

Classic problems at the syntax-morphology
interface: whose are they?

Greville G. Corbett

Surrey Morphology Group
University of Surrey

Proceedings of the HPSG10 Conference

Université Paris Diderot, Paris 7, France

Stefan Müller (Editor)

2010

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

There are fascinating problems at the syntax-morphology interface which tend to be missed. I offer a brief explanation of why that may be happening, then give a Canonical Typology perspective, which brings these problems to the fore. I give examples showing that the phenomena could in principle be treated either by syntactic rules (but these would be complex) or within morphology (but this would involve redundancy). Thus ‘non-autonomous’ case values, those which have no unique form but are realized by patterns of syncretism, could be handled by a rule of syntax (one with access to other features, such as number) or by morphology (with resulting systematic syncretisms). I concentrate on one of the most striking sets of data, the issue of prepositional government in Latvian, and outline a solution within Network Morphology using structured case values.

1 Background¹

Syntacticians have devoted considerable effort to understanding the constraints on the distribution of features. Less effort has gone into justifying the feature inventories for particular languages. This was a concern of members of the Set-theoretical School, a tradition which is of continuing relevance (see van Helden 1993 and Meyer 1994 for an overview). The work of Zaliznjak is particularly useful for our topic (e.g. Zaliznjak 1973), since he highlights problems whose solution involves complicating either the syntax or the morphology. Two later trends have conspired to background the

¹ The support of the European Research Council (grant ERC-2008-AdG-230268 MORPHOLOGY) is gratefully acknowledged. I wish to thank especially Matthew Baerman and Axel Holvoet for several very helpful discussions of the issues, Dunstan Brown for his insights on the Network Morphology analysis, and Patricia Cabredo Hofherr, Marina Chumakina, Sebastian Fedden, Andrew Hippisley, Uwe Junghanns, Aleksandr Krasovitsky, Marianne Mithun, Enrique Palancar, Adam Przepiórkowski, Ivan Sag, Lameen Souag, Greg Stump, Claire Turner and Martin van Tol, for various comments and suggestions. The paper is an interim report on an ongoing project. Versions were read at MOWL (Morphology of the World’s Languages), Leipzig, June 2009, at the Fourth conference of the Slavic Linguistics Society, University of Zadar, September 2009, at the Meeting of the International Commission on the Grammatical Structure of the Slavonic Languages, University of Kraków, September 2009 and at the Workshop on morphology and formal grammar, 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, 8-10 July 2010. My thanks to those present for their reactions. Errors are mine.

problems I address. The first is the trend within formal grammar towards simpler syntax. This started with work on Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum & Sag 1985), which demonstrated how much can be achieved using a leaner theoretical apparatus. It was also explicit in appropriately limiting the scope of syntax, which means that the issues I shall raise appear to some to fall outside syntax. The second is the growing acceptance of morphology as a component deserving of separate study, with its own issues. Some therefore concentrate on core morphological issues, leaving aside boundary problems. Hence the potential for a crack, down which complex and interesting issues may fall.

2 An example

As a brief illustration, the Russian preposition *po*, which expresses a wide span of meanings, has the following behaviour for some speakers/writers (there is ongoing variation). Specifically in the phrase *skučat' po* 'to long for, miss', we find *po* (in one system at least) with the dative of nouns and the locative of pronouns (see also Iomdin 1991):²

Russian (from the writings of Andrej Platonov 1899-1951)

- (1) skuča-l-a po rebenk-u (not: *po rebenk-e* in this corpus)
 miss-PST-SG.F for child-SG.DAT
 'missed (her) child'
- (2) skuča-et po nem (not: *po nemu* in this corpus)
 miss-3SG for 3SG.LOC
 'is missing him'

Other prepositions do not behave in this way; thus *k* 'towards' governs the dative, of nouns and pronouns alike, while *o* 'about, concerning' governs the locative, of both nouns and pronouns. What then can we make of (1) and (2)? There are at least two analyses. According to the morphological approach, we can say that there is an extra case value (call it the DAT-LOC). It has no unique form, being syncretic with the dative for nouns and the locative for pronouns. The disadvantage of this analysis is that we have introduced an extra case value just for a few such expressions; moreover the extra case value has no separate form, it is 'non-autonomous' (Zaliznjak 1973: 69-74). The alternative, the syntactic approach, requires a rule of government which is certainly not simple, since it needs to specify different values for phrases

² *Po* is challenging in its various senses and in different Slavonic languages; see for example Przepiórkowski (2008) on Polish.

according to the type of their head. In contrast, the normal situation in Russian is that government operates ‘canonically’, without reference to a noun/pronoun distinction, or to any other part of speech.

Thus we have two types of analysis, and it is not self-evident which is to be preferred. This is one example of several such interface problems, which are our topic.

3 Canonical typology

As a tool for identifying and highlighting such examples, we adopt the approach of Canonical Typology. Adopting a canonical approach means that we take definitions to their logical end point, and this enables us to build theoretical spaces of possibilities. Only then do we investigate how this space is populated with real instances. Canonical instances are those that match the canon: they are the best, clearest, the indisputable ones. Given that they have to match up to a logically determined standard, they are unlikely to be frequent. They are more likely to be rare, and may even be non-existent. This is not a difficulty. The convergence of criteria fixes a canonical point from which the phenomena actually found can be calibrated.

4 Canonical morphosyntactic features and their values

We set out an idealized world, and then concentrate on phenomena that “ought” not to happen, particularly those where there are two solutions, both troublesome.

Canonical morphosyntactic *features and values* have been described in terms of two overarching principles (covering ten converging criteria). The important part for our analysis is the two principles given here (detail on the criteria can be found in Corbett 2010).

Principle I (morphological):

Features and their values are clearly distinguished by formal means (and the clearer the formal means by which a feature or value is distinguished, the more canonical that feature or value).

Principle II (syntactic):

The use of canonical morphosyntactic features and their values is determined by simple syntactic rules.

5 Classic problems: the two principles in conflict

We find interesting problems when our two principles are in conflict; consider first this paradigm from Classical Armenian:

SINGULAR PLURAL		
azg	azgk'	NOMINATIVE
azg	azgs	ACCUSATIVE
azgi	azgs	LOCATIVE
azgi	azgac'	DATIVE

Classical Armenian *azg* ‘people’ (from Baerman 2002)

Figure 1: Non-autonomous case value

In this example, which is more general than the limited Russian instance above, there is no unique form for the accusative; its forms are always syncretic. There are two alternatives:

- we recognize an accusative case value. At the same time, we accept that it is a less canonical feature value than the nominative or dative. It is non-autonomous, and so it goes against Principle I, the morphological principle.
- we have a rule of syntax, which states that transitive verbs govern the nominative for singular NPs and the locative for plural NPs. This avoids having a non-autonomous case value, but it goes against Principle II, the syntactic principle, in requiring a complex syntactic rule.

Faced with such issues, the more usual choice in recent times has been to opt for simple syntax, and thus to accept a non-autonomous case value. There are fully analogous instances with other morphosyntactic features: gender, and person. For explicit discussion of alternative analyses in comparable but not identical circumstances see Goddard (1982) and Fedden (2007).

6 A canonical space for morphosyntax

We now move on to some new morphosyntactic criteria, in addition to the ten covered by the two principles above, hence numbered 11 to 15. Each of these criteria in different ways can be seen as exemplifying and maintaining the principle of **simple syntax**. They are listed for completeness; the most important for present purposes is Criterion 13.

6.1 Canonical government: governors govern

Criterion 11: A canonical rule of government consists of what the governor requires and the domain of government (and only that).

6.2 Canonical agreement: controllers control agreement

Criterion 12: A canonical rule of agreement consists of the feature specification of the controller and the domain of agreement (and only that).

6.3 Canonical interaction: morphosyntactic features ‘mind their own business’

Criterion 13: The distribution of morphosyntactic feature values is constrained by the rules of government and agreement; it is not canonical for the values of other morphosyntactic features to have a role.

6.4 Canonical interaction of part of speech classifications and features: no effect on feature values

Criterion 14: Part of speech classification is accessible to morphosyntactic features; it is not canonical for it to be accessible to determine their values.

6.5 Canonical limit on lexical eccentricity

Criterion 15: Lexical items may have idiosyncratic inherent specification but may not canonically have idiosyncratic contextual specification.

7 The classic morphosyntactic problem: Latvian

The Baltic language Latvian deserves special attention since there are several conflicting lines of argument. To get to grips with the issues, it makes sense to start from the way in which the data are typically presented:

- (3) Latvian noun paradigm (typical presentation: Veksler & Jurik 1978: 25)

<i>galds</i> 'table'	SINGULAR	PLURAL
NOMINATIVE	<i>gald-s</i>	<i>gald-i</i>
GENITIVE	<i>gald-a</i>	<i>gald-u</i>
ACCUSATIVE	<i>gald-u</i>	<i>gald-us</i>
INSTRUMENTAL	<i>gald-u</i>	<i>gald-iem</i>
DATIVE	<i>gald-am</i>	<i>gald-iem</i>
LOCATIVE	<i>gald-ā</i>	<i>gald-os</i>

The key point is that the instrumental singular is syncretic with the accusative, while the instrumental plural is syncretic with the dative. This is not something special about this class of noun; the same pattern of syncretism runs right through the language, including the personal pronouns. In fact there are no uniquely instrumental forms, hence if we assumed an instrumental case value it would be non-autonomous.

The instrumental, if recognized, is almost always found together with the preposition *ar* 'with'. If we do not recognise the instrumental, then we have a preposition, *ar* 'with', which takes different case values according to whether the governed element is in the singular or the plural. Such a situation is not what we expect, and it is not 'simple syntax'. Now consider these examples (Veksler & Jurik 1978: 87, and compare the discussion in Fennell 1975 and Holvoet 1992):

- (4) Grūti dzīvot bez draug-a
hard live.INF without friend-SG.GEN
'It's hard to live without a friend.'
- (5) Grūti dzīvot bez draug-iem
hard live.INF without friend-PL.DAT/INS
'It's hard to live without friends.'

We see that other prepositions, according to the traditional account, take different case values in the singular and plural. In fact all prepositions take the dative (=instrumental) in the plural.

There are good arguments for not recognising an instrumental case value. We could simply say that *ar* is a preposition which takes the accusative in the singular and which, when it governs a plural, behaves like all other prepositions in taking the dative plural (as do those which everyone agrees take the accusative in the singular). However, this approach flies in the face of the notion of simple syntax, since it goes against Criterion 13 (§6.3).

If, however, we wish to maintain a simple rule of government, we need to recognize a non-autonomous case value, governed by prepositions like *ar* ‘with’ and *par* ‘about’; we could even call it ‘instrumental’, but for clarity here let us label it ACC-DAT. This looks like the traditional position. Left like this, the analysis is hardly tenable. The problem is the prepositions like *bez* ‘without’, as in (4) and (5), which similarly take the dative in the plural. To have a simple rule of government we need to recognize a further case value, the GEN-DAT. We do not, of course, need a third for the dative, since here the same value is found in the singular and the plural. Thus our rule of government can be simple, provided we accept the cost of having an additional two non-autonomous case values. The issues are interesting in their own right, but also more generally, as an illustration of interface problems which need to be considered from the perspective of simple syntax and a clearly-defined morphology.

8 Towards an analysis

There have been several attempts to analyse the Latvian data, based on different (often implicit) assumptions about syntax and morphology. The previous sections have clarified our assumptions somewhat, and we should attempt to tackle the problem from both the syntactic and the morphological direction.

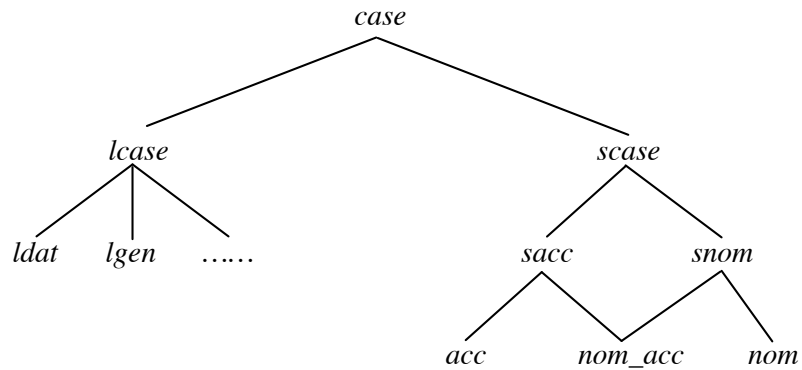
8.1 Syntax: HPSG

There are ideas within the HPSG literature that appear promising and relevant (thanks to Ivan Sag for pointing these out). First, Levine, Hukari & Calcagno (2001: 205) investigate parasitic gap examples like this:

- (6) Robin is someone who even good friends of believe should be closely watched.

They need to allow an item like English *who* to be both accusative and nominative; their solution involves a novel sort hierarchy for case (2001: 207-210):

(7)

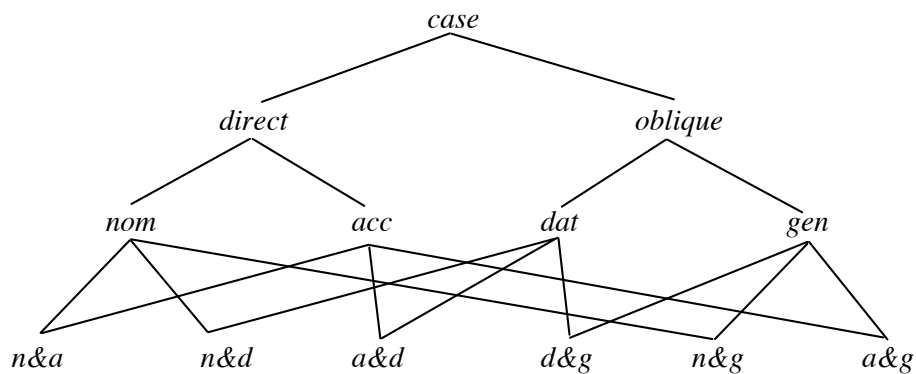


The interest is in the *scase* (structural case) part of the hierarchy. It includes an additional case value *nom_acc* and this satisfies any selectional requirement for nominative and accusative. In this approach, saying that a verb assigns *snom* to its subject is an abbreviation for saying it takes *nom* or *nom_acc*. The relevant forms have these specifications:

- (8) whom [CASE *acc*]
 who [CASE *nom_acc*]

Sag (2003) takes this further, when analysing coordinate structures where the conjuncts have different feature specifications (including the well-known German examples involving different case values). For these he proposes the following hierarchy of types (Sag 2003: 278):

(9)



If we think of the Russian *po* construction (§2), and stay with the simplest scenario assuming a rigid distinction between nouns and pronouns, we could propose an additional case value *dative_locative*; this would be the case value required by *po* in the construction we examined. But this leaves a substantial problem: *po* does not simply take any item that is dative or locative; we still need to specify that *dative_locative* is identical to pure dative for nouns and pure locative for pronouns.

There are two important points for our purposes. First, these analyses involve adding feature values; the syntax is kept simple, and there are additional feature values which introduce complications into the morphology. Moreover there is a relaxation of the standard HPSG assumptions; the requirement that feature structures be sort-resolved is abandoned (Sag 2003: 274). And second, the examples we have been examining are in one respect more challenging than those which have figured to date in the HPSG literature cited in this section: the extra dimension is that the additional values do not apply generally. Thus the Russian problem of government of *po* involved part of speech (noun versus pronoun in the simplest instance), while Latvian involved number.

More generally, the issue is not one of special syntactic constructions, as have figured in the instances those authors deal with, it is one of getting the right inflectional form. In some instances this form is clear-cut and not subject to variability. The particular problems we have concentrated on involve prepositions (there are comparable examples in other languages which do not, however); we could look for a ‘weakened’ featural requirement specifically for prepositions, which need not bring with it a general relaxation of the feature system. In other words, an analysis that pins the difficulty on the governor would be attractive.

Thus we should consider: (a) how we set up the features; (b) whether we can tie any special device uniquely to the case controller. With these possibilities in mind we turn to the morphology.

8.2 Morphology: Network Morphology

We look for an analysis within Network Morphology, which is an inferential-realizational theory; see, for example, Corbett & Fraser (1993), Evans, Brown & Corbett (2002), Brown & Hippiusley (in progress). A bibliography of work in this framework can be found at: [http://www.surrey.ac.uk/LIS/SMG/Network Morphology Bibliography.htm](http://www.surrey.ac.uk/LIS/SMG/Network%20Morphology%20Bibliography.htm). Network Morphology gives a central place to defaults, which are layered,

and analyses are represented and implemented in the lexical knowledge representation language DATR (Evans & Gazdar 1996).

Idea 1: In Latvian, certain prepositions take the *accdat* case, but *nothing* has an *accdat* case value, that is, no lexical entry includes a form with this featural description. (We use *accdat* to make it clear that this is an atomic value.) High in the morphological hierarchy, we could have statements of this type:

MOR_NOMINAL:

```
<mor pl accdat> == "<mor pl dat>"
<mor sg accdat> == "<mor sg acc>"
```

The effect is that any nominal (noun, pronoun or adjective) for which the *accdat* is required, will “provide” the dative if plural, and the accusative if singular.

The architecture of Network Morphology theories involves different hierarchies, related to each other by defaults. The morphological hierarchy just mentioned accounts for the lexeme’s purely morphological behaviour, while the lexemic hierarchy takes care of its interface to syntax. (They are comparable to the content paradigm and form paradigm of Paradigm Function Morphology, earlier known as the morphological and syntactic paradigm, see Stump 2002: 149-153, 178.)

An alternative (Dunstan Brown, personal communication) is to state the regularity in the lexemic hierarchy:

NOMINAL:

```
<syn pl accdat> == "<mor pl dat>"
<syn sg accdat> == "<mor sg acc>"
```

In both, a similar rule is necessary for the *gendat* of course. This means that we miss the generalization that all prepositions take the dative in the plural.

This has the advantage of placing the statement right on the syntax-morphology interface. The Latvian data do not offer any unambiguous pointer as to which hierarchy is the preferable place; this is another instance of how uniquely tricky the Latvian data are. (Some other comparable instances may prove more helpful here in having specific morphological quirks, which would suggest the correct place is the morphological hierarchy.)

Idea 2: We have a ‘structured case’; prepositions in Latvian take *prep acc*, *prep gen* or *prep dat* (for discussion of structured case values see Brown 2007 and Corbett 2008: 17-22). No noun, adjective or pronoun has a prepositional case form. In the lexemic hierarchy, we have these equations:

NOMINAL:

```
<syn pl prep> == "<mor pl dat>"  
<syn sg prep> == "<mor sg>"
```

The first line states the surprising fact: all plural nominals governed by a preposition stand in the dative. The second states compactly that government in the singular is fully usual: any extension of the path on the left will also occur on the right. Thus, from the second line can be inferred (it is not stated explicitly):

```
<syn sg prep acc> == "<mor sg acc>"  
<syn sg prep gen> == "<mor sg gen>"  
<syn sg prep dat> == "<mor sg dat>"
```

This analysis has several advantages. The feature system is made more complex for one case value only, the structured prepositional case, which exists alongside the remaining simple case values (nominative, accusative, genitive, dative, instrumental, locative). Structured cases are established as necessary in analyses of other languages. In Latvian, the structured prepositional case can be governed only by prepositions. It is non-autonomous: the realization of its values is mediated through the lexemic hierarchy, which locates the issue appropriately at the syntax-morphology interface. Thus we recognize the additional values (available for government by prepositions only), in order to keep the syntax simple, but they are dealt with by the lexemic hierarchy; no lexical item has a separate form for these values, as shown by the fact that they do not appear in the morphological hierarchy.

9 Conclusion

These data at the syntax-morphology interface present remarkable analytical challenges. They are thrown into relief by the Canonical Approach. The general point is that these unusual but recurring interface phenomena require a combined approach, rather than being allowed to escape the attention of both syntacticians and morphologists. The specific outcome is that we can

treat the Latvian case problem using structured case values: the syntax remains simple, there is a complication of the feature system, and this is linked specifically to the case governor, the preposition. The structured case values have no additional morphological forms and the patterning of forms is handled, in the morphology, using a Network Morphology approach.

References

- Baerman, Matthew. 2002. Armenian. In: Surrey Syncretisms Database. Available at: <http://www.smg.surrey.ac.uk/>
- Brown, Dunstan. 2007. Peripheral functions and overdifferentiation: the Russian second locative. *Russian Linguistics* 31.61-76.
- Brown, Dunstan & Andrew Hippisley. In progress. *Network Morphology*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2008. Determining morphosyntactic feature values: the case of case. In: Greville G. Corbett & Michael Noonan (eds) *Case and grammatical relations: Studies in honor of Bernard Comrie*, 1-34. Oxford: Oxford University Press.
- Corbett, Greville G. 2010. The penumbra of morphosyntactic feature systems. In: Jonathan Bobaljik, Uli Sauerland & Andrew Nevins (eds) *Markedness in the Morphosemantics of ϕ -Features* (special issue of *Morphology*). Published online 19 June 2010.
- Corbett, Greville G. & Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian inflectional morphology. *Journal of Linguistics* 29.113–42. [Reprinted 2003 in: Francis X. Katamba (ed.) *Morphology: Critical Concepts in Linguistics, VI: Morphology: Its Place in the Wider Context*. London: Routledge, 364-396.]
- Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 2002. The semantics of gender in Mayali: Partially parallel systems and formal implementation. *Language* 78.111-155.
- Evans, Roger & Gerald Gazdar. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22.167–216.
- Fedden, Sebastian. 2007. Women, houses, and plural objects? – Homophony in the Mian gender system. In: Robyn Loughnane, Cara Penry Williams & Jana Verhoeven (eds) *In Between Wor(l)ds. Transformation and Translation* (=School of Languages and Linguistics Postgraduate Research Papers on Language and Literature, vol. 6), 183-198. Melbourne: University of Melbourne, School of Language and Linguistics.
- Fennell, Trevor G. 1975. Is there an instrumental case in Latvian? *Journal of Baltic Studies* 6.41-48.

- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Blackwell: Oxford
- Goddard, Cliff. 1982. Case systems and case marking in Australian languages: a new interpretation. *Australian Journal of Linguistics* 2.167-196.
- Helden, W. Andries van. 1993. *Case and gender: Concept formation between morphology and syntax (II volumes)* (Studies in Slavic and General Linguistics 20). Amsterdam: Rodopi.
- Holvoet, Axel. 1992. Bemerkungen über die Entwicklung des lettischen Kasussystems: der Instrumental. In: Wojciech Smoczyński & Axel Holvoet (eds) *Colloquium Pruthenicum Primum*, 143-149. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego,
- Iomdin, Leonid L. 1991. Slovarnaja stat'ja predloga "PO". *Semiotka i informatika* 32.94-120.
- Levine, Robert D., Thomas E. Hukari & Michael Calcagno. 2001. Parasitic Gaps in English: Some overlooked cases and their theoretical consequences. In: Peter W. Culicover & Paul M. Postal (eds) *Parasitic Gaps* (Current Studies in Linguistics 35), 181-222. Cambridge, MA: MIT Press.
- Meyer, Peter. 1994. Grammatical categories and the methodology of linguistics: Review article on van Helden, W. Andries (1993) *Case and gender: concept formation between morphology and syntax*. *Russian Linguistics* 18.341-377.
- Przepiórkowski, Adam. 2008. Generalised quantifier restrictions on the arguments of the Polish distributive preposition PO. *Studia Kognitywne* 8.159-177.
- Sag, Ivan A. 2003. Coordination and underspecification In: Jong-Bok Kim & Stephen Wechsler (eds) *Proceedings of the Ninth International Conference on Head-Driven Phrase Structure Grammar, Kyung-Hee University, Seoul*. Stanford: CSLI Publications.
- Stump, Gregory T. 2002. Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In: Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 2001*, 147-180. Dordrecht: Kluwer.
- Veksler, Bunim X. & Jurik, Vladimir A. 1978. *Latyškij jazyk (samoučitel')*. Riga: Zvajgzne. [Third edition.]
- Zaliznjak, Andrej A. 1973. O ponimanii termina 'padež' v lingvističeskix opisanijax. In: Andrej A. Zaliznjak (ed.) *Problemy grammatičeskogo modelirovanija*, 53-87. Moscow: Nauka. [Reprinted in: Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*, 613-647. Moscow: Jazyki slavjanskoj kul'tury.]