sg masc> = (perv _ij).
Pervij: <animate mor acc sg masc> = (perv _ovo).
Pervij: <mor acc sg fem> = (perv _uju).
Pervij: <mor acc sg neut> = (perv _ojo).
Pervij: <mor gen sg masc> = (perv _ovo).
Pervij: <mor gen sg fem> = (perv _oj).
Pervij: <mor gen sg neut> = (perv _ovo).
Pervij: <mor dat sg masc> = (perv _omu).
Pervij: <mor dat sg fem> = (perv _oj).
Pervij: <mor dat sg neut> = (perv _omu).
Pervij: <mor inst sg masc> = (perv _im).
Pervij: <mor inst sg fem> = (perv _oj).
Pervij: <mor inst sg neut> = (perv _im).
Pervij: <mor loc sg masc> = (perv _om).
Pervij: <mor loc sg fem> = (perv _oj).
Pervij: <mor loc sg neut> = (perv _om).
Pervij: <mor nom pl> = (perv _i _e).
Pervij: <inanimate mor acc pl> = (perv _i _e).
Pervij: <animate mor acc pl> = (perv _i _x).
Pervij: <mor gen pl> = (perv _i _x).
Pervij: <mor dat pl> = (perv _i _m).
Pervij: <mor inst pl> = (perv _i _m'i).
Pervij: <mor loc pl> = (perv _i _x).
```