

## MORPHOSYNTACTIC FEATURES: THE SPECIAL CONTRIBUTION OF THE SLAVONIC LANGUAGES<sup>1</sup>

### 1. Introduction

This paper is concerned with the typology of features, particularly that of morphosyntactic features. This is a topic which has been often discussed in Slavonic linguistics. And this is with good reason because, as we shall see, the Slavonic languages provide key data for making progress in this area. From the considerable literature on the subject we shall take one major concept, obligatoriness (§2) and then discuss one important body of work, that of the Set-theoretical School (§3). Then we consider a way forward, based on the canonical approach in modern typology (§4). And then we shall examine three sets of Slavonic data (§§5–7) for what they show us for a typology of features. We shall treat gender, number and so on as ‘features’ – they are often called ‘categories’ in the Slavonic tradition, but the latter term is used increasingly for lexical categories (parts of speech). When we deal with particular genders (like feminine), or numbers (like dual) we shall term these ‘values’.

### 2. Obligatoriness

A notion which has an important place in the discussion of features, and in the distinction between inflectional and derivational morphology, is obligatoriness. The well-known quotation belongs to Jakobson: “Thus the true difference between languages is not in what may or may not be expressed but in what must or must not be conveyed by the speakers.” (1959[1971]:492). The point is this: it is not remarkable that, for example, Russian can make a distinction between *žurnal* ‘magazine’ and *žurnaly* ‘magazines, more than one magazine’. The interesting thing is that speakers of Russian are forced to make this distinction. There is no natural way to avoid number in Russian. We say therefore that number is inflectional in Russian. Moreover, since both morphological and syntactic rules (agreement) have access to it, we say that it is a morphosyntactic feature, on the strict definition of morphosyntactic.

Jakobson made the point in discussing Boas, and he gives Boas (1938:132) as the source. The idea surfaces at different times. For instance, Mel’čuk (1960[1974]) discusses this criterion and in (1974:111) points out that he, Mel’čuk, wrote the article in 1958, before seeing Jakobson’s article. It has been suggested that there is an earlier source, in Maspero’s work on Chinese (see Maspero 1934:35). It is Percov (1996:40, 2001:71) who drew attention to Maspero’s early discussion of obligatoriness. However, I think Jakobson is right to give precedence to Boas, since Boas discusses the notion in the Introduction to the *Handbook of American Indian languages* (1911:35–43, especially 40–43). Some of the difficulties with the notion are raised in Corbett (1999). In what follows we consider two of these in §5 and §6.

---

<sup>1</sup> The support of the ESRC under grant RES-051-27-0122 and of the AHRC under grant B/RG/AN4375/APN18306 is gratefully acknowledged. A version of the paper was read at the Perspectives on Slavistics conference, held in September 2006 in Regensburg. I would like to thank the organizers for the invitation, those present for interesting and wide-ranging discussion and Anna Kibort for helpful comments on the written version.

### 3. The Set-theoretical School

At various places in the literature, for instance in Zwicky (1986:988–989) there is the suggestion that natural languages have a restricted inventory of features on which they can draw, and that these features in turn have a restricted inventory of values. Progress towards validating this claim has been limited. However, we should note the interesting and valuable work done by members of the loose grouping sometimes known as the Set-theoretical School. There is interesting material on the beginnings of the group in Uspenskij (1957), notably on the early discussions which were stimulated by questions posed by Andrej Kolmogorov: see van Helden (1993:138) for sources. This work in the Set-theoretical School led to important contributions by Zaliznjak (1967; 1973), which are of continuing value. Indeed, some substandard work in typology might have been avoided if Zaliznjak's contribution had been more widely known. One reason for his continuing relevance is his modern approach to morphology (see Krylov 2002:705). Several of the issues raised by Zaliznjak are taken up by Mel'čuk (1986[2006]). An extensive and sympathetic technical survey of the work of the Set-theoretical School can be found in van Helden (1993). In turn there is an insightful review of van Helden (1993) by Meyer (1994).

### 4. The canonical approach in typology

A typical result from set-theoretical approaches is that the expected values are confirmed, and that in addition other values, of less clear standing emerge. How can we go further in this situation? One way forward is to adopt a 'canonical' approach. We extrapolate from what there is to what there might be, to set up a consistent theoretical schema. And within that theoretical space we can situate the real instances we have found. One effect of this canonical approach is to separate out coincidental overlappings in the examples that exist; we may then ask which characteristics coincide by chance and which cooccur of necessity.

Consider the Russian form *mal'čika* 'boy'. According to context this could be an instance of the accusative or of the genitive. If every form were like that, we would have no evidence for distinguishing genitive from accusative. And if we go along that route to the logical end point, if we had no means of distinguishing cases we would have no case system. Contrast that with a form like *mal'čiku* 'boy'. Irrespective of context this must be an instance of the dative case. It is easy to recognize that the situation found with *mal'čiku* '(to a) boy' is canonical. It shows a unique mapping from form to function, which is canonical. If we found a system in which every form of every nominal were like *mal'čiku* in this respect, we would have no difficulty in agreeing that this was an instance of a case system, and we could readily establish the number of values. It is immaterial at this stage in the argument whether such a system exists: the essential thing is that we can define it, we could recognize it with certainty if we found it, and so it gives us a measure of canonicity according to which we can calibrate the instances of values in the system which we are examining.

Having looked at the form side of the issue we can consider the function side in the same way. Imagine a language whose case values were all determined by simple syntactic rules equivalent to: 'the direct object of a transitive verb stands in the accusative case'. We could recognize and agree about such a system, and it would indeed be canonical. It is trickier to analyse the numerous deviations from such simple syntactic rules: it is easy to see the deviation, but it is much harder to demonstrate whether we have a slightly different syntactic structure, or a semantic condition, and so on. The

point is that having the logically possible canonical system in mind as a standard can be of value whether or not we find examples of it.

The canonical approach requires clear definitions. By taking these to their logical end points, we construct a theoretical space. The convergence of criteria fixes a canonical point from which we can calibrate the phenomena we find. The instances which would qualify as canonical according to our definition will almost certainly not be frequent. This is fully expected: in the neat formulation by Johanna Nichols (personal communication): “Canonical constructions are all alike; each non-canonical construction is non-canonical in its own way.”<sup>2</sup> The canonical approach has shown its effectiveness for both syntax and morphology (see Corbett 2007 for references). It allows us to handle gradient phenomena in a principled way; particular values can be treated as more or less canonical, rather than having an ‘all or nothing’ requirement.

## 5. Case in Russian

Russian case raises numerous difficulties and has deservedly been the subject of numerous studies. A detailed recent study is Corbett (forthcoming), which includes a survey of some of the literature. The main points of the article are summarized here. Ten criteria for canonicity are proposed. These can be summarized under two overarching principles:

Principle I: 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)

When formal means are ‘clear’ this means that they allow a straightforward and regular mapping from form to function. It follows that in the canonical situation we have clear evidence both for the case feature and for each of its values.

Principle II: The use of canonical morphosyntactic features and their values is determined by simple syntactic rules

This principle relates to the interface between syntax and morphology. In the canonical situation, the rules of syntax determine the required case value, and the morphology realizes this value without further complications.

Given such a set of requirements, the traditional case values of Russian are confirmed, though some are more canonical than others. For instance, the accusative case is in many instances not clearly distinguished by formal means. And beyond the traditional case values, there are other values of less certain status: the second locative, the second genitive, the vocative and the adnumerative. These have various non-canonical characteristics. Instead of a homogeneous case system which is the way in which the Russian case system is sometimes presented, we have rather a set of values of differing status, some being close to canonical, and some showing non-canonical and sometimes even quite exotic behaviour.

In terms of obligatoriness too, we find a mixed picture. While case as a feature is indeed obligatory in Russian, not all of the values are. For instance, for several nouns which have a second genitive its use is facultative.

---

<sup>2</sup> Compare: “All happy families are alike; each unhappy family is unhappy in its own way.” (Lev Tolstoy, *Anna Karenina*).

## 6. Number in Russian

Number is often treated as a straightforward morphosyntactic feature; however, there is a tradition of discussion within Slavonic linguistics, which shows that its status is not so clear. If we consider just the notion of obligatoriness, we soon see the difficulty. Taking Russian as an example, it is evident that not all nouns distinguish number. There are many nouns outside the number distinction, the non-count nouns. This is well known, of course. The question is whether the existence of the non-count nouns is as unproblematic as is often assumed.

We are used to lexical exceptions of various types. However, here we are dealing with something more important and more systematic. It is not just an issue of which nouns show or do not show number; in addition we have agreement of various targets. In the latter instance we are dealing with contextual inflection, in Booij's terms (1996). Hence we are dealing with a genuine morphosyntactic feature. I suggest that, when contextual inflection is involved, splits in the lexicon must be 'principled'. Thus we find nouns like *sani* 'sledge', which are inherently plural, and so control plural agreement. But we do not find the converse, which would be a noun which had the full range of morphological distinctions but arbitrarily controlled only plural agreement.

The fact that the number opposition is not available for a substantial part of the nominal lexicon is indeed principled. It is constrained by the Animacy Hierarchy (see Smith-Stark (1974), Corbett (2000)):

(1) The Animacy Hierarchy

speaker > addressee > 3 person > kin > human > animate > inanimate

The basic claim from the Animacy Hierarchy concerning number is that the items which distinguish number must form some top segment of the hierarchy. Thus if in a language the nouns denoting (non-human) animates distinguish number, then all the items to the left on the hierarchy will also distinguish number. The cut-off point varies dramatically from language to language. The Animacy Hierarchy is thus the basis for splitting the noun lexicon into count and non-count nouns. This is a simplification since, as Allan (1980) points out, number is a property of noun phrases rather than of nouns. He goes on nevertheless to point out that nouns have number preferences (and it is these which are constrained by the Animacy Hierarchy).

For Slavonic languages the boundary comes within the inanimate nouns, and there has been interesting work on the treatment of nouns denoting fruit and vegetables. Simplifying somewhat, we may say that Russian *kartofel'* 'potatoes', *vinograd* 'grapes', *kljukva* 'cranberries', *gorox* 'peas', *izjum* 'raisins', and many more like them, do not distinguish singular and plural (in some cases there are derived forms which do). On the other hand, *frukt* 'fruit' has singular and plural forms. Russian then sets the boundary for number-differentiability somewhat higher than English, and indeed a little higher than some other Slavonic languages (see Ivić 1982). There is interesting detail on this topic, for which see Mel'čuk (1979; 1985:257–64), Polivanova (1983), Jarvis (1986) and Wierzbicka (1988:503–506).

Beyond the simple split between nominals which distinguish number and those which do not, there is the interesting phenomenon of what Kulikov (2004:127) calls 'number-orientedness', referring to work by Polivanova (1983). The latter paper is significant; further data can be found there and in Ljaševskaja (2004). Consider these data:

Russian (Kulikov 2004:127, following Polivanova 1983)

- (2) rep-a                    doroža-et  
 turnip-SG.NOM    get\_dearer-3SG  
 'turnips are getting dearer'
- (3) ogurc-y                    dorožaj-ut  
 cucumber-PL.NOM    get\_dearer-3PL  
 'cucumbers are getting dearer'

Both *repa* 'turnip' and *ogurec* 'cucumber' are count nouns. They have singular and plural forms available, particularly for contexts where a quantity of items is specified. In more general contexts, however, these and similar nouns have a number preference: for *repa* 'turnip' it is the singular, while for *ogurec* 'cucumber' it is the plural. These examples are again at the boundary of count and non-count, which is in accord with the Animacy Hierarchy. Whether this is always the case is a question worthy of further research.

Before leaving the Animacy Hierarchy we should note that it determines the distribution of meaning, too. Number values are not fully consistent in semantic terms. Sometimes, for instance, we find associative readings. These are found particularly at the top of the Animacy Hierarchy. And more generally, the different possible readings are determined by the hierarchy (Corbett 2000:83–87). Thus, though number in Russian is far from straightforward, the complications we find are not mere quirks, but are principled, in that they are constrained by the Animacy Hierarchy.

## 7. Number in Slovene

Slovene has grammatical number which is comparable to that of Russian in being obligatory, and as in Russian it is relevant both to morphology and to syntax (through agreement) and hence is a genuinely morphosyntactic feature. It differs from Russian in having a dual value, and this third value allows much fuller insight into the feature. While number is obligatory, it might be argued that not all values are. Consider this example:

Slovene (Priestly 1993:440–441)

- (4) nóg-e    me            bolijo  
 foot-PL    1SG.ACC    hurt.3PL  
 'my feet hurt'

One solution is to say that number is indeed obligatory, but that the dual value is facultative. That is essentially the solution proposed in Corbett (2000:93–94). The dual, like the other number values, is involved in agreement (it has contextual uses). We predict that any limitations on its lexical range (that is, which nominals control which agreements) will be principled. That is, there will not be arbitrary lexical restrictions. This appears to be the case here: the dual is facultative for nouns lower on the Animacy Hierarchy:

- (5) Ranges of number values in Slovene (Corbett 2000:94).<sup>3</sup>

	1	>	2	>	3	>	kin	>	human	>	animate	>	inanimate
range of plural	■		■		■		■		■		■		■
range of dual	■		■		■		■		■		■		■

<sup>3</sup> Speaker and addressee are represented as '1' and '2' for reasons of space.

This is a reasonable solution. We have a split on the Animacy Hierarchy, in terms of number availability for the plural. There is a second split, involving the obligatory (■) or facultative (□) nature of the dual.

There is an alternative solution, which involves giving the dual a slightly more complex semantics. Since the semantics of number is generally relatively transparent, we may tend to assume that we know what a dual is. While we have no expectation that the neuter gender or genitive case will be identical from language to language, we may expect number values to coincide. In contrast to this view it may be that the Slovene dual, besides referring to two entities, also has an element ‘newsworthy’. In other words, it means not just ‘two’, but rather ‘two – and it matters’. We would have to explain the effect of (5) by saying that the items at the top of the Animacy Hierarchy are so significant that the number involved is always newsworthy. Lower on the hierarchy, this is not the case, and in examples like (4) the fact of there being two is not newsworthy at all, and so the dual is not used. For most instances, the two accounts yield the same predictions. However, the second predicts more variability according to situation than does the first (which implies a stronger link to specific lexical items according to their position on the Animacy Hierarchy). The evidence available does not allow us to choose between them; we must hope for further research on Slovene, which would help us to take the issue further. There are pointers in Derganc (2003) which could be taken as favouring the latter interpretation. For instance, she shows that the noun *starši* ‘parents’ was previously typically used in the plural (2003:174); in the modern language, the dual is also commonly used. The earlier use was an instance of number orientedness, between dual and plural, of a kind we could understand as an interaction of lexical meaning with the ‘newsworthy’ element of the feature value. This orientedness is weakening for this noun in the modern language. While such a line of reasoning looks promising, we are in danger of heaping too much theorizing on too few examples.

## Conclusion

Our discussion has highlighted the importance of Slavonic in making progress in understanding morphosyntactic features. In particular, we have seen the significance and difficulty of the notion ‘obligatoriness’, the value of a canonical approach, the interest of the idea of ‘orientedness’, and the possibility that number values should be given a more complex semantics than is usually allowed. Given the resources available for research into Slavonic languages, we may hope for further progress in these areas.

## References

- Allan, K. 1980. Nouns and countability. *Language* 56, 541–567.
- Boas, F. 1911. Introduction. In: Boas, F. (ed.): *Handbook of American Indian languages: I*, (Bureau of American Ethnology, Smithsonian Institution, Bulletin 40, part 1). Washington, DC, 1–83
- Boas, F. 1938. Language. In: Boas, F. (ed.): *General Anthropology*. Boston, Mass., 124–145.
- Booij, G. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In: Booij, G.; van Marle, J. (eds): *Yearbook of Morphology 1995*. Dordrecht, 1–16.
- Corbett, G. G. 1999. Prototypical inflection: implications for typology. In: Booij, G.; van Marle, J. (eds): *Yearbook of Morphology 1998*. Dordrecht, 1–22.
- Corbett, G. G. 2000. *Number*. Cambridge.
- Corbett, G. G. 2007. Canonical typology, suppletion and possible words. *Language* 83, 8–42.
- Corbett, G. G. forthcoming. Determining morphosyntactic feature values: the case of case. To appear in: Corbett, G.G.; Noonan, M. (eds.): *Case and grammatical relations: papers in honour of Bernard Comrie*. Amsterdam.
- Derganc, A. 2003. The dual in Slovenian. *Sprachtypologie und Universalienforschung* 56, 165–182.
- van Helden, W. A. 1993. *Case and gender: Concept formation between morphology and syntax* (II volumes). Amsterdam.

- Ivić, M. 1982. Slavic fruit and vegetable names and countability. *International Journal of Slavic Linguistics and Poetics* 25–26 (= Naylor, K.E. et al. (eds.): *Slavic Linguistics and Poetics: Studies for Edward Stankiewicz on his 60th Birthday, 17 November 1980*). Columbus, Ohio, 209–211.
- Jakobson, R. O. 1959[1971]. Boas' view of grammatical meaning. In: Goldschmidt, W. (ed.): *The anthropology of Franz Boas: Essays on the centennial of his birth* (Memoirs of the American Anthropological Association 89), Menasha, Wis., 139–145. [Reprinted in: Jakobson, R. 1971. *Selected Writings II*. The Hague, 489–496, and cited from the 1971 publication.]
- Jarvis, D. K. 1986. Some problems with noun number choice. *Slavic and East European Journal* 30, 262–270.
- Krylov, S. A. 2002. “Russkoe imennoe slovoizmenenie” A. A. Zaliznjaka tridcat' let spustja: opyt retrospektivnoj recenzii s pozicij neostrukturalistskoj morfologii. In: Zaliznjak, A.A. 2002. *Russkoe imennoe slovoizmenenie*. Moskva, 699–748.
- Kulikov, L. 2004. Review of Corbett, G.G. 2000, “Number”. *Southwest Journal of Linguistics* 23, 124–129.
- Ljaševskaja, O. N. 2004. *Semantika russkogo čisla*. Moskva.
- Maspero, H. 1934. La langue chinoise. *Conférences de l'Institut de Linguistique de l'Université de Paris, année 1933*, Paris, 33–70.
- Mel'čuk, I. A. 1974. Grammatical meaning in interlinguas for automatic translation and the concept of grammatical meaning. In: Rozenčevjg, Ju. V. (ed.): *Essays on Lexical Semantics: I*. Stockholm, 95–113. [original version: Mel'čuk, I. A. 1960. K voprosu o “grammatičeskom” v jazyke-posrednike. *Mašinnyj perevod i prikladnaja lingvistika* 4, 25–45.]
- Mel'čuk, I. A. 1979. Countability vs. non-countability. In: Clyne, P.R.; Hanks, W.F.; Hofbauer, C.L. (eds): *Papers from the Fifteenth Regional Meeting Chicago Linguistic Society, April 19–20, 1970*. Chicago, Ill., 220–227.
- Mel'čuk, I. A. 1985. *Poverxnostnyj sintaksis russkix čislovyx vyraženi*. Wien.
- Mel'čuk, I. A. 1986. Toward a definition of case. In: Brecht, R.D.; Levine, J. (eds.): *Case in Slavic*. Columbus, Oh, 35–85. [revised version in Mel'čuk 2006, 110–179].
- Mel'čuk, I. A. 2006. *Aspects of the Theory of Morphology*, edited by David Beck. Berlin.
- Meyer, P. 1994. Grammatical categories and the methodology of linguistics: Review article on van Helden, W. A. 1993, “Case and gender: concept formation between morphology and syntax”. *Russian Linguistics* 18, 341–377.
- Percov, N. V. 1996. Grammatičeskoe i objazatel'noe v jazyke. *Voprosy jazykoznanija* 1996(4), 39–61.
- Percov, N. V. 2001. *Invarianty v russkom slovoizmenenii*. Moskva.
- Polivanova, A. K. 1983. Vybor čislovyx form suščestvitel'nogo v russkom jazyke. In: Grigor'ev, V.P. (ed.): *Problemy strukturnoj lingvistiki*. Moskva, 130–145.
- Priestly, T. M. S. 1993. Slovene. In: Comrie, B.; Corbett, G.G. (eds.): *The Slavonic Languages*. London, 388–451.
- Smith-Stark, T. C. 1974. The plurality split. In: La Galy, M.W.; Fox, R.A.; Bruck, A. (eds.): *Papers from the Tenth Regional Meeting, Chicago Linguistic Society, April 19–21, 1974*. Chicago, Ill., 657–671.
- Uspenskij, V. A. 1957. K opredeleniju padeža po A. N. Kolmogorovu. *Bjulleten' Ob'edinenija po problemam mašinnogo perevoda* 5, 11–18. [Web version, with a postscript from 2001, available at: <[http://www.kolmogorov.pms.ru/uspensky-k\\_opredeleniyu\\_padezha\\_po\\_kolmogorovu.html](http://www.kolmogorov.pms.ru/uspensky-k_opredeleniyu_padezha_po_kolmogorovu.html)>, viewed 20 May 2006].
- Wierzbicka, A. 1988. *The Semantics of Grammar*. Amsterdam, 503–506.
- Zaliznjak, A. A. 1967. *Russkoe imennoe slovoizmenenie*. Moskva. [Reprinted in: Zaliznjak, A.A. 2002. Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju. Moskva, 1–370.]
- Zaliznjak, A. A. 1973. O ponimanii termina ‘padež’ v lingvističeskix opisanijax. In: Zaliznjak, A.A. (ed.): *Problemy grammatičeskogo modelirovanija*. Moskva, 53–87. [Reprinted in: Zaliznjak, A. A. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*. Moskva, 613–647.]
- Zwicky, A. M. 1986. German adjective agreement in GPSG. *Linguistics* 24, 957–990.