1998

Indexed Stems and Russian Word Formation: A Network Morphology Account of Russian Personal Nouns

Andrew R. Hippisley
University of Kentucky, andrew.hippisley@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Follow this and additional works at: https://uknowledge.uky.edu/lin_facpub

Part of the Linguistics Commons

Repository Citation

Indexed Stems and Russian Word Formation: A Network Morphology Account of Russian Personal Nouns

Digital Object Identifier (DOI)
https://doi.org/10.1515/ling.1998.36.6.1093

Indexed stems and Russian word formation: a network-morphology account of Russian personal nouns[1]

ANDREW HIPPISLEY

Abstract

Recent lexeme-based models have proposed that a lexeme carries an inventory of stems on which morphological rules operate. The various stems in the inventory are associated with different morphological rules, of both inflection and derivation. Furthermore, one stem may be selected by more than one rule. For this reason stems in the inventory are labeled with indexes, rather than being directly associated with a particular morphological function. It is claimed that an indexed-stem approach captures generalizations in the morphological system that would otherwise be missed. We argue that such an approach provides for greater generalization in the Russian morphological system. One area of Russian derivation that particularly lends itself to an indexed-stem approach is the highly productive system of personal-noun formation. We present a declarative account of Russian personal nouns that assumes indexed stems and show how such an account on the one hand obviates the need to posit either compound suffixes or "concatenators" and on the other dispenses with truncating/deleting rules. The account is couched within network morphology, a declarative lexeme-based framework that rests on the concept of default inheritance and is expressed in the computable lexical knowledge representation language DATR.

Introduction

It has been proposed in a number of recent lexeme-based models such as Aronoff (1992, 1994), Stump (1998a), and Zwicky (1992: 335–336) that a lexeme carries not just one stem on which morphological rules operate, but an inventory of stems. We show how such an approach captures generalizations in the Russian derivational system that would otherwise be missed. Indexed-stem accounts fit naturally within declarative frameworks, and we demonstrate this by couching our account within the declarative framework of network morphology.

In section 1 we briefly look at the implications of a lexeme-based morphology, the approach presupposed by the notion of stem indexing. In section 2 the focus shifts to the phonological level of description of the lexeme, and the motivation for indexed stem inventories. In section 3 we turn to the area of the Russian morphological system we are investigating, Russian personal-noun formation. We look at a range of affixation rules that we argue operate on indexed stems. The framework within which our account is couched is outlined in section 4, and the account itself is presented in section 5. For the full theory and its theorems, the reader is referred to Hippisley (1997).

1.   Word formation as lexeme formation

In a lexeme-based model,[2] derivational morphology is the mapping of form and meaning at the level of the word, or more specifically the "lexeme." Lexemes are vocabulary items belonging to the major lexical categories of verb, noun, and adjective (as well as adverb) (Aronoff 1994: 10). The lexeme is an assembly of an item's phonology (form), syntax, and meaning. Thus any manifestation of an item is covered by the lexemic representation of that item. As such, the lexeme constitutes the minimal sign of the language. Inflectional morphology is the realization of morphosyntactic categories in that it is used in the spelling-out of the lexeme's collection of grammatical words into word forms; derivation, on the other hand, is the realization of derivational categories and is used to spell out "lexeme formation." More specifically, it concerns the change in shape of the lexeme's stem as it becomes the stem of another lexeme, and the corresponding change in the syntactic and semantic information content specified by the derivational category. We can take for example *dobr(ij)* 'kind' and its derivative *dobrot(a)* 'kindness'.[3] We can state the relationship as one between the base lexeme DOBRIJ and the derivative DOBROTA 'kindness'. Morphologically the stem change of /dobr/ to /dobrot/ corresponds to the syntactic change of adjective to noun, and the semantic change of 'kind' to 'quality of being kind'. This example of lexeme formation is shown in (1).

(1)   Base   →   Derivative

DOBRIJ $\begin{cases} \text{syntax:} \\ \quad \text{adjective} \\ \text{semantics} \\ \quad \text{'kind'} \\ \text{phonology (stem):} \\ \quad \text{/dobr/} \end{cases}$ DOBROTA $\begin{cases} \text{syntax:} \\ \quad \text{noun} \\ \text{semantics} \\ \quad \text{'quality of being} \\ \quad \text{kind'} \\ \text{phonology (stem):} \\ \quad \text{/dobrot/} \end{cases}$

From (1) we see changes at the syntactic, semantic, and phonological levels of the base. Because lexeme formation is broken down into three levels of change, associated with the three levels of description of the lexeme, it allows for the possibility of the information content of one of the levels to remain unaltered, corresponding to the types of derivation variously described as transposition, zero derivation, and (syntactic) category-preserving derivation.[4] In transposition, there is a change at the syntactic level of the lexeme, that is, change in syntactic category, but no (independent) change at the semantic level. A productive area of transposition in Russian is relational adjective derivation, such as *serebr'an(ij)* 'silver (adj)' from *serebr(o)* 'silver (noun)'.[5] Transposition is represented straightforwardly in a lexeme-based approach as in (2), where the information that is preserved is semantic.

(2) Base   →   Derivative

$$
\text{SEBERO} \begin{cases} \text{syntax:} \\ \quad \text{noun} \\ \text{semantics} \\ \quad \text{'silver'} \\ \text{phonology (stem):} \\ \quad \text{/serebr/} \end{cases}
\qquad
\text{SEREBR'ANIJ} \begin{cases} \text{syntax:} \\ \quad \text{adjective} \\ \text{semantics} \\ \quad \text{'silver'} \\ \text{phonology (stem):} \\ \quad \text{/serebr'an/} \end{cases}
$$

It would be perfectly feasible for a language to use lack of the alteration in the stem to encode a derivation. This is easily described in a lexeme-based model by stating that one level of the lexeme, the phonological level, has been preserved in the derivation. Zero derivation is a possible morphological operation type used in Russian derivation. Relational adjective derivation again represents a good example of this, such as *zolot(o)* 'gold' and its derivative adjective *zolot(oj)*. Preservation of the stem is represented in (3).

(3) Base   →   Derivative

$$
\text{ZOLOTO} \begin{cases} \text{syntax:} \\ \quad \text{noun} \\ \text{semantics} \\ \quad \text{'gold'} \\ \text{phonology (stem):} \\ \quad \text{/zolot/} \end{cases}
\qquad
\text{ZOLOTOJ} \begin{cases} \text{syntax:} \\ \quad \text{adjective} \\ \text{semantics} \\ \quad \text{'relating to} \\ \quad \text{gold'} \\ \text{phonology (stem):} \\ \quad \text{/zolot/} \end{cases}
$$

In examples such as these, the directionality needs some motivation, and

in this case we follow Isačenko (1969: 52–3), who uses the fact that, in the noun, stress is on the first syllable of the stem (*zólot[o]*), and in the adjective, the stress is on the ending (*zolot[ój]*). Since this type of stress shift is analogous to other affixal adjective derivation, for example *béreg* 'shore' > *beregov(ój)*, where the suffix *-ov* is being used to mark the derivation, *zolot(oj)* is a derived word based on the noun *zolot(o)*. Zero derivation in lexeme-based morphology can therefore be thought of as "omissive" morphology (see for example Beard 1995: 69).

A seemingly straightforward example of syntactic category preservation is the derivation of personal nouns from noun bases. For example *barabanšč'ik* 'drummer' is derived from the noun base *baraban* 'drum', and *mogil'šč'ik* 'gravedigger' from the noun base *mogil(a)* 'grave'. In cases such as these Isačenko notes that the nature of the semantic change is not completely clear. For example, as well as 'person who beats drums' *barabanšč'ik* can also denote 'person who makes drums'. In both there is the sense of 'person who relates to X', but the relationship is not specified by the derivation; hence the term "relators" that Isačenko gives for suffixes such as *-šč'ik* (1969: 54). An example of a "relator" in Czech is also given: in the derivative *vředář* derived from *vřed(y)* 'ulcers', the meaning is either 'person who relates to ulcers as a doctor' or 'person who relates to ulcers as a patient'.

As we have said, in zero derivation (or omissive morphology) it is the stem of the lexeme's stem that is preserved, that is, no change has taken place at the phonological level. In the next section we look more closely at this phonological level.

2.   Lexemes and indexed stems

Aronoff (1992: 14–16) notes that the formal part of a lexeme carries with it a number of notions that are important to disentangle. First, the "root" is the form that is left when all morphologically added structure has been "wrung out." Second, a lexeme's "citation form" is the special form used in lexicography as a place-holder or address which we can think of as the entire lexeme in shorthand. (For Russian nouns this is the nominative singular form.) Third, the "lexical representation" is the analogue in the mental lexicon of the citation form.[6] Fourth, and finally, the stem is "that form of a lexeme to which a given affix is attached or on which a given realization rule operates" (1992: 14). As we noted above, the realization may be inflectional or derivational. The stem is distinguished among the sound forms of a lexeme in that it alone is morphologically relevant: morphological rules act on the stem, and not on the other sound forms.

Furthermore, unlike the other sound forms, a lexeme may have more than one stem.

To illustrate this last point, consider the Russian noun *kot'onok* 'kitten', whose paradigm we give in Table 1. For this lexeme there are clearly two forms on which inflections may be realized. Moreover, the distribution of the two forms is systematic, since the /kot'at/ form only appears in the plural subparadigm and the /kot'onok/ form only in the singular subparadigm.

To capture first the fact that there are two forms, and second that these forms are systematically distributed in the way that they are, the lexeme is given an inventory of two stems, /kot'onok/ and /kot'at/. To ensure that the right stem is used for every morphosyntactic word, the stems are labeled such that the first is the one referred to in inflectional rules realizing singular and the second in inflectional rules realizing plural. The lexeme KOT'ONOK would be represented as in (4). The lexeme is complex and is made up of three levels of description. It has a syntactic level, where the syntactic category "noun" is specified, a semantic level, where for simplicity we have given the gloss rather than the precise lexical semantics, and a phonological level. This is given as an inventory of stems on which morphological rules operate.

(4)

$$\text{KOT'ONOK} \begin{cases} \text{syntax:} \\ \quad \text{noun} \\ \text{semantics:} \\ \quad \text{'kitten'} \\ \text{phonology (stem inventory):} \\ \quad \text{stem sg /kot'onok/; stem pl /kot'at/} \end{cases}$$

Note that an alternative would be to have only /kot'onok/ as the stem, and then a series of operations associated with the rule for plural inflections: first truncate the stem to yield /kot/; then attach the -*'at* formative

Table 1. *Paradigm of word forms of* kot'onok.

	Singular	Plural
nom	kot'onok	kot'at-a
acc	kot'onk-a	kot'at
gen	kot'onk-a	kot'at
dat	kot'onk-u	kot'at-am
inst	kot'onk-om	kot'at-am'i
loc	kot'onk-e	kot'at-ax

to yield /kot'at/,[7] then supply the affixes. The two approaches have been neatly summarized by Zwicky (1996) as a "trade-off between multiple operations and multiple stems." The multiple-stem approach will be seen to allow for greater generalization to be stated in the morphological system.

2.1. *The need for indexed stems*

The stems in (4) each have a special function attached to them, hence the labels "sg" and "pl." There are, however, situations where a function label appears to be inappropriate since the stem has more than one function. Aronoff (1992) claims that this is true for Latin verbs, and he argues that indices rather than functions should distinguish the set of stems that belong to a lexeme.

Latin verbs are traditionally treated as having a stem inventory, in the same way as KOT'ONOK in (4) above, since they display three distinct shapes throughout their paradigms. Taking the regular first-conjugation verbs *am(ō)* 'love' and *laud(ō)* 'praise', Table 2 shows how the stems are distributed among the present infinitive active, the perfect active, and the perfect passive participle.[8]

Given the sample of categories in Table 2, one may attempt to associate each stem with one particular function, in the same way as /kot'onok/ was associated with singular and /kot'at/ with plural above. This is even more appealing when the other categories are included. For example, only perfect active tenses take the stem in the middle column: perfect, pluperfect, future perfect (all in indicative and subjunctive), as well as perfect infinitive.[9] A one-stem-to-one-function mapping seems to be what holds here, if imperfect can be counted as a function. However, we hit upon a severe problem with the third column. Aronoff convincingly shows that the stem in the third column of Table 2 is multifunctional since, in addition to the perfect passive participle, it is used to realize other unrelated categories, namely future active participle and supine, as well as a number of derivational categories, for example agentives and abstracts. Since no sole function can be identified with this stem, Aronoff

Table 2. *Stems of* am(ō) *and* laud(ō)

Present active infinitive	Perfect active	Perfect passive participle
amā-re	amāv-i	amāt-us
laudā-re	laudāv-i	laudāt-us

proposes simply indexing it as the "third" stem, on which the appropriate inflectional and derivational rules will operate. The conclusion to be drawn from the above is that in the stem inventory of verbs, one stem, the third stem, is associated with multiple functions, which can be summarized in (5). In other words, stems may have multiple functions.

(5)

third stem
$\Big\{$
perfect passive participle (infl)
future active participle (infl)
supine (infl)
agent (deriv)
abstract (deriv)
intensive (deriv)
desiderative (deriv)
iterative (deriv)

In fact if a stem carries an index, then what characterizes it is not that it has many functions, but rather that it has no function at all (Aronoff 1992: 18). This is an example of "pure" morphology, or "morphology by itself" (Aronoff 1994).

3. Personal nouns and indexed stems

Russian personal-noun formation is highly productive. For example, a study of all new formations in the four-volume Ušakov dictionary (1935–1940), conducted by Likova (1959), found that of the 900 new formations, over one-third represent personal nouns.[10] There are numerous suffixes used to derive personal nouns. Cubberley (1994) identifies well over 50 such suffixes in the 1980 Academy Grammar (Švedova 1980). Out of these, the suffixes in Table 3 are generally recognized in the literature as the more productive ones.

Table 3. *Productive personal-noun-forming suffixes*

Suffix	Stem	Gloss	Personal noun	Gloss
-tel′	grab´i(t´)	'steal'	grab´itel′	'thief'
-´ist	traktor	'tractor'	traktor´ist	'tractor driver'
-´ik	frontov(oj)	'front-line'	frontov´ik	'front-line soldier'
-n´ik	vest′	'news'	vestn´ik	'herald'
-šč´ik	baraban	'drum'	barabanšč´ik	'drummer'
-(l′)šč´ik	r´isova(t´)	'draw'	r´isoval´šč´ik	'draughtsman'
-ec	skup(oj)	'stingy'	skupec	'skinflint'

Derivation using the affixes in Table 3 meets with a number of structural problems that a stem-indexed approach resolves. First, as will be seen in the following sections most of the suffixes have what appear to be complex alternants. For example, *-n'ik* has what appears to be an alternant *-'atn'ik* in the derivation *los'* 'elk' > *los'atn'ik* 'elk hunter'. One alternative analysis to complex suffixes posits a formative /'at/ intervening between the root and the deriving suffix, that is, a "concatenator" or "interfix" (*interfiks*). We will see that our approach is to view derivation taking place on a complex stem in /'at/. Second, the deverbal suffix *-tel'* is always preceded by a vowel (e.g. *grab'itel'*). We will show that this is naturally accounted for if we assume the "multiple-stem" approach to Russian verb inflection and allow it to carry over into derivation: the suffix *-tel'* selects a vocalic stem in the verb's inventory of stems, as we shall see. Finally, personal-noun derivation may involve truncation. This is the case, for example, with derivation in *-ec*. This suffix is highly productive in deriving nouns from relational adjectives in *-sk*, for example *makedonsk(ij)* 'Macedonian' > *macedonec* 'Macedonian person'. We will posit a special "truncated stem" in the stem inventory of the base on which derivation in *-ec* operates.

3.1. *"Concatenators" and indexed stems*

Most of the denominal plural suffixes in Table 3 appear to have complex alternants. For example, *-šč'ik* has the alternant *-ovšč'ik* in the personal noun *lesovšč'ik* 'forester', *-n'ik* has the alternant *-'atn'ik*, and *-ec* has the alternant *-ovec*. Examples are given in Table 4.

The complex, or compound, suffix analysis of the examples in Table 4 is the traditional one and the one assumed in the Academy Grammars. Vinogradov et al. (1953) has, for example, one section on suffixation in *-šč'ik* (section 290) and another on suffixation in *-ovšč'ik* (section 291). However, it was noted by Zemskaja (1964) and Shapiro (1967), among others, that this analysis misses generalizations since the alternant complex suffix performs the same function as the simple suffix and has the same distributional properties (Zemskaja 1964: 41; Shapiro 1967: 55, 56). The two authors independently posit a formative intervening between the root and the suffix, what Zemskaja terms an "interfix" and Shapiro calls a "concatenator." Both authors attribute such formatives in Russian morphophonemics to Trubeckoj's (1934: 14) *Verbindungsmorphemen*. Thus the analysis of *grobovšč'ik* 'coffin builder' would be *grob-ov-šč'ik* (where /ov/ is a concatenator).

Table 4. *Complex alternants*

Suffix	Gloss	Alternant	Gloss
les	'forest'	les-ovšč'ik	'forester'
starj(o)	'old clothes'	starj-ovšč'ik	'old clothes dealer'
skob(a)	'clamp, shackle'	skob-ovšč'ik	'clamp builder'
kalandr	'calander' (machine)	kalandr-ovšč'ik	'calander operator'
los'	'elk'	los'-atn'ik	'elk hunter'
medved'	'bear'	medvež-atn'ik	'bear hunter'
volk	'wolf'	volč-'atn'ik	'wolf hunter'
tel'onok	'calf'	tel-'atn'ik	'calf-herd'
kružok	'circle (study)'	kružk-ovec	'circle member'
vuz	'vuz'	vuz-ovec	'student of *vuz*'
rabfak	'workers' school'	rabfak-ovec	'*rabfak* student'

Apart from the complex suffix and concatenator approach, Shapiro mentions a third possible approach, which he later rejects, where the alternation is not found in the suffix but in the stem itself (Shapiro 1967: 55). For example, in *pervenec* 'first-born', derived from the adjective *perv(ij)*, the suffix *-ec* is attached to *perven-*, rather than the basic stem *perv-*. In other words, the basic stem *perv-* is alternating with the less-basic stem *perven-*. Importantly, alternation in stem in such cases is not, however, morphophonological, as Zemskaja is at pains to point out (1964: 36).[11] Instead, the stem alternant(s) would have to be lexically listed with the stem. This is in fact the approach taken in a later work in Isačenko (1969) to account for derivation from indeclinables, such as *kino > kinošn'ik* 'cinema buff', where the deriving noun is listed with two alternant stems *k'ino-* and *kinoš-* (1969: 64–65). Our approach follows on from Isačenko's. Lexemes at their phonological level have an inventory of related, but not necessarily morphophonologically related, stems, which are selected depending on the nature of the morphological rule.

Thus in our stem-indexed approach for the nouns in -*'atn'ik* in Table 4 we posit a separate stem in the formative /'at/ that is selected for derivation in -*n'ik*. For example *los'* 'elk' derives *los'atn'ik* 'elk hunter' with the formative -*'at* intervening between the root and the deriving suffix. The formative /'at/ is viewed as part of the stem. Examples of this are given in Table 4. A noun such as *los'* 'elk' has (at least) two stems in this analysis. One stem is the default, used in all inflection and the majority of derivation. We label this stem 0. It is a form that (by default) coincides with the root. The second stem, stem 1, is the root plus the /'at/ formative /los'at/. This is represented in (6). Thus for this lexeme the syntactic

level is specified with the category "noun," the semantic level with the item's gloss, and the phonological level with a set of two stems.

(6) LOS′ $\left\{\begin{array}{l} \text{syntax:} \\ \quad \text{noun} \\ \text{semantics:} \\ \quad \text{'elk'} \\ \text{phonology (stem inventory):} \\ \quad 0 \text{ /los′/; } 1 \text{ /los′at/} \end{array}\right.$

The same approach is taken for the other examples in Table 4. Thus for example the lexeme LES contains the stem alternant /lesov-/, which is selected for personal-noun derivation in -*šč'ik*, and the lexeme KRUŽOK contains the alternant /kružkov/. From the examples in Table 4 it should be clear that the stem-building formative and index used for one lexeme do not necessarily coincide with the stem formative and index used for another. Index 1 is assigned to a stem built in /ov/ for LES and for a stem built in /'at/ for the lexeme VOLK. Note also that though in this case personal-noun derivation is derived in two different suffixes, one should be careful about viewing different suffixes always being associated with different stem-building formatives, for example /'at/ with derivation in -*n'ik* and /ov/ with derivation -*šč'ik*. For a number of lexemes in -*n'ik*, the complex stem 1 on which they are based is a stem in /ov/, for example *sadovn'ik* 'gardener', *baxč'ovn'ik* 'watermelon watchman'. It is therefore an important theoretical move to keep indexation separate from stem formation, as has been argued in some detail by Greg Stump (1998a: chapter 6).

Note also that the complex stem alternants serve for derivation other than personal-noun derivation. For example, /los′at/ is selected not only for personal-noun derivation in -*n'ik* but also derivation in -*'in(a)*, a suffix used to derive nouns denoting the meat of an animal, among other things (see Cubberley 1994: 45). It is therefore important that stem alternants are indexed, expressing that they are semantically empty and map onto several functions at once. Of course for the majority of noun bases the simple stem is used. This is captured naturally within our default inheritance framework by simply stating that by default stem 1 inherits from stem 0, as we will see in section 5.

3.2.   *Verb lexemes and indexed stem*

In descriptions of the Russian verbal system it is traditional to identify two stems on which rules realizing inflectional categories are based.

Timberlake (1993: 850), for example, talks about the two-stem approach as being "convenient" for describing Russian verbs (though at the same time noting Jakobson's famous alternative one-stem account [1948]). He outlines this approach as follows: "Verbs commonly display two major stem alternants, the present allostem, used for the present tense, imperative and present participles, and the past/infinitive allostem, used for past, infinitive, past (active) participle and (past) passive participle."

Table 5 shows both stems of three verbs belonging to three different groups, which are inflected for nonpast first person singular and for infinitive. Note that Roman numerals denote the two conjugation classes; in the first conjugation there are a number of subgroups, and these are labeled in terms of the present stem.[12]

In *-tel'* suffixation we find that out of the stems contained in the verb lexeme, one type is (nearly) always selected over the others. The stem selected is the "infinitive" stem; all the bases in Table 6 are given in the infinitive form and we can clearly see that *-tel'* attaches to this same stem.

In a lexeme-based model, this is simply a matter of having an indexed-stem inventory for verbs. Indexed-stem inventories have been proposed for Russian verbs in a recent article by Sadler et al. (1997). Leading on from the two-stem tradition, they propose an indexed-stem account of Russian verbs based on Aronoff (1994). Though their paper chiefly concerns the inheritance of the verb's argument structure in deverbal nominalizations, their account relies on assigning to verbs four indexed

Table 5. *Verb allostems*

Class	Root	Infinitive stem	Present stem	Gloss
I -aj	del-	dela-(t')	delaj-(u)	'do'
I -ov	tolk-	tolkova-(t')	tolkuj-(u)	'demand'
II	reš-	reši-(t')	reš-(u)	'decide'

Table 6. *Nouns in -tel'*

Infinitive	Gloss	Personal noun	Gloss
dar'i(t')	'give (present)'	dar'itel'	'donor'
grab'i(t')	'steal'	grab'itel'	'thief'
xran'i(t')	'preserve'	xran'itel'	'custodian'
terza(t')	'torment'	terzatel'	'tormentor'
podžiga(t')	'set on fire'	podžigatel'	'arsonist'
zauša(t')	'abuse'	zaušatel'	'abuser'
tolkova(t')	'interpret'	tolkovatel'	'interpreter'

stems. In other words, they add two stems to the inventory assumed in traditional accounts. The four stems (when stem 0 is counted) are given in Table 7. If we compare Table 7 with Table 5 we see that "infinitive" and "present" stem correspond to stem 1 and stem 2. What is novel is the additional stems stem 0 and stem 3.

Stem 0 resembles the root in Table 4, but the authors are at pains to point out in an earlier version of the paper (1994) that there is a distinction (e.g. 1994: 22). To illustrate with the examples in the tables above, the root to *tolkova(t')* is *tolk-*; but in Table 7 we see that its stem 0 is the more complex *tolkov-*. This is because in derived verbs stem 0 is the root plus a verb-forming suffix. Subsequent derivation from the verb takes place on this stem, rather than the root; in other words, stem 0 is selected for derivation. For example, *tolkova(t')* 'interpret' has as well as *tolkovatel'* the alternative derivative *tolkovn'ik*, both meaning 'interpreter'. The derivation in *-n'ik* is based on stem 0 /tolkov/, and not the root /tolk/, which would yield **tolkn'ik*. It should be noted, however, that the root will by default coincide with stem 0 in underived verbs. This is the case for example with *dela(t')*, which derives the result noun *del(o)* 'deed' (through zero derivation).

Stem 3 is a radical proposal that takes its lead from Aronoff's account of Latin and his use of the "third" stem (see section 2). Sadler et al. propose lifting the category past participle passive, for example *sdelan(o)* 'done (neut sg)', out of the domain of the infinitive stem (their stem 1) and associating it with another stem, stem 3. This is a logical step if the /n/ is viewed as a stem-forming element (see above), rather than as part of the inflection realizing the category, or a concatenator. Hence *sdelano* is viewed as *sdelan-o* and not *sdela-no* (or *sdela-n-o*). This same stem is used for another category, namely productive deverbal nominalization: examples such as *delan'jio* 'doing', *tolkovan'ijo* 'interpreting' are analyzed as *delan-'ij(o)*, *tolkovan-'ij(o)*, where the /n/ is part of the stem and the /'ij/ is the nominalizing suffix. Though they do not make the observation, a priscianic approach, where one form is derived from the other, will not work. In Russian, imperfective verbs cannot form as past passive participle, yet they freely nominalize, for example **trebovan(o)* but

Table 7.   *Sadler et al.'s indexed verb stems*

Stem 0	Stem 1	Stem 2	Stem 3
del-	dela-	delaj-	delan-
tolkov-	tolkova-	tolkuj-	tolkovan-
reš-	reši-	reš-	rešen-

trebovan'ij(o) 'demanding'. An attempt to derive the participle from the nominalization will also fail. Russian has alternative ways of nominalizing, and verbs capable of forming the participle may have nominalizations in a rival suffix. For example *pobel'i(t')* 'whitewash' derives *pobelk(a)* 'whitewashing' with the rival nominalizing suffix *-k* (which is based on stem 0); nonetheless it is able to form the participle *pobe'en(o)*.

Adopting a stem-indexed approach based on Sadler et al., we can represent a verb such as *tolkova(t')* 'interpret' as in (7). At the phonological level we see a stem inventory of four indexed stems. Derivation in *-n'ik* takes place on stem 0, *tolkovn'ik*; derivation in *-tel'* is based on stem 1, *tolkovatel'*; and nominalization on stem 3, *tolkovan'ij(o)*. Stem 2 is reserved for the inflectional categories nonpast, imperative, and present participle.

(7) TOLKOVAT' $\left\{\begin{array}{l}\text{syntax:}\\\quad\text{verb; imperfective; transitive}\\\text{semantics:}\\\quad\text{'interpret'}\\\text{phonology (stem inventory)}\\\quad 0\ /\text{tolkov}/;\ 1\ /\text{tolkova}/;\ 2\ /\text{tolkuj}/;\ 3\ /\text{tolkovan}/\end{array}\right.$

3.3. Truncation and indexed stems

The relationship between adjectives in *-sk* and corresponding person nouns in *-ec* is highly productive. It is widely found for example in terms for countries and peoples, such as *makedonsk(ij)* 'Macedonian' and *makedonec* 'person from Macedonia'. It is commonly assumed that the relationship is that of base adjective in *-sk* deriving the personal noun in *-ec*, where the relational adjective suffix *-sk* is truncated in the process.[13]

In an interesting article by Isačenko it is claimed that truncation (*usečenie*) is a productive feature of some Russian derivation. However, when it occurs it is restricted to the deletion of material that can be specifically identified as a suffix (Isačenko 1972: 98).[14] Isačenko distinguishes truncation due to the avoidance of juxtaposing like with like ("ustranenie udvoennyx suffiksov," 1972: 97) from truncation as a regular feature of the derivation. An example of the former would be the qualitative adjective suffix *-ovat(ij)*, as in *krasnovat(ij)* 'reddish', based on a relational adjective in *-ov*. For example *rozov(ij)* 'pink' derives *rozovat(ij)* 'pinkish', and not **rozovovat(ij)*. An example of the latter would be derivation of person nouns from nominalizations in *-k*, such as *zabas-*

tovk(a) 'strike' > *zabastovšč'ik* 'striker' where the suffix *-k* is truncated.[15] Isačenko's term for this is "morphological truncation" ("morfologičeskoe usečenie," 1972: 103).

Truncation of *-ec* is, of course, only necessary if it is assumed that X-*ec* is derived from X-*sk*, and the literature does not agree on this.[16] One could try to account for this productive relationship without truncating *-ec*. For example in the Macedonian example, one could view *Makedon'ij(a)* as a base that simultaneously derives the *-sk* and the *-ec* items (where the element /'ij/ is then truncated for both derivations). Another approach is to posit an abstract item {MAKEDON}, which "coderives" *makedonec* and *makedonsk(ij)*. This is proposed for example in Worth (1967: 2280). The approach we favor is X-*sk* > X-*ec*, with truncation of *-ec*, based on the convincing directionality arguments outlined in Darden (1988). His main arguments are as follows. First, *-ec* is anyway productively added to adjective bases, for example *podl(ij)* 'ignoble' > *podlec* 'scoundrel', *mudr(ij)* 'wise' > *mudrec* 'wise man', *znakom(ij)* 'familiar' > *znakomec* 'acquaintance'. Second, stress behavior tends to favor derivation from a *-sk* base. For example, in the set *árm'ij(a)* 'army', *arméjsk(ij)* 'army (adj)', *arméec* 'soldier', clearly stress in the *-ec* form follows the adjective and not the noun. And third, velar palatalization appears to indicate derivation from the adjective (1988: 93–94). Derivation in *-sk* based on stems in a velar may or may not cause mutation. Where they do not, the derivative in *-ec* does not usually either, for example *Arkadak, arkadaksk(ij), arkadakec*. However, on occasion it does, as in *Viborg, viborgsk(ij), viborgec*, or *viboržec*. The choice is only open to the *-ec* form, though, suggesting that it is derived from the adjective in *-sk*, and not the other way around.

Having argued for truncation, the question is how truncation can be fitted into a stem-indexed approach. We propose that, as part of their stem inventory, lexemes may contain a special "truncated" stem, which we label stem −1. This stem is defined as stem 0 (the default stem) minus the suffix that derived it. In this way we limit truncation to applying to morphologically introduced structure only, following Isačenko, as well as giving it a declarative interpretation. We will return to truncation in section 5.4.

4. Network morphology

Network morphology[17] is a framework representing the lexicon in terms of a network of nodes, capturing the generalizations that can be made about morphology, as well as providing for the idiosyncrasies characteris-

tic of morphology. The central concept that lies behind network morphology is that complexities characteristic of morphological systems can be more parsimoniously described by distributing information declaratively across a network, and in such a way that generalizations may be expressed.[18] Network morphology should therefore be seen as a member of the set of declarative frameworks that make use of inheritance hierarchies. Morphological facts are distributed across nodes that can be arranged hierarchically, and the flow of information between them is given the interpretation of inheritance. The more general a fact, the higher will be its place on the hierarchy to ensure widest possible application. Conversely, rarer facts will appear lower down, and their inheritance will be more limited.[19] The inheritance from nodes in network morphology has a nonmonotonic, default interpretation. In a default-inheritance hierarchy, what are stated at higher-level nodes are not absolutes, but generalizations. These generalizations may be overridden locally. A default-inheritance hierarchy naturally captures, for example, the fact that Russian class I nouns generally form the genitive plural in *-ov*, but in the case of *soldat*, the bare stem is used. Thus default inheritance goes hand in hand with network morphology's aim of expressing morphological generalizations.

In network morphology we may say that we have a highly structured lexicon, in the sense of a component that accounts for lexemes. This component comprises a lexemic hierarchy where generalizations relating to word classes are stored, an inflectional hierarchy for inflectional morphology, and a derivational hierarchy for derivation. Lexical idiosyncrasies are encoded in the lexical entries, and the lexical entries appear as leaf nodes of the lexemic hierarchy. From this follow two things: first, lexical entries are partially specified items and are "filled out" by inheriting information from the network; and second, lexicalized exceptionality is naturally encoded by local information overriding information in the hierarchies, possible due to inheritance from nodes being nonmonotonic, or default. Lexical entries are lexemes; in other words, network morphology assumes that the lexeme is the minimal sign.

Figures 1 and 2 show the interaction of the lexemic hierarchy with the inflectional and derivational hierarchies respectively.[20] In the lexemic hierarchy, nodes representing the various word classes are arranged under the top node LEXEME. The node NOMINAL generalizes over ADJECTIVE and NOUN, which in turn act as mothers to adjective and noun lexical items. The lexemic hierarchy makes generalizations about lexemes: their phonology, syntax, and semantics. For example, adjectives and nouns normally have stems ending in a hard constant, hence this is a phonological generalization that can be stated at NOMINAL. General semantic
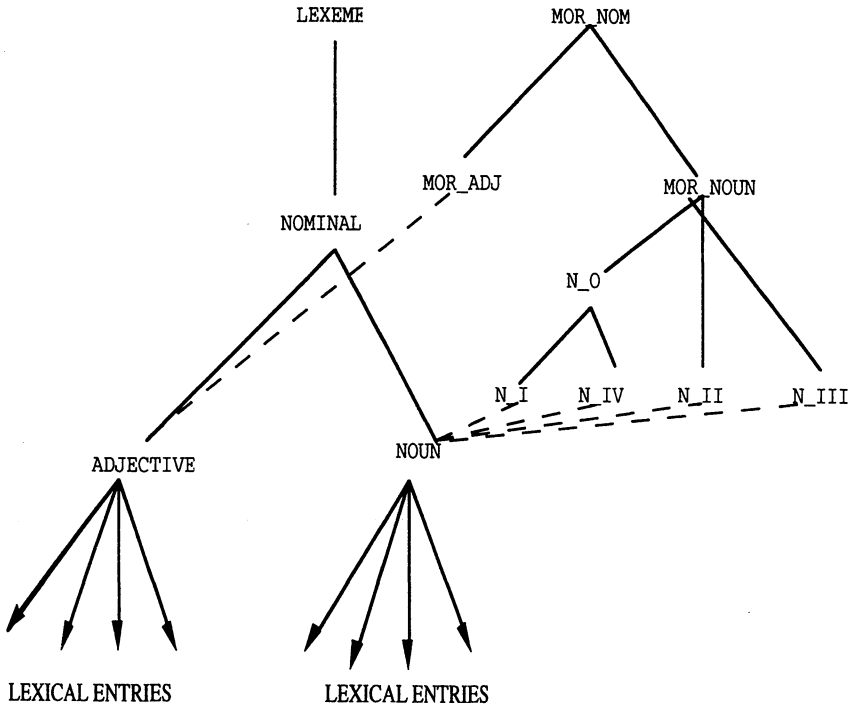
Figure 1.   *The lexemic hierarchy and the inflectional hierarchy*

information, such as the fact that nouns are normally undifferentiated for sex, is stated at NOUN.

   The inflectional hierarchy is orthogonal to the lexemic hierarchy, where the lexical items are the terminal nodes. This encodes the way in which morphology is orthogonal to the word, thereby constituting a distinct, but connected, level of linguistic description. Affix homonymy is captured as the sharing of morphological facts expressed by mother and daughter node relationships, where the information at a mother node is made available to a daughter node. The node MOR_NOUN acts as a mother to the declension class nodes. Information shared between all classes will be made available from this node; for example the dative, instrumental, and locative plurals.[21] Where sharing is restricted to only certain declension classes, a node is set up as mother over only those classes. For example, sharing the oblique singular inflections is restricted to N_I and N_IV. A node N_O is set up that will store these facts, from which N_I and N_IV will draw. With specifically default inheritance, facts can be shared across some but not all nodes as defaults, which are overridden

by the nodes not participating in the fact sharing. For the nominative plural -*i* ending, MOR_NOUN generalizes over N_I, N_II, and N_III, but N_IV is made to override with a more specific fact that the nominative plural is in -*a*.

Figure 2 shows the derivational hierarchy, which, like the inflectional hierarchy, is orthogonal to the lexemic hierarchy. Recall that the leaf nodes of the lexemic hierarchy are the lexical entries themselves, representing lexemes. Since the derivational hierarchy describes additional information about a lexeme's derivational patterns, the shape of the derivational hierarchy is to some degree determined by the lexemic hierarchy and is a projected image of it.

The formalism used to represent network-morphology theories is the lexical knowledge representation language DATR, which was developed by Evans and Gazdar at Sussex in the late 1980s. DATR defines networks
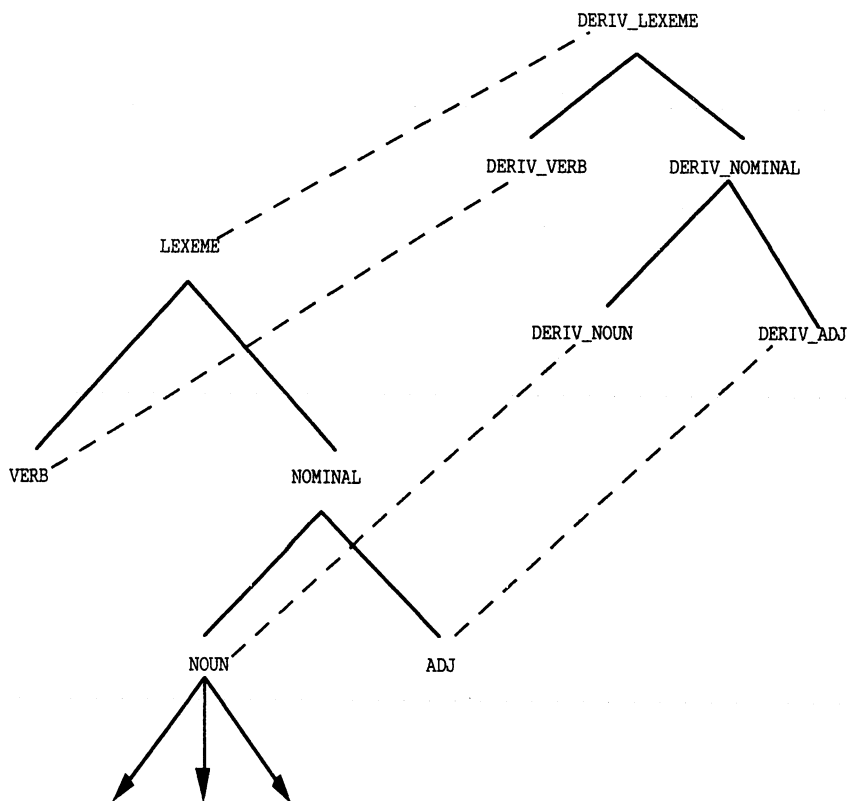

Figure 2. *The derivational hierarchy as a projected image of the lexemic hierarchy*

by links typed by attribute paths through which information is inherited (Gazdar 1990: 1). Its interpretability comes from an explicit theory of inference (see Evans and Gazdar 1989) and an explicit declarative semantics (see Keller 1995). It is implementable on computer (a number of computer interpreters exist for DATR descriptions; see for example Evans 1990), which allows for a given theory's predictions to be verified by computer. Lastly, as a main feature DATR has a nonmonotonic notion of inference by default (Gazdar i.p.: 15). All these features make DATR highly appropriate for default-inheritance frameworks, such as network morphology, as noted in Evans and Gazdar (1996: 207).

5. Network morphology and indexed stems

Having introduced network morphology, we are in a position to show how indexed stems are naturally accommodated within a declarative framework such as this. We first show how the derivation of stem in terms of affixation is expressed as inheritance of a formative from the derivational hierarchy, and inheritance of the base lexical entry's stem. We then discuss stem indexation in the framework and the relationship of the stems within the stem inventory (stem "formation"). Finally we look at how truncation can be treated in terms of stem formation.

5.1. *Derivation as inheritance of the stem*

The derivational hierarchy stores information regarding the stem of a derived lexeme. Thus for derived lexical entries the stem is inherited information. For inheritance from the derivational hierarchy one must begin with the lexemic hierarchy from where underived and derived lexemes inherit. The stem is there analyzed as consisting of two components, the stem of its base lexeme, and the affix that is used in the derivation. Corresponding to these two elements, inheritance of the stem is from two sources: first the derivational hierarchy, via the node LEXEME in the lexemic hierarchy for the suffix component, and second, the lexical entry's base for the root component. This is represented in Figure 3 for the derivational chain *atom* > *atomn'ik* 'atomic scientist'. Inheritance from the derivational hierarchy is from the top node LEXEME, which, it will be recalled from Figure 2, is the top node in the lexemic hierarchy. Nodes representing derived lexical entries multiply inherit from LEXEME, and thence DERIV_LEXEME, and the lexical entry representing the base.
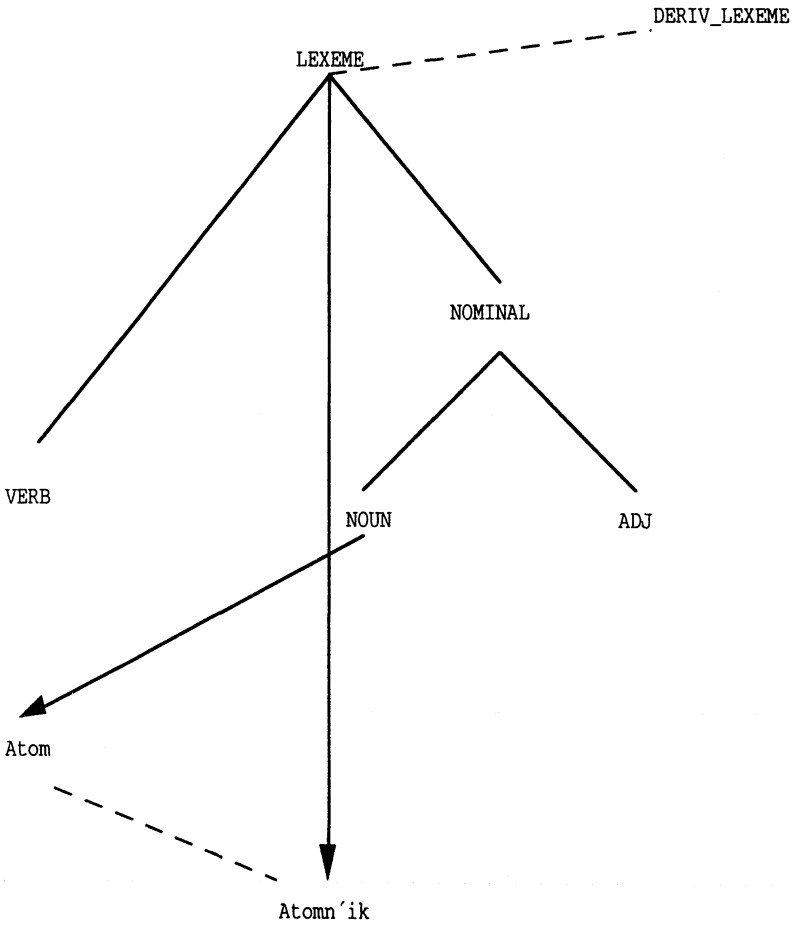
Figure 3. *The derivative* Atomn'ik *inheriting from LEXEME and* Atom *for stem information*

Underived noun, adjective, and verb lexical entries inherit from the appropriate node in the lexemic hierarchy, NOUN, ADJ, VERB.

The DATR representation of Figure 3 is given in (8) to (10). In (8) we see a partial representation of LEXEME that expresses that the stem is defined as a complex value: the stem of the base plus the deriving suffix. In the case of a complex noun such as *atomn'ik* this corresponds to /atom/ plus /n'ik/. It should therefore be carefully noted that Atomn'ik is the lexeme's label and does not denote the fact that for this item the suffix *-n'ik* is assumed to be lexically specified.

(8) LEXEME:
     <stem>=="<base stem>" "<deriv suffix>"
     . . .

The inheritance of Atomn´ik from LEXEME is represented in (9), where the empty angled brackets represent in DATR the main inheritance source, the inheritance of all paths at the mother node by default.[22] The secondary source of inheritance is from its base Atom, (10). The equation <base>=="Atom:<>" expresses that all paths at Atom, including the path <stem>, are inherited by Atomn´ik and relabelled with the prefix attribute base. Hence the value for the path <stem> at Atom, (10), is inherited as the value of the labelled path <base stem> for Atomn´ik. At Atomn´ik the value for <base stem> will therefore be atom. In this way a partial value is provided for the path <stem> at LEXEME, (8).

 (9) Atomn´ik
      <>==LEXEME
      <base>=="Atom:<>"
      . . .

(10) Atom:
      <>==NOUN
      <stem>==atom
      . . .

The remainder of the value, the inheritance of <deriv suffix>, is more complex. As we said, lexeme formation specifies a change in a base's syntactic category, and concomitant change in its semantic makeup, which corresponds to a change in stem shape. The derivational hierarchy matches the suffix used in the derivation with derived syntactic and semantic information. This is shown in (11) and (12) for derivation in the suffix *-n´ik*.

(11) NODE_1:
      <deriv suffix>==NODE_2:<"<deriv syn cat>"
      "<deriv sem feature>">
      . . .

(12) NODE_2:
      <noun person>==n´ik
      . . .

The paths in (11) express the three types of information introduced in a derivation, the suffix, the syntactic category, and the semantic feature. Note that the attribute deriv identifies this information as derived information. Derivative lexical entries are specified for the values for the

paths `<deriv syn cat>` and `<deriv sem>`, that is, the syntax category and semantics introduced in derivations are listed in the lexical entry. These values are then retrieved from the query lexical entry to evaluate what suffix the query lexical entry will inherit. The change brought about by the derivation is thus encoded by matching the derived syntactic and semantic features that are recorded at a derivative lexical entry with a particular suffix. (11) shows that the derivational hierarchy represents the match as an evaluation, where (12) expresses that a query lexical entry that has been specified as derived with the syntactic class noun and the semantic feature person will inherit the suffix *-n'ik*, that is, *-n'ik* is a possible personal-noun-forming suffix.

5.2. *Representing stem indexing*

We demonstrated above that verbs can be represented as lexemes with an inventory of stems. Each stem is indexed, and an index corresponds to a range of morphological functions, following Aronoff. For verbs, we argued for an inventory of four stems (section 4.2). Deverbal person derivation is based on stem 1. To capture this, one approach would be to list all the stems in the base lexical entry and have the derivative lexical entry select the appropriate stem. The actual suffix is inherited from the hierarchy, as we showed above. This is represented in (13) and (14).

```
(13)  Tolkovat´:
      a.   <>==VERB
      b.   <root>==tolk
      c.   <stem 0>==tolkov
      d.   <stem 1>==tolkova
      e.   <stem 2>==tolkuj
      f.   <stem 3>==tolkovan
           . . .
(14)  Tolkovatel´:
      a.   <>==LEXEME
      b.   <base>=="Tolkovat´:<>"
      c.   <base stem>=="Tolkovat´:<stem 1>"
      d.   <deriv sem feature>==person
           . . .
```

In (13) and (14) the selection of stem 1 by the derivative is expressed by the fact (14c) at `Tolkovatel´`. Stem 1 is lexically listed in the base `Tolkovat´`, (14d).

We argued that the complex suffix -*'at-n'ik* could be interpreted as derivation in simple -*n'ik* based on a special stem in /'at/ belonging to the stem inventory of the base. Stems in /'at/ are posited in any case for plural inflection of nouns denoting young of animals (e.g. *kot'onok*, *kot'at[a]*), as we saw in section 2. The derivative *los'atn'ik* 'elk hunter' and its base *los'* 'elk' are represented in (15) and (16). (15) expresses the base lexeme LOS' with its inventory of stems /los/ and /los'at/. (16) represents the derivative LOS'ATN'IK where reference is made to the stem labeled stem 1 at LOS' to express derivation based on the stem /los'at/.

(15) Los´:
      <>==NOUN
      <stem 0>==los´
      <stem 1>==los´at
      . . .

(16) Los´atn´ik:
      <>==LEXEME
      <base>=="Los´:<>"
      <base stem>=="Los´:<stem 1>"
      <deriv sem feature>==person
      . . .

5.2.1. *Lexicalized and nonlexicalized stem selection.*   One crucial difference between derivation in -*n'ik* and derivation in -*tel'* with regard to stem selection is that stem selection is regular with the suffix -*tel'* since almost without exception -*tel'* attaches to stem 1, whereas there is only a handful of -*n'ik* words based on the /'at/ stem. To encode stem selection as productive in this way we specify it in the derivational hierarchy, as represented in (17) and (18).

(17) LEXEME:
      a. <stem>=="<base stem "<index>">"
          "<deriv suffix>"
      b. <index>==STEM_SELECTION:<"<base syn cat>"
                                    "<deriv sem feature>">
      c. <base stem>=="<root>"
          . . .

(18) STEM_SELECTION:
      a. <verb person>==1
      b. <>==0
          . . .

Comparing (17a) with the earlier (8) we note the additional evaluable

path "`<index>`". This expresses that a derived stem[23] is defined not as a suffix added to simply the base's stem, but more specifically it is defined as a suffix added to a particular indexed stem of the base. This can be stem 0, 1, 2, 3, etc. In other words, it assumes that base lexical entries have a range of indexed stems. Which indexed stem is selected is expressed by (17b) and (18). In (17b) we see that the index of the stem depends on the type of derivation, that is, the syntactic category of the base and the semantics of the derivation. In (18a) we can see that for example deverbal personal-noun formation (which is productively realized by the suffix -*tel'*) will be based on stems with index 1. Note the default, (18b), that any other type of derivation, represented by the empty angle brackets, will be based on stem 0. Recall from section 2 that for example the -*k* nominalizations such as *pobelk(a)* 'whitewashing' are based on stem 0 (i.e. /pobel/).

An important point to note is that stem selection is expressed independently of -*tel'* suffixation. Stem 1 is associated with the function "deverbal person formation," and not directly with the suffix -*tel'*. Since deverbal person formation is productively realized by -*tel'*, in one sense the stem selection could be viewed as part of a productive rule of -*tel'* affixation. Instead we have been less direct: a morphological category is first indexed with one of the various stems of the lexeme, just as Aronoff does with Latin verbs where a number of categories are associated with the verb's third stem (see section 2.1). If we wished to express correspondence of stem with inflectional categories, we would simply add a referral at `STEM_SELECTION` such as `<mor inf>==<verb person>`. In this way we capture disparate functions being united by a single index.

5.3. *Representing stem formation*

The base lexical entries in (13) and (15) specify their inventory of stems in a somewhat redundant fashion. To capture the similarity of the stems in the inventory we can make generalizations about "stem formation" in the hierarchy. Derivation based on stem 0 is the default, and we saw how this is represented in (18b), which we can think of as the "stem selection default." An example was given in section 3.2 where the derivation of the result noun *del(o)* from the verb *dela(t')* is based on stem 0. In terms of stem formation, on the other hand, the first generalization we need to capture is that by default stem 0 coincides with the root. This is represented in (19), the "stem formation" default. For example, stem 0 of `Atom` will be its root. (19) expresses that by default a noun's "basic" stem (stem 0) coincides with its root.

(19) NOUN:
      `<stem 0>=="<root>"`

      . . .

      VERB:
      `<stem 0>=="<root>"`

      . . .

Given this, (15) can be modified as (20) where what is lexically specified is the root, and not stem 0. To account for the complex stem in /'at/, stem formation is viewed as a referral. It is specified as a referral to the basic stem `<stem 0>` plus the formative ´at.

(20) Los´:
      `<>==NOUN`
      `<root>==los´`
      `<stem 1>==<stem 0> ´at`

      . . .

We noted in section 3.2 that for some verbs stem 0 must be distinguished from the root. In other words it is important to see stem 0 coinciding with the root not as an absolute, but as a default generalization that may be overridden. This is the case for *tolkova(t´)*, which derives the alternative personal noun *tolkovn´ik* based on /tolkov/, where the stem /tolkov/ is distinct from the root stem /tolk/. This override must therefore be specified as such in the lexical entry, that is, there must be a lexically specified value for `<root>` as well as for `<stem 0>`. But there is nonetheless a formal relationship between stems in a lexeme's stem inventory. This is captured simply by expressing stems as referrals to one another. For `Tolkovat´` we modify (13) as (21). This expresses that stem 0 /tolkov/ and stem 2 /tolkuj/ are both based on the root stem /tolk/; in turn stem 1 /tolkova/ is based on stem 0 /tolkov/, and stem 3 /tolkovan/ is based on stem 1 /tolkova/.

(21) Tolkovat´:
      a. `<>==VERB`
      b. `<root>==tolk`
      c. `<stem 0>==<root> ov`
      d. `<stem 1>==<stem 0> a`
      e. `<stem 2>==<root> uj`
      f. `<stem 3>==<stem 1> n`

          . . .

5.3.1. *Lexicalized and nonlexicalized stem formation.* The node represented in (21) expresses lexically specified stem formation. This, however,

misses an important generalization since the pattern in (21) is repeated for a whole class of verbs, what we labeled the I-*ov* class in Table 5. To capture this generalization we simply push the facts in (21) up the lexemic hierarchy from the leaf lexical entry node `Tolkovatel´` to a node under VERB, the node VERB_I_OV, in other words, a node representing a class of verbs from which `Tolkovatel´` is specified to inherit. This is represented in (22) and (23). Ultimately we would establish nodes for all the classes in this fashion. In fact a network morphology account of verbal inflection along these lines has already been put forward in Brown (i.p.).

(22) VERB:
    a. `<>==LEXEME`
    b. `<syn cat>==verb`
    c. `<stem 0>=="<root>"`
        . . .

(23) VERb_I_OV:
    `<>==VERB`
    `<stem 0>=="<root>" ov`
    `<stem 1>=="<stem 0>" a`
    `<stem 2>=="<root>" uj`
    `<stem 3>=="<stem 1>" n`
    . . .

It should be noted that the overriding of the generalization that stem 0 coincides with the root is itself a generalization for the I-*ov* class. Hence this can be stated at the node VERB_I_OV, rather than being lexically specified. Thus stem 0 coinciding with the root stem is the default of verbs, (22c), but this is overridden in the I-*ov* class where the formative /ov/ added to the root encodes stem 0.

5.4. *Representing truncation as a type of stem "formation"*

In section 3.3 we argued for an analysis of *kanadec* that relied on a truncation rule. The personal noun is derived from the relational adjective *kanadsk(ij)*, where the suffix *-sk* is truncated. Our novel approach is to view truncation in terms of derivation taking place on one of the stems in the base lexeme's stem inventory, a special "truncated" stem. This truncated stem we label stem −1. Thus for the lexeme KANADSKIJ we posit (at least) two stems, stem 0 /kanadsk/ and stem −1 /kanad/.

The indexed stem solution to truncation is represented within our formalism as in (24) to (26), which show the derivational chain *kanad(a) > kanadsk(ij) > kanadec*. Note that for ease of exposition we

have also listed a number of extensional sentences, that is, expressions of facts belonging to lexical entries that are inferred from the theory (the facts that appear in a theorem list). These are represented by a single equals sign in DATR, and for expository purposes we have placed them in square brackets.

(24) Kanada:
    a. `<>==NOUN`
    b. `<stem 0>==kanad`

     . . .

(25) Kanadskij:
    a. `<>==LEXEME`
    b. `<base>=="Kanada:<>"`
    c. `<deriv sem feature>==rel`
    d. `[<base stem>=kanad]`
    e. `[<stem 0>=kanadsk]`
    f. `[<deriv suffix>=sk]`
    g. `<stem -1>==kanad`

     . . .

(26) Kanadec:
    a. `<>==LEXEME`
    b. `<base>=="Kanadskij:<>"`
    c. `<base stem>=="Kanadskij:<stem -1>"`
    d. `<deriv sem feature>==person`
    e. `[<stem 0>=kanadec]`

     . . .

In (26c) we see lexically specified stem selection, as we did for Los´atn´ik in (16). The stem selected from the base Kanadskij is what we have termed stem −1, the truncated stem, and this is lexically specified in (25g). However, if we look at the extensional sentences at the node Kanadskij we see that the value of `<stem -1>` coincides with that of `<base stem>`, (25d). This is because stem −1 is simply a derived stem minus the suffix it is derived in. This can be rather neatly captured in our approach by the referral `<stem -1>==<base stem>`.

We view this as a type of stem "formation" similar to what we did for *tolkovatel'*. Its productivity is captured by pushing it further up the hierarchy for general application. In fact, we can push this generalization right to the topmost node LEXEME, to have widest possible application. Note that Isačenko talks about morphological truncation as a phenomenon occurring in all major word classes (Isačenko 1972). We therefore state stem −1 formation at LEXEME as in (27). Recall that derived lexemes inherit directly from LEXEME. As we have seen, stem is defined

as the stem of the base plus the suffix, (27a). Stem -1 is then viewed as simply the stem deprived of the suffix, (27b).

(27) LEXEME:
     a.   `<stem>=="<base stem "<index>">"`
         `"<deriv suffix>"`
     b.   `<stem -1>=="<base stem 0>"`
         `. . .`

6. Conclusion

We have demonstrated that greater generalization can be located in the derivational system of Russian by taking a lexeme-based approach where lexemes are equipped with an inventory of indexed stems. In this way we account for the common phenomenon in Russian derivation of concatenators or interfixes. We also account for what we could call the converse phenomenon, truncation. This approach has been shown to fit naturally within declarative frameworks such as network morphology, where the theory has been computationally verified for the predictions it makes.

We showed that lexemes may contain more than one stem on which morphological rules operate. Following Sadler et al. (1997) we identify four stem types for Russian verbs. For the verb lexeme TOLKOVAT' 'interpret' the indexed stem inventory was given as in Table 8.

A WFR's phonological condition may be viewed as the selection of one stem over other stems in the inventory. The -*tel'* WFR was specified with the phonological condition on stem 1 of the verb base's set of stems. To incorporate phonological conditions as stem selection we revised the definition of stem as a particular indexed stem of the base plus suffix. The indexed stem chosen then depends on the value for `<index>`. This is evaluated on the basis of the type of derivation: with deverbal person derivation the index is 1, and with any other type the index is 0. Selection may also be lexically specified, as in the case of *los'* 'elk' > *los'atn'ik* 'elk hunter'. The similarity of stems in an inventory was captured in terms of referrals, where the basic referral is to the root. This was either lexically specified, or expressed as generalizations in the hierarchy, though the

Table 8.   *Indexed stem inventory of TOLKOVAT'*

Stem 0	Stem 1	Stem 2	Stem 3
tolkov-	tolkova-	tolkuj-	tolkovan-

basic stem, stem 0, was stated at the syntactic nodes in the lexemic hierarchy as coinciding with the root by default.

Finally, truncation was accounted for in terms of indexed stems. In this case the morphological structure of the base lexeme is accessed. The base is seen as containing a "truncated stem" that is analyzed as the stem of the base minus the suffix. This was represented as in (28) where the truncated stem <stem -1> is defined as the base's stem, without the base's suffix.

(28)   LEXEME:
```
<stem>=="<base stem "<index>">"
"<deriv suffix>"
<stem -1>=="<base stem 0>"
. . .
```

Notes

2.   For lexemes and lexeme-based morphology, see discussions in Matthews (1991 [1974]: chapter 2), Zwicky (1992: 333–338), Aronoff (1994: chapter 1), Beard (1995: 45–50), and Stump (1998b: 13–14).
3.   Russian orthography closely follows phonemic representation, and the phonemic transcription we use is therefore close to standard transliteration, with a few minor points of difference (based on Corbett and Fraser 1993: note 2, which is itself based on Timberlake 1993: 828–832). The paired palatalized (soft) and unpalatalized (hard) consonants are distinguished by an acute ('), which marks the soft member of the pair. For example, in the minimal pair *l'uk* 'hatchway', and *luk* 'onion' the first form has the soft /l'/. Note that consonants are always soft before the phoneme /e/, hence there is no need to mark them with an acute in this context. For example, the locative singular of *zakon* 'law' is represented as *zakone* since the stem-final /n/ is automatically soft. The velars /g/, /k/, and /x/ are hard except when preceeding the /i/ and /e/ phonemes; in these contexts they are automatically softened. We therefore do not use an acute on the velars in these contexts since they are automatically softened. Compare the nominative singular form *ručka* 'handle' with the genitive singular *ručki*, where the /k/ is

soft before the -*i* ending but not indicated as such. Note that unpaired soft /č′/ and /šč′/ are redundantly marked with an acute when preceding a vowel, but unpaired soft /j/ is never marked with an acute. We recognize five vowel phonemes (under stress), which are /a/, /e/, /i/, /o/, /u/. The phoneme /i/, standardly transliterated as '*i*', has an allophone [ɨ], standardly transliterated as '*y*'. The allophone [ɨ] is automatically used when following a hard consonant. The correct version of /i/ will therefore be implied by the nature of the preceding consonant.

4. That is, the word class appears to remain unaltered due to the fact that the syntactic class of the input to the WFR coincides with the syntactic class introduced by the WFR. It is important to distinguish this from "true" category-preserving derivation, which refers to the preservation of the features of all inherent morphosyntactic categories of the base, particularly gender (Stump 1993). An example of true category-preserving derivation in Russian is the rich system of expressive derivation. In network morphology the preservation of all features in expressive morphology is simply viewed as inheritance by default, as shown in Hippisley (1996).

5. Relational adjectives "... designate a relationship which characterises the noun modified as being of, from, or connected with something ..." (Townsend 1975: 209). Compare this to qualitative adjectives where there is a definite semantic content: either 'having X' or 'similar to X' as a quality, where X is the semantics of the noun base (Beard 1976: 108–9). From Beard an example of each type is *kamen'ist(ij)* 'rocky' from *kamen'* 'rock', i.e. 'having rock(s)', and *repčat(ij)* 'turnip-like' from *rep(a)* 'turnip'.

6. However, unlike the citation form, a lexical representation may end up never actually occurring as a surface form of the lexeme (Aronoff 1992: 16).

7. Note that we assume that the formative carries with it a "softening" feature, represented by the acute, which acts to soften any preceding consonant, in this case the /t/ of the stem *kot-*.

8. Note that vowel length is phonemic in Latin, and long vowels are indicted as such with a macron.

9. Though the first column includes past and present tenses, it is nonetheless associated with one function if we consider that all the forms that attach to it are imperfect (active and passive): present, future, imperfect of indicative mood; present, imperfect of subjunctive mood; present of imperative, infinitive, and participle (Aronoff 1992: 29–30). Note that Latin lacks a future subjunctive.

10. Cited in Panov (1968: 171, note 1).

11. The inflectional and derivational stem alternants posited in Stankiewicz (1962: 11) were shown later in Worth (1967) to be alternants that could be adequately accounted for by rules of allomorphy.

12. There are a number of ways of classifying Russian verbs. For a network morphology account of Russian verb classification, see Brown (i.p.).

13. For example Švedova (1980: section 287).

14. "Usečeniju podležat liš′ morfologičeskie edinicy a ne prosto cepo čki fonem," 'Only morphological units are subject to truncation, and not just any phoneme string'.

15. For discussion of the directionality suggested here, and other examples like this, see Isačenko (1969: 51).

16. See for example Vinogradov et al. (1953: 214), Vinogradov (1971: 99–100).

17. For the network morphology theory, see Corbett and Fraser (1993), Brown and Hippisley (1994), Fraser and Corbett (1995), Brown et al. (1996), Fraser and Corbett (1997), and Brown (1998). For an overview of the network morphology literature, see chapters 3 and 4 in Hippisley (1997).

18. On this see for example Corbett and Fraser (1993: 116); Fraser and Corbett (1997: 26).

19. The notion of inheritance hierarchy has been imported into linguistics from representations of knowledge common in the artificial intelligence literature (see Gazdar 1987: 45–48). Daelemans et al. offer a brief survey of the important inheritance-based frameworks used for accounts of syntax, morphology, phonology, and semantics (1992: 210–213).
20. Figure 1 is based on Figure 6 in Brown et al. (1996: 72); Figure 2 is from Hippisley (1997: 149).
21. These facts may be stated at a higher node MOR_NOM, which acts as a mother to MOR_ADJ as well as MOR_NOUN. The details are not important at this point of the discussion.
22. This is because the empty path is taken to be a leading subpath of every path at the mother node (Evans and Gazdar 1996: 172).
23. Recall that derived lexical entries inherit directly from LEXEME, hence <stem> here refers to a derived stem. Stem information for underived items will be specified at the syntactic class nodes in the lexemic hierarchy, from which underived lexical entries directly inherit.

References

Aronoff, Mark (1992). Stems in Latin verbal morphology. In *Morphology Now*, Mark Aronoff (ed.), 5–32. New York: State University of New York Press.
—(1994). *Morphology by Itself: Stems and Inflectional Classes.* Cambridge, MA: MIT Press.
Beard, Robert (1976). A semantically based model of a generative lexical word-formation rule for Russian adjectives. *Language* 52, 108–120.
—(1995). *Lexeme–Morpheme Base Morphology: A General Theory of Inflection and Word Formation.* Albany: State University of New York Press.
Brown, Dunstan (1998). From the general to the exceptional: a network morphology account of Russian nominal inflection. Unpublished doctoral dissertation, University of Surrey.
—(i.p.). Stem indexing and morphonological selection in the Russian verb. In *Models of Inflection*, Ray Fabri, Albert Ortmann, and Teresa Parodi (eds.), 196–221. Tübingen: Niemeyer.
—; Corbett, Greville; Fraser, Norman; Hippisley, Andrew; and Timberlake, Alan (1996). Russian noun stress and network morphology. *Linguistics* 34, 53–107.
—; and Hippisley, Andrew (1994). Conflict in Russian genitive plural assignment: a solution represented in DATR. *Journal of Slavic Linguistics* 2, 48–76.
Corbett, Greville; and Fraser, Norman (1993). Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29, 113–142.
Cubberley, Paul (1994). *Handbook of Russian Affixes.* Columbus, OH: Slavica.
Daelemans, Walter; De Smedt, Koenraad; and Gazdar, Gerald (1992). Inheritance in natural language processing. *Computational Linguistics* 18, 205–218.
Darden, Bill (1988). Truncation and/or transderivational constraints in Russian word-formation. In *Papers from the 24th Annual Regional Meeting of the Chicago Linguistic Society*, Lynn MacLeod, Gary Larson, and Diane Brentari (eds.), 89–99. Chicago: Chicago Linguistic Society.
Evans, Roger (1990). An introduction to the Sussex Prolog DATR system. In *The DATR Papers*, vol. 1, Roger Evans and Gerald Gazdar (eds.), 63–71. Cognitive Science Research Paper CSRP 139. Brighton: University of Sussex.

—; and Gazdar, Gerald (1989). Inference in DATR. In *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*, 66–71. Manchester: Association for Computational Linguistics.

—; and Gazdar, Gerald (1996). DATR: a language for lexical knowledge representation. *Computational Linguistics* 22(2), 167–216.

Fraser, Norman; and Corbett, Greville (1995). Gender, animacy, and declensional class assignment: a unified account for Russian. In *Yearbook of Morphology 1994*, Geert Booij and Jaap van Maarle (eds.), 123–150. Dordrecht: Kluwer.

—; and Corbett, Greville (1997). Gender assignment in Arapesh: a network morphology analysis. *Lingua* 103, 25–57.

Gazdar, Gerald (1987). Linguistic applications of default inheritance mechanisms. In *Linguistic Theory and Computer Applications*, Peter Whitelock, Mary Wood, Harold Somers, Rod Johnson, and Paul Bennett (eds.), 37–67. London: Academic Press.

—(1990). An introduction to DATR. In *The DATR Papers*, vol. 1, Roger Evans and Gerald Gazdar (eds.), 1–14. Cognitive Science Research Paper CSRP 139. Brighton: University of Sussex.

—(i.p.). Ceteris paribus. In *Aspects of Computational Linguistics*, Hans Kamp and Christian Rohrer (eds.), 2–35. Berlin: Springer.

Hippisley, Andrew (1996). Russian expressive derivation: a network morphology account. *The Slavonic and East European Review* 74(2), 201–222.

—(1997). Declarative derivation: a network morphology account of Russian word formation with reference to nouns denoting 'person'. Unpublished doctoral dissertation, University of Surrey.

Isačenko, A. V. (1969). Morpheme classes, deep structure and Russian indeclinables. *International Journal of Slavic Linguistics and Poetics* 12, 48–72.

—(1972). Rol' usečenija v russkom slovoobrazovanii. *International Journal of Slavic Linguistics and Poetics* 15, 95–125.

Jakobson, Roman (1948). Russian conjugation. *Word* 4, 155–167.

Keller, William (1995). DATR theories and DATR models. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 55–62. Cambridge, MA: Association for Computational Linguistics.

Likova, A. G. (1959). Obrazovanie imen suščestvitel'nyz so značeniem lica v sovremennom russkom jazyke. Unpublished doctoral dissertation, Moscow State University.

Matthews, Peter (1991 [1974]). *Morphology*, 2nd ed. Cambridge: Cambridge University Press.

Panov, M. V. (1968). *Russkij jazyk i sovetskoe obščestvo: slovoobrazovanie sovremennogo russkogo literaturnogo jazyka*. Moscow: Nauka.

Sadler, Louisa; Spencer, Andrew; and Zaretskaya, Marina (1994). The "third stem" in Russian. In *Working Papers in Language 44*. Colchester: University of Essex.

—; Spencer, Andrew; and Zaretskaya, Marina (1997). A morphomic account of a syncretism in Russian deverbal nominalizations. In *Yearbook of Morphology 1996*, Geert Booij and Jaap van Marle (eds.), 181–215. Dordrecht: Kluwer.

Shapiro, Michael (1967). Concatenators and Russian derivational morphology. *General Linguistics* 7(1), 50–66.

Stankiewicz, Edward (1962). The interdependence of paradigmatic and derivational patterns. *Word* 18, 1–22.

Stump, Gregory (1993). How peculiar is evaluative morphology? *Journal of Linguistics* 29, 1–36.

—(1998a). Paradigm function morphology. Unpublished manuscript, University of Kentucky.

—(1998b). Inflection. In *Handbook of Morphology*, Andrew Spencer and Arnold Zwicky (eds.), 13–43. Oxford: Blackwell.

Švedova, N. J. (ed.) (1980). *Russkaja grammatika*, vol. 1. Moscow: AN SSSR.

Timberlake, Alan (1993). Russian. In *The Slavonic Languages*, Bernard Comrie and Greville Corbett (eds.), 827–86. London and New York: Routledge.

Townsend, Charles (1975). *Russian Word-Formation*. Columbus, OH: Slavica.

Trubeckoj, N. S. (1934). *Das morphonologische System der russischen Sprache*. Prague: Travaux du cercle linguistique de Prague.

Ušakov, D. N. (1935–1940). *Tolkovyj slovar' russkogo jazyka*, 4 vols. Moscow.

Vinogradov, V. V. (1971). *Russkij jazyk (grammatičeskoe učenie o slove)*. Moscow: Academic International.

—; Istrina, E. S.; and Barxudarov, S. G. (eds.) (1953). *Grammatika russkogo jazyka*, vol. 1: *Fonetika i morfologija*. Moscow: AN SSSR.

Worth, Dean (1967). The notion of "stem" in Russian flexion and derivation. In *To Honor Roman Jakobson: Essays on the Occasion of his Seventieth Birthday*, vol. 3, 2269–2288. The Hague: Mouton.

Zemskaja, E. A. (1964). Interfiksacija v sovremennom russkom slovoobrazovanii. In *Razvitie grammatiki i leksiki sovremennogo russkogo jazyka*, I. P. Mučnik and M. V. Panov (eds.), 36–62. Moscow: AN SSSR.

Zwicky, Arnold (1992). Some choices in the theory of morphology. In *Formal Grammar; Theory and Implementation*, Robert Levine (ed.), 327–371. Oxford: Oxford University Press.

—(1996). The architecture of morphology. Paper presented at the sixth ESRC seminar series "Frontiers of Research in Morphology," University of Sussex, April.