

# The penumbra of morphosyntactic feature systems<sup>1</sup>

Greville G. Corbett  
Surrey Morphology Group  
[g.corbett@surrey.ac.uk](mailto:g.corbett@surrey.ac.uk)

Draft of 10 July 2009: comments very welcome.

To appear in *Markedness in the Morphosemantics of  $\phi$ -Features* (special issue of *Morphology*), edited by Jonathan Bobaljik, Uli Sauerland & Andrew Nevins

## 1 Introduction

Features are often presented as clean, neat and simple. This is understandable, since it is the contrast with the widely varying behaviour of lexical entries which gives the intuitive justification for features. But real feature systems are more complex. For instance, linguists do not agree on how many case values Russian has, after years of study and argument. Yet Russian is no recent discovery, no obscurity pieced together from scraps of evidence. At least it is clear that Russian has a case feature. A whole feature may be in question: for the Daghestanian language Archi the issue is whether it has person at all. More generally, there are many examples where we need to ask whether we *really* have a proper feature, or feature value. We have to recognize that feature systems vary according to how well founded they are, and here the work of the Set-theoretical School is still of value. They also vary in how features and their values are distributed across the lexicon, sometimes in principled ways, sometimes almost randomly.

In order to analyse this murky area, the penumbra of feature systems, a canonical approach proves helpful (§2). Having justified this approach, we consider a set of converging criteria for canonical features and values (§3). This gives us a point in the theoretical space from which to calibrate the difficult instances. We ask whether the problems we find are feature-specific or whether they recur in the different morphosyntactic features (§4). This leads to asking whether they can co-exist (§5).

---

<sup>1</sup> This paper was first presented at the meeting on ‘Markedness and Underspecification in Morphology and Semantics’, Harvard, February-March 2008. I am grateful to the participants, and particularly to David Pesetsky, the discussant, for useful suggestions. A new version was given at the Max Planck Institute for Evolutionary Anthropology Leipzig, where there was also a good discussion: I thank Juliette Blevins, Bernard Comrie and Martin Haspelmath for the points they raised. I also wish to thank Matthew Baerman, Dunstan Brown, Marina Chumakina, Anna Kibort, Alison Long and Claire Turner for helpful comments on a draft. The paper has been improved following constructive suggestions by anonymous referees and Jonathan Bobaljik: my thanks to them also. The support of the AHRC (grant AH/D001579/1), the ESRC (grant RES-062-23-0696) and the ERC (grant ERC-2008-AdG-230268 *MORPHOLOGY*) is gratefully acknowledged.

Finally we check the instances we have found against the set of theoretical possibilities, and find that the picture is remarkably close to complete (§6).

## **2 The canonical approach**

Imagine we found a language in which every last noun had robust morphology distinguishing singular and plural, and every verb, adjective and adposition showed clear agreement in number. We would propose a morphosyntactic feature number, with the values singular and plural, without hesitation. Any alternative would make the syntax highly redundant. Conversely, if we found a language in which the only trace of number was a distinction between the equivalent of ‘I’ and ‘we’, then any case for a number feature would take some careful preparation. Of course, many languages fall between these two extremes. Yet we are perhaps too ready to treat them as though they were instances of the first type. Morphosyntactic features, like number, often have a ‘penumbra’ where the data are not clear-cut, and we need to be careful in our analysis.

I contend that each time a morphosyntactic feature is proposed in an analysis it needs to be argued for. If a parallel distinction proves useful in the semantic analysis, or the distinction is reflected in a purely morphological form, this is no more than suggestive. We need an argument specifically for proposing a morphosyntactic feature (and we may later wish to consider the degree of overlap between the different features). One part of the argument will be the range of the feature across the lexicon: a potential feature available to a few items is to be distinguished from a feature available to the majority of items in the lexicon (thus case in English is not on a par with case in Russian). It seems evident too that the analysis proposed should not depend on geography. Yet that is what we find: Guugu Yimidhirr has been analysed as having split ergativity (simple morphology, complex syntax), because it is spoken in Australia. If it were spoken in Siberia it would have been analysed as having a more complex case system, with simpler syntax. The paper will elucidate these issues, asking in particular whether they relate to individual features or apply more generally.

Those of us who work in typology tend to be attracted to clusterings of properties. And indeed, when we find features which are problematic, they are often problematic in more than one way. We need to see whether these are significant clusterings or accidental coincidences. One route towards that goal is to extend the theoretical space so that the clusters can be pulled apart: it turns out that the space is rather larger than is generally imagined. A way of anchoring this space is to take the type of instance we started with – the indisputable instance of a feature. We can use it to establish the properties of a canonical feature and its values, and then take it as the yardstick from which we measure the actual examples we find. And of course, the closer our example is to being canonical, the easier it is to justify the use of a morphosyntactic feature.

The aims of the paper are to elucidate the penumbras of features, to bring out the issues that require attention, to clarify some of the terminology on the way, and to establish whether the morphosyntactic features are different or similar when we look at their less clear manifestations (their penumbras). In some instances non-canonical behaviour has been given different terms for the different features, which perhaps leads us to expect that they are indeed different. This might be a useful pointer, or just a case of the tail wagging the dog.

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. This approach has been worked out particularly for inflectional morphology, as well as for syntax. Inflectional morphology has been treated by Baerman, Brown & Corbett (2005: 27-35), Spencer (2005), Stump (2005, 2006), Corbett (2007a), Nikolaeva & Spencer (2008), Stump & Finkel (2008) and Thornton (2008). In syntax, agreement has occupied centre stage, for instance in Corbett (2003, 2006), Comrie (2003), Evans (2003), Polinsky (2003), Seifart (2005: 156-74) and Suthar (2006: 178-98). A working bibliography of this growing body of research can be found at <http://www.surrey.ac.uk/LIS/SMG/CanonicalTypology/index.htm>. Most relevant to the current issue is the discussion of morphosyntactic features in Corbett (2008). There is a practical point to canonicity: since the examples nearest to canonical are those which are 'indisputable', when defining a canonical use of a term we should be able to assume it covers the canonical core; in the ideal scenario, differences in use of terms can be specified in terms of how far out from the canonical point different researchers allow particular terms to apply.

Thus canonical is not identical to prototypical (as normally used) since we have no requirement to produce a canonical exemplar; rather we need to be able to define and so identify the canonical point. We should also not confuse canonicity with being easy to find: the example which is frequently cited may not be a fully canonical instance of a phenomenon.

Why adopt this approach? We are tackling an area that is relatively new, where the detailed work has been mainly on individual features rather than taking the broad view. Hence, effort devoted to the basic intellectual housekeeping is worthwhile:

Главный признак точности в научном исследовании – не использование математического аппарата, а четкое, не допускающее различных толкований определение понятий. Гладкий (2007: 29)

Precision in research is characterized not primarily by the use of a mathematical apparatus but by the sharp definition of concepts, precluding different interpretations. Gladkij (2007: 29); translation GGC

Since we need to tackle the less clear cases in feature systems, not just the neat ones, a canonical approach will prove valuable in this tricky area.

### 3 Canonical features and feature values

In brief, a canonical morphosyntactic feature is one that has robust formal marking and is manipulated or constrained by simple rules of syntax. These two general principles cover sets of criteria, and weakenings of the criteria define a space for situating features that are ‘less good’, arguable or marginal, and for systematizing various earlier observations in the literature.

We now consider in more detail what a canonical morphosyntactic feature and its values would look like. We are focussing on the morphosyntactic features, strictly defined (Corbett forthcoming b): that is, features which have a role in both morphology and syntax. We shall start with case, since case has provoked interesting thinking on the topic, but we shall see that similar issues arise with the other features. We need to look first at how case values are identified, in order then to investigate how issues of canonicity relate to such case values. As a point of reference, we give traditional paradigms of two types of Russian noun:

(1) Paradigm of two Russian nouns

	<i>žurnal</i> ‘magazine’ (inflectional class I)		<i>komnata</i> ‘room’ (inflectional class II)	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOMINATIVE	žurnal	žurnaly	komnata	komnaty
ACCUSATIVE	žurnal	žurnaly	komnatu	komnaty
GENITIVE	žurnala	žurnalov	komnaty	komnat
DATIVE	žurnalu	žurnalam	komnate	komnatam
INSTRUMENTAL	žurnalom	žurnalami	komnatoj	komnatami
LOCATIVE	žurnale	žurnalax	komnate	komnatax

These are fully regular nouns: there are many thousands which inflect similarly.<sup>2</sup> However, these nouns represent only two inflectional classes; there are two other major classes and several smaller subclasses (Corbett & Fraser 1993). Moreover, there are some less secure case values, as we shall see in §5.

---

<sup>2</sup> Animates like *drug* ‘friend’ will be discussed below (see (18)); these inflect like *žurnal* ‘magazine’ but have the accusative syncretic with the genitive.

The case values proposed in (1) are relatively uncontroversial. But it is still worth asking, as Kolmogorov and the members of the Set-theoretical School did, how we justify such an analysis. For instance, given that *žurnal* has the same forms, singular and plural, for the case values nominative and accusative, how do we justify claiming there are two case values here? The method is set out in Zaliznjak (1967/2002: 36-42), but see also Goddard (1982) and Comrie (1986, 1991). We start from the idea of contexts. We take different syntactic contexts, such as *ja vižu ...* ‘I see ...’, *u menja net ...* ‘I haven’t got a ...’ and collate the forms which fit appropriately into these contexts (by tradition, the contexts are the rows, and the items examined – nouns in this instance – are arranged in columns, giving a table, as in (2)).

(2) Establishing features and values: an example from Russian

	Item1 <i>žurnal</i> ‘magazine’	Item2 <i>gazeta</i> ‘newspaper’	Item3	...	...	...
Context1 <i>Na stole ležit ...</i> ‘on table lies ...’	<i>žurnal</i>	<i>gazeta</i>				
Context2 <i>Ona думаet o ...</i> ‘she thinks about ...’	<i>žurnale</i>	<i>gazete</i>				
Context3 <i>Ona čitaet ...</i> ‘she reads ...’	<i>žurnal</i>	<i>gazetu</i>				
...						
...						
...						

If we had only the evidence of the first noun *žurnal* ‘magazine’, we would have to say that the contexts 1 and 3 provided no evidence for different values. However, when we put *gazeta* ‘newspaper’ in the same two contexts, this provides evidence for distinct feature values (the traditional nominative and accusative). Hence we could claim that the two different instances of *žurnal* have different case values. If two contexts produce exactly the same results for every item we test, then we can discard one of the two contexts.

So far this makes intuitive sense. However, the context must also be semantically constrained. Thus the context *ona pišet ...* ‘she is writing ...’ would allow both *pis’mo* ‘a letter’ and *karandašom* ‘with a pencil’. We do not want to suggest these two nouns are in the same case, rather that the apparently single context is not adequate here. In the terms of Comrie (1986: 91) we also require identity of function. And more generally, we may require alternative contexts in order to allow natural readings for different semantic classes of noun.

The procedure works well while we constrain the contexts (consciously or unconsciously). But suppose that like Zaliznjak we take the procedure seriously and include contexts like *ja risuju svoju* ... 'I am drawing my own ...' ? *Svoju* 'one's own' is feminine, and there is no possible form of *žurnal* 'magazine' that could fit into this context, since it is masculine. The next step in the procedure is to eliminate contexts like this one, which produce gaps in the table, provided that in all other respects (i.e. apart from the gap) the context gives results which are identical to those of another context. This has the neat effect of allowing us to abstract away from the features of number and gender while investigating case. For a fuller account see Zaliznjak (1967/2002: 36-42); the issues are well summarized in Blake (1994: 29-30), and the work of the Set-theoretical School is documented and evaluated by van Helden (1993). Quite often, the result of such an analysis is that the expected features and values are established, but that less clear instances emerge too, as we shall see.<sup>3</sup>

Given this general approach to determining feature values, let us now consider a canonical morphosyntactic feature and its values. There are several criteria, which we group under more general principles. Once the theoretical space is clear, we can identify and better understand some interesting deviations from the canonical ideal. The criteria for canonical features and values have been worked out previously (Corbett 2008). Going further, we find that those criteria give us a way for talking of various interesting phenomena described in the literature but not so far in this systematic way. As we tackle them it is worth keeping in mind that some of the phenomena have been represented by a couple of well-known instances, which confound more than one 'symptom'; we shall try to separate these out.

The first set of criteria, which is the set we shall concentrate on, is covered by this general principle:

*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).*

Formal means are 'clear' to the extent that they allow a transparent and regular mapping from form to function. The general point is straightforward: in the canonical situation there is clear evidence for the feature and its values. Principle I covers four more specific criteria.<sup>4</sup> We sketch these four criteria, to give the general picture, and then we consider each in greater detail below.

---

<sup>3</sup> It is worth remembering, as noted early, that features are not 'given' but have to be argued for and justified; see Pullum & Tiede (forthcoming) for clear demonstration of this point.

<sup>4</sup> By convention, principles are given Roman numerals, and criteria are labeled with Arabic numerals.

### 3.1 Criterion 1: Canonical features and their values have dedicated forms (are ‘autonomous’)

It is natural to assume that in order to postulate a feature, and its various values, we should be able to point to at least one inflected form and to show that it can be explained only in terms of the particular feature and value. For instance, in (1) above, the form *žurnalom* requires reference to case, and to the value instrumental: it cannot result from any other specification. This is what Zaliznjak (1973: 69-74) and Mel’čuk (1986: 66-70) treat as ‘autonomous’. There is a question, however, as to what the standard of comparison is. We might state the criterion in absolute terms, that is, we look for some marker with a unique function. This would be appropriate if the inflectional morphology in question was also canonical (Corbett 2007a, forthcoming a). However, Zaliznjak and Mel’čuk both treat autonomy relative to a particular lexeme; if for a given lexeme there is a unique form, then the feature value is autonomous.

The following schema represents an autonomous feature value.

SINGULAR PLURAL		function
a	d	NOMINATIVE
a	e	ACCUSATIVE
b	e	LOCATIVE
c	f	DATIVE




Fig. 1: Autonomous feature values

In Figure 1, the realizations ‘a’, ‘b’ and so on represent fully inflected forms, and that labelled ‘b’ uniquely marks the locative case value. Since it has one unique marker, the locative is ‘autonomous’. The nominative and dative are similarly autonomous. There is no form, however, which uniquely signals accusative case, which is therefore **non-autonomous**. Compare the situation in Figure 2:

SINGULAR	PLURAL	function
<i>azg</i>	<i>azgk'</i>	NOMINATIVE
<i>azg</i>	<i>azgs</i>	ACCUSATIVE
<i>azgi</i>	<i>azgs</i>	LOCATIVE
<i>azgi</i>	<i>azgac'</i>	DATIVE

Classical Armenian *azg* 'people'

Fig. 2: Non-autonomous case value

In this example (from Baerman 2002a) there is no unique form for the accusative; its forms are always syncretic. We still recognize an accusative case value, as is necessary to maintain simple rules of syntax. At the same time, we recognize that it is a less canonical feature value than the nominative or dative. We shall return to further instances of non-autonomous values once we complete our sketch of the criteria for canonicity.

### ***3.2 Criterion 2: Canonical features and their values are uniquely distinguished across other logically compatible features and their values***

In the canonical situation, we can distinguish a feature and its values irrespective of other features and their values. In other words, we do not have to select particular combinations: any of them will serve.<sup>5</sup>

Consider these examples from Italian, concentrating on the adjective. They show canonical expression of number and gender, since the values of each can be uniquely identified by the formal contrast, independently of the other:

---

<sup>5</sup> Zaliznjak (1973: 84-86) and Mel'čuk (1986: 61) talk of 'complete' versus 'partial' case values, where a complete case value 'embraces all nouns of a language throughout the whole paradigm', while a partial case value 'functions for a subset of the nouns or for a subset of the paradigms only' (Mel'čuk 1986: 61). I separate out the notions of embracing all relevant lexical items and that of applying throughout the paradigm in criteria 4 and 2 respectively.



Italian (Pier Marco Bertinetto, personal communication)<sup>6</sup>

- (3)    *il*                    *nuov-o*                *libr-o*  
        DEF.M.SG        new-M.SG            book(M)-SG  
        ‘the new book’
- (4)    *i*                        *nuov-i*                *libr-i*  
        DEF.M.PL        new-M.PL            book(M)-PL  
        ‘the new books’
- (5)    *la*                        *nuov-a*                *rivist-a*  
        DEF.F.SG        new-F.SG            magazine(F)-SG  
        ‘the new painting’
- (6)    *le*                        *nuov-e*                *rivist-e*  
        DEF.F.PL        new-F.PL            magazine(F)-PL  
        ‘the new magazines’

At this point we are interested in the evidence provided by the adjective for number and gender. This would be the natural analysis:

- (7)    Gender and number in Italian adjectival forms

GENDER	NUMBER	
	SINGULAR	PLURAL
MASCULINE	<i>nuovo</i>	<i>nuovi</i>
FEMININE	<i>nuova</i>	<i>nuove</i>

Each form in this paradigm is unique; each provides evidence of gender and number. The two sets of information are realized cumulatively, and so the morphology is not canonical. What matters here is that any of the forms is sufficient to provide evidence for a gender value and a number value. We have two number values and two gender values, each of which is autonomous. Contrast this with the cognate Russian adjective, in these examples (at this stage we consider a small part of the paradigm):

Russian

- (8)    *nov-yj*                    *žurnal*  
        new-M.SG.NOM        magazine(M)[SG.NOM]  
        ‘a new magazine’

<sup>6</sup> Glossing follows the Leipzig Glossing Rules. Inherent features are given in parentheses. Thus gender is glossed with the noun stem. While *-a* on the noun often implies feminine gender by the assignment rules of Italian, this is not necessarily so, as with *poet-a* ‘poet’ (masculine); similarly *-o* often implies masculine, but this is not always the case, as with *mano* ‘hand’ (feminine). These distinctions are important precisely because we are dealing with features.

*Penumbra of morphosyntactic feature systems*

- (9) nov-aja knig-a  
 new-F.SG.NOM book(F)-SG.NOM  
 ‘a new book’
- (10) nov-oe pis’ m-o  
 new-N.SG.NOM letter(N)-SG.NOM  
 ‘a new letter’
- (11) nov-ye žurnal-y  
 new-PL.NOM magazine(M)-PL.NOM  
 ‘new magazines’
- (12) nov-ye knig-i  
 new-PL.NOM book(F)-PL.NOM  
 ‘new books’
- (13) nov-ye pis’ m-a  
 new-PL.NOM letter(N)-PL.NOM  
 ‘new letters’
- (14) Gender and number in Russian adjectival forms (partial)

GENDER	NUMBER	
	SINGULAR	PLURAL
MASCULINE	novyj	novye
FEMININE	novaja	
NEUTER	novoe	

In this instance, in order to establish the need for a gender feature at all, and to establish the feature values, we need to look at forms which are singular. Russian clearly has gender, with three gender values, but its expression is less canonical than in Italian. Gender is not uniquely distinguished across number, as Criterion 2 requires; rather it is distinguished in one number value only.

Paying attention to the combination of features is a key part of the procedure for defining features and their values (see Zaliznjak 1973: 59 on case). The non-canonicity produced by various types of syncretism can produce difficult problems, as we shall see. Indeed, Meyer (1994: 360) suggests that successful set-theoretic modelling of a feature in a given language requires full knowledge of all the other features. Some put great store by the patterns of syncretism; the message of Baerman, Brown & Corbett’s (2005) survey of syncretism is that this is unwise: the specific patterns found in a few familiar examples are not representative of the syncretisms discovered in a larger sample of the world’s languages.

Deviations with respect to criterion 2 give **subvalues** (including structured values). At this stage we will restrict ourselves to one example of a subvalue, the animate

subgender of Russian. This requires us to extend our coverage of Russian adjectives to include case:

(15) Paradigm of the long form adjective *novyj* ‘new’ (traditional presentation)

	MASCULINE	FEMININE	NEUTER	PLURAL
NOMINATIVE	novyj	novaja	novoe	novye
ACCUSATIVE	as NOM / GEN	novuju	novoe	as NOM / GEN
GENITIVE	novogo	noj	novogo	novyx
DATIVE	novomu	noj	novomu	novym
INSTRUMENTAL	novym	noj(u)	novym	novymi
LOCATIVE	novom	noj	novom	novyx

In (15) the adjectival paradigm is given in traditional layout (a truly traditional presentation has the accusative ordered after the dative, but this obscures some of the regularity of the paradigm). There are many more cells than there are distinct phonological forms, owing to pervasive syncretisms. These syncretisms are brought out in (16) by reordering the gender and case values:

(16) The paradigm of the long form adjective *novyj* ‘new’ (showing syncretisms)

	FEMININE	NEUTER	MASCULINE	PLURAL
NOMINATIVE	novaja	novoe	novyj	novye
ACCUSATIVE	novuju		as NOM / GEN	
GENITIVE	noj	novogo		novyx
LOCATIVE		novom		
DATIVE		novomu		novym
INSTRUMENTAL	noj(u)	novym		novymi

A first thing to notice is that looking at the case values we have added in here (as compared with (14)) does not give any further differentiation for the three genders in the plural. Second, the masculine and neuter are less well differentiated than the feminine; they are identical in the oblique cases and so are less canonical in this respect. Thus *noj* is uniquely feminine while *novogo* could be masculine or neuter. (From the point of view of case, however, the feminine does not distinguish the oblique cases, and the adjectival instrumental inflection *-ju* is now largely limited to poetry.) However, the most interesting of these syncretisms in the adjectival paradigm concerns the accusative case. If we concentrate on the masculine singular (all the plurals behave in the same way too), then the form of the accusative is identical to the nominative for inanimates, as in (17), and identical to the genitive for animates, as in (18) and (19):

Russian

- (17) ja viž-u star-yj dom  
 I see-1SG old-M.INAN.SG.ACC house(M.INAN)[SG.ACC]  
 ‘I see an old house’

(18) ja viž-u star-ogo drug-a  
I see-1SG old-M.ANIM.SG.ACC friend(M.ANIM)-SG.ACC  
'I see an old friend'

(19) ja viž-u star-ogo dedušk-u  
I see-1SG old-M.ANIM.SG.ACC grandfather(M.ANIM)-SG.ACC  
'I see (my) old grandfather'

For a formal account of this syncretism, which goes over paradigm boundaries, see Corbett & Fraser (1993), and for the typological implications see Baerman, Brown & Corbett (2005: 206-217). The key point is that we cannot claim that the animate form in (18) is simply a genitive, as we see if we look carefully at (19). The form of the noun in (19) is uniquely accusative (distinct from both nominative and genitive), and the form of agreeing adjective must also be accusative. Since the adjective carries the specification 'accusative, masculine, and animate', and it has no unique form, its form is identical to the genitive.

Animacy in Russian is a good illustration of why we need to separate out the different criteria. On the one hand, the animacy distinction is severely limited in that it is found within just one case (and it is non-canonical in this respect). On the other hand, it is a central part of the system, affecting nouns, pronouns, almost all adjectives (those that can occur in attributive function) and some numerals.

We will consider further deviations in §4 below. First we should complete our initial pass through the criteria.

### ***3.3 Criterion 3: Canonical features and their values are distinguished consistently across relevant parts of speech (word classes)***

In a sense it is true to say that German and English both have case, and indeed that German and English both have gender. However, these statements are somewhat misleading. In German we have evidence for the case feature in articles, adjectives and pronouns (as well as limited evidence in nouns). For gender we have evidence in articles, adjectives and pronouns. In English the evidence for case and gender is restricted to pronouns.<sup>7</sup> Criterion 3 draws the distinction between the two systems, German being largely canonical here and English clearly not.

---

<sup>7</sup> For those who do not accept pronominal gender systems, as discussed in Corbett (1991: 169-170), the gender example is not relevant here, but the argument from case is clear. Numerous further examples of case systems where the inventories of different parts of speech differ can be found in Iggesen (2005).

A problem with criterion 3 gives a **limited** system. In the extreme instance, where there is a distinction in only one part of speech, we have for instance a pronominal gender system or a pronominal case system. More interesting problems arise if the realization of the values is not *consistent*. We return to that issue in §4 below.

### 3.4 Criterion 4: *Canonical features and their values are distinguished consistently across lexemes within relevant parts of speech*

In the canonical situation, given the morphosyntactic specification determined by the syntax and the part of speech (word class) of the target, no more is required. Each member of the part of speech marks the feature and all its values consistently. When more information *is* required, that is, when not all members of the part of speech behave consistently, deviations from the canonical situation may be seen in two different ways:

**In terms of level**, the deviation may be in terms of the *feature* as a whole<sup>8</sup> or only in terms of its *values*. In Macedonian, while most adjectives mark gender and number, some mark number but do not mark gender (Friedman 1993: 266-267). That is, they do not mark the *feature* gender; it is not simply that they fail to distinguish particular values.

- (20) Macedonian adjectives (Victor Friedman 1993: 266-267 and personal communication)

MASCULINE	FEMININE	NEUTER	PLURAL	gloss
nov	nova	ново	novi	new
kasmetlija			kasmetlii	lucky
taze				fresh

Typical native adjectives, like *nov* ‘new’, distinguish three genders and two numbers. Adjectives like *kasmetlija* ‘lucky’ agree in number but not in gender. On the other hand, *taze* ‘fresh’ and adjectives like it are indeclinable, they are unable to agree (see Baerman, Brown & Corbett 2005: 30-33). This means that we cannot necessarily say that in a given language, adjectives (or whatever part of speech we have in mind) agree in particular features. There may be variation within the part of speech, so that we need information about particular lexical items.

This is an example of particular lexical items failing to mark a *feature* which the other members of the part of speech do mark. At the level of *values* consider these Latin adjectives:

<sup>8</sup> This criterion takes up ‘lexical generality’, as in Bybee (1985: 84-86), and goes further in clearly distinguishing its application to features and to their values.

(21) Latin adjectives of inflectional class 3 (singular)

MASCULINE	FEMININE	NEUTER	gloss
acer	acris	acre	sharp
facilis	facilis	facile	easy
vigil	vigil	vigil	alert

Latin has three gender values, as shown by many adjectives (though not in all of the cases). Adjectives belonging to the third inflectional class show various possibilities. *Acer* ‘sharp’, and others like it distinguish three gender values; adjectives like *facilis* ‘easy’, on the other hand, show no distinction between masculine and feminine. *Vigil* ‘alert’ and similar adjectives show no evidence of gender agreement in the nominative singular (though certain other forms distinguish neuter from the other genders).

**In terms of range**, the deviation may affect different numbers of lexemes. The canonical situation is that each lexeme marks the feature and its values. Deviations may involve larger or smaller subclasses. We may find inflectional classes which distinguish the particular feature and its values to varying degrees (as in point 1). There may be subclasses at various levels, right down to individual lexical exceptions. The latter may be overdifferentiated (marking ‘too many’ distinctions compared with the other members of their subclass) or they may show additional syncretism and so show too few distinctions. For such situations, approaches like that of Network Morphology (as in Corbett & Fraser 1993 and Evans, Brown & Corbett 2002), which rely on default inheritance, are helpful, since they readily capture generalizations which apply to large classes and are overridden by specifications which are more and more restricted, right down to the idiosyncrasies of individual words. Moreover, this sort of deviation may be motivated to varying degrees. We may find that a split is principled, as when a distinction is available for, say, all animates. At the other extreme a distinction may be (synchronically) unprincipled, so that each lexical item affected requires a special indication.

The different types of non-canonicity of lexemes (such as marking too few or too many distinctions) fall under a principle that we shall not discuss further here, namely that canonical morphosyntactic features and their values are realized through canonical inflectional morphology (Corbett forthcoming a). That principle is concerned with canonicity from the point of view of the lexeme. At this point we are taking the perspective of the feature and its values, suggesting that a canonical feature will not be subject to the restrictions we are considering.

The essential point of the fourth criterion, consistency across lexemes within relevant parts of speech, is that in the canonical situation it is sufficient to have a syntactic rule (of the type: in Polish ‘the preposition *ku* ‘towards’ requires the dative’) and the part of speech of the target (e.g. noun). Any requirement for additional information about the particular lexeme(s) making up the target is non-canonical.

The set-theoretical approach highlights the issues. Subsequent analysis is likely to involve a trade-off between regularities stated high in an inheritance tree and statements that belong lower down either as sub-regularities, or even in particular lexical entries. Suppose that the set-theoretical approach to a particular language demonstrates the need for both dative and instrumental case values; however, some nouns lack distinct forms. If such nouns are in a minority, we would probably favour an analysis in which the distinction is stated high up, and the particular items showing syncretism are treated as an override to this general statement. However, if the nouns making the distinction are in the minority, we might propose a high-level generalization that the two forms are identical, with a lower level statement of distinct forms. In each instance, we need to make a special statement for a larger or smaller group of lexical items (this is taken up in the discussion of (29) and (30) below).

In the examples of inconsistency across parts of speech discussed so far, the status of the feature or feature value was not in doubt, since the problems involved relatively few lexemes. If the problems affect a large number of lexemes, so as to threaten the status of the feature or feature values, we come to the issue of **minor** features and values, which we shall discuss below.

So far, then, we have surveyed Principle I: which concerns clear formal marking; we have considered briefly the four criteria it covers. We should also have in mind Principle II, which requires that morphosyntactic features be manipulated or constrained by simple rules of syntax. This is relevant to the current discussion since certain non-canonical aspects of features could be avoided by baroque rules of syntax, if these were allowed. For instance, the case problem in Classical Armenian (Figures 1 and 2) could be analysed away by claiming that transitive verbs take the nominative with singular direct objects and the locative with plural direct objects. Such a rule, which I assume we should avoid, is clearly more complex than the rule stating that transitive verbs take the accusative case of their object. It is this simpler rule which then leads us to propose a non-autonomous case value. A general requirement of simple syntax is that it should be ‘morphology-free’ (Zwicky 1996: 301, Corbett & Baerman 2006). Thus Principle II prevents us from shipping out our difficult examples into the syntax. And third, as has already been mentioned, another perspective on canonical features and their values, which we shall not take further here, is that they be realized by canonical inflectional morphology (see Corbett 2007b for more on canonical inflection).<sup>9</sup>

## 4 Recurring types of non-canonical feature values

Let us now go back and consider deviations from the four criteria (under Principle I) in more detail, asking in particular how they affect different features and their values.

---

<sup>9</sup> A very observant reader may wonder about fully canonical morphosyntactic features and how we could distinguish them. That is a good question, one whose answer is too large to fit here; an account is given in Corbett (forthcoming c).

Recall that Principle I is concerned with the morphological realization of the features. There are two rather different types of prediction we might have. We could look at the evident differences in the semantics and syntax of the different features and assume that these will be reflected in comparable differences in their morphology (specifically in the possibilities for the non-canonical behaviour we are investigating). Alternatively we might have in mind the arguments that morphology is autonomous, and predict that the realization of the different featural requirements specified by the syntax will be similar across the features. In an intriguing fashion, as we shall see, issues of non-canonicity with feature values apply to the different features in comparable ways, giving a picture closer to the second prediction.

At this stage we are concerned with the existence or not of all the types of non-canonical behaviour which we have defined. To this end it makes sense to look wherever the chances are best: many languages were investigated, and no language was excluded because of any sampling considerations.

**4.1 Criterion 1: Canonical features and their values have dedicated forms (are ‘autonomous’)**

The problematic example we considered earlier (Figure 2) involved case. If we look more abstractly at the relevant part of the paradigm, we have this schema (Figure 3).

forms		functions
a	c	I
a	d	II
b	d	III

Fig. 3: Non-autonomous feature value

Given this abstract schema, we can see an important aspect of the analysis: it can apply equally to other features, since similar deviations from canonicity can be found with them. Consider these data from Romanian:

Romanian (Anca Sevcenco, personal communication)

(22) student      înalt student(M)[SG] tall[M.SG] ‘a tall (male) student’	(25) student-i      înalt-i student(M)-PL tall-M.PL ‘tall students’
---	---

(23) scaun      înalt chair(N)[SG] tall[M.SG] ‘a tall chair’	(26) scaun-e      înalt-e chair(N)-PL tall-F.PL ‘tall chairs’
--	---



- (24) student-ă      înalt-ă      (27) student-e      înalt-e  
 student(F)-SG tall-F.SG      student(F)-PL tall-F.PL  
 ‘a tall (female) student’      ‘tall (female) students’

If we had just the evidence in (22), (23) and (24), we would propose two agreement classes, and hence two gender values. There was a long debate about whether there are indeed two genders or three (see references in Corbett 2007c: 245-249). On the one hand, nouns like *scaun* have no agreement forms which are used uniquely for them; on the other, they do have a unique set of agreement requirements, as we see when we include the evidence of (25)-(27). In terms of agreement classes, there are clearly three, as in Figure 4.

SINGULAR PLURAL		
înalt	înalt-i	MASCULINE
înalt	înalt-e	NEUTER
înalt-ă	înalt-e	FEMININE

Romanian *înalt* ‘tall’

Fig. 4: Non-autonomous gender value

The feature has been changed here – we are now dealing with *gender* – but the pattern is the one we have just seen (in Figure 3). In the singular in Romanian there is no evidence for a third, neuter gender, nor in the plural; it is the combination which gives the third value. We recognize it as a gender value, but one that is less canonical than the other two. A gender value of this type is sometimes known as ‘genus alternans’ (Igartua 2006); see this source and references there for discussion of the development of such instances in Indo-European. The important point for us is that this situation is fully analogous with that illustrated for case above. Therefore we do not need the term ‘genus alternans’. Similarly, Corbett (1991: 164-165) introduced the term ‘dependent target gender’ for gender values realized only through syncretic forms; this term too can be replaced by ‘non-autonomous gender value’, thus making the terminology consistent across the different features. While saying that Romanian has three gender values, we want to distinguish it from languages like German, Russian or Tamil. It is typologically different, in various ways, including the fact that one of the values is non-autonomous.

A clear instance of a non-autonomous *person* value is found in Old Nubian (a Nilo-Saharan language, with texts dating from the eighth to the fifteenth centuries, according to Browne 2002: 1). Here are the key data (from Browne 2002: 50, see also Bechhaus-Gerst 1996: 237, cited in Baerman, Brown & Corbett 2005: 75):

SINGULAR	PLURAL	
dollire	dolliro	1 PERSON
dollina	dolliro	2 PERSON
dollina	dollirana	3 PERSON

Old Nubian present indicative (*doll-* ‘wish’)

Fig. 5: Non-autonomous person value

Old Nubian is cited here because it had regular syncretism, as in Figure 5, through its verbal paradigms. In Nobiin, its descendant, second person forms were innovated in some parts of the verbal paradigm, so that the second person became autonomous.<sup>10</sup>

Thus in general, problems with criterion 1 give **non-autonomous** values. We shall see a further example for case when we consider the Russian second genitive below, though the effect there is less striking. A more dramatic example involving person is found in Archi. It may seem obvious that we can argue for a feature only if at least two values are autonomous (so there is a distinction in form). However, even that is not clear-cut; see Chumakina, Kibort & Corbett (2007) where it is argued that a person feature is required in the grammar of Archi, even though there is no dedicated form to support it.

#### 4.2 *Criterion 2: Canonical features and their values are uniquely distinguished across other logically compatible features and their values*

Problems with criterion 2 give **sub**-values. For example, see this definition for subgender (Corbett 1991: 163)

Subgenders are agreement classes which control minimally different sets of agreements, that is, agreements differing for at most a small proportion of the morphosyntactic forms of any of the agreement targets.

---

<sup>10</sup> Other instances of non-autonomous person found to date are typically less systematic, varying according to verbal categories such as tense, aspect and mood. Thus in the Omotic language Benchnon, the second person is non-autonomous in the indicative final and in the medial forms, but has unique forms in polar question forms (Rapold 2007: 70-71). In Amahuaca (Panoan) the second person is again non-autonomous, with the syncretisms varying according to tense (Baerman 2002b, citing Sparing-Chávez 1998: 449). And Marc-Olivier Hinzelin points out (personal communication) that certain Francoprovençal varieties show relevant patterns, but only in particular tense-mood combinations: see Fankhauser (1911: 151-153) for examples.

There is a detailed account of Serbian/Croatian/Bosnian (Corbett 1991: 161-165) to show what is covered by ‘minimally different’. The intuition behind the definition is that in larger feature systems there can be distinctions which are minimal, when compared to those of other feature values, and these are to be specifically noted.<sup>11</sup>

Sub-values often have other non-canonical behaviours. For instance, they may be non-autonomous too (Corbett 1991: 164, where the term used was ‘dependent’), but this is not required (Brown 1998b: 198); these two types of non-canonical behaviour may co-occur but are not bound to do so. Sub-values may also not be distributed across other relevant parts of speech. If we are to extend the notion to other morphosyntactic features, it is the notion of *minimal difference* that is relevant. Thus Brown (2007) treats the Russian second locative as a *subcase*. There are also possible instances of a *subnumber*, where the dual is restricted to the first person. However, such instances are open to other interpretations. A clearer instance is found in Biak (an Austronesian language, of the South Halmahera-West New Guinea subgroup, spoken on the islands of Biak and Numfor and on several small islands near the Bird's Head, and along part of the north coast of the Bird's Head), described by Steinhauer (1985) and discussed also in Cysouw (2003: 201-202, including footnote 12). We follow the most recent source, van den Heuvel (2006):

(28) The independent personal pronouns of Biak (van den Heuvel 2006: 66)

PERSON GENDER		NUMBER			
		SINGULAR	DUAL	PAUCAL	PLURAL
1	EXCLUSIVE	aya	nu	inko	
	INCLUSIVE	—	ku	ko	
2		aw	mu	mko	
3	ANIMATE	i	su	sko	si
	INANIMATE				na

As can be seen, Biak has a paucal number, but in the third person only. These features and values are not restricted to the independent pronoun (as in (28)); their use extends to the pronominal affixes and pronominal articles. We may view this as an instance of a sub-number.

The instances presented so far have all involved a restriction according to a *different* feature/value (for instance, a number value restricted by person, as in Biak). Now we

<sup>11</sup> We need the notion of comparison within the paradigm: we do not want, for instance, to say that because they are distinguished in one number only the Russian genders are subgenders (see examples (8)-(13) above); particularly since the syncretism fits into a widespread typological pattern (Greenberg 1963).

turn to instances where the restriction is to a limited number of values – usually one – within the same feature. Such instances are sometimes treated as **structured** values.

One such treatment is Brown's (2007) account of the Russian second locative. If we look back at the basic system given in (1), we need to add that there is an additional value, the second locative, which is non-canonical in more than one respect. What concerns us here is that it is distinct for a small proportion of the available paradigm cells, namely within the locative singular only. That is, there is no distinction in the plural, and where there is a distinction it is within the locative only. There is a handful of prepositions which take the locative, and just two of them (*v* 'in' and *na* 'on') require the second locative, as in *v sadú* (second locative) 'in the garden', in contrast with *o sade* (ordinary locative) 'concerning the garden'. When it is available (and it is also non-canonical in that few items have a distinct form, see §4.4), the second locative is thus a subdivision of the locative.

This sort of non-canonicity has a satisfying account. Brown's solution is to treat the second locative as a 'structured' case (1998a: 198-200). In his Network Morphology analysis, which relies heavily on the notion of defaults, the default realization of the second locative will be as the normal locative. This default is overridden for the relative few nouns with a distinct second locative. Inflectional morphology in this model is specified according to ordered paths, where more specific information is ordered after more general. For instance the specification of the ordinary locative singular may be given as:<sup>12</sup>

(29) <singular locative>

This ordering is justified by the fact that some nouns have different stems for singular and plural (and these are specified within the lexical entry), while none have a special stem for locative. Then the specification of the second locative is an extension of this path:

(30) <singular locative locative2>

The effect is that when the second locative is in principle required by the preposition, then: (a) if the noun has a form matching the specification (30) completely, a distinct second locative, this is the appropriate form; (b) if not, by default, the closest matching path specifies the form, and that path is as in (29), giving a normal locative. This latter will always occur with adjectives, which Brown covers elegantly with no extra machinery. The analysis in effect says that all nouns have an appropriate realization for second locative (if the syntax requires it, there is an appropriate form, usually the ordinary locative), but very few nouns *really* do (that is, few nouns have a distinct morphological form for the second locative). As far as the main issue here is

---

<sup>12</sup> Brown (1998a, 2007) uses 'prepositional' and 'locative' rather than 'locative' and 'locative2'; I have retained locative and locative2 for consistency.

concerned, this approach clearly treats the second locative as a subcase of the locative, as the ordered path in (30) shows.

Two important points need to be made here. First, Brown's solution for this particular type of non-canonicity is not just an elegant idea. It is worked out as a full implementation within Network Morphology (one of the inferential-realizational approaches to inflectional morphology). Brown's implementation gives the right outcomes and others can test it to ensure that it is indeed a valid analysis (the fragment is provided in Brown 1998a). And second, it develops an idea found in Comrie (1991: 102), that of having a hierarchical feature analysis for the second case,<sup>13</sup> one which captures the specific nature of these case values and does not involve invoking other artificial features for other case values.

We may think of clusivity in similar terms. On one analysis the distinction is not logically compatible with the singular; otherwise it can be found across all values of other features, and yet it is restricted to occurring within *person*. See, for instance, the system of Belhare (a language of the Kiranti group within Sino-Tibetan, spoken in Nepal) as Bickel & Nichols present it (2005: 51):

(31) Belhare (Bickel & Nichols 2005: 51): intransitive verb forms

		SINGULAR	DUAL	PLURAL
1	EXCLUSIVE	-ŋa	-chi-ŋa	-i-ŋa
	INCLUSIVE	—	-chi	-i
2		-ga	-chi-ga	-i-ga
3		Ø-	N-...-chi	N-

Here we see a clear marker of the exclusive *-ŋa*, while inclusive remains without a marker; clusivity is restricted to the first person. Note, however, that not all would agree with this analysis of clusivity; see for instance Daniel (2005) and the review by Bobaljik (2008).

Another example is provided by Sursurunga, which is convincing because of the size of the number system. The data are from Hutchisson (1986, and personal communications). Sursurunga has some 4 000 speakers in southern New Ireland. It is one of the South New Ireland/West Solomonic languages, which form part of the New Ireland Network, that being a branch of Melanesian, within Oceanic, in turn part of Austronesian (Ross 1988: 258). The meanings of the values are discussed in Corbett (2000: 26-29):

<sup>13</sup> Brown discusses the locative and Comrie the genitive, but in both instances the account generalizes to the other second case.

(32) The emphatic pronouns in Sursurunga

	SINGULAR	DUAL	PAUCAL	GREATER PAUCAL	PLURAL
1 EXCLUSIVE	iau	giur	gimtul	gimhat	gim
1 INCLUSIVE	—	gitar	gittul	githat	git
2	iáu <sup>14</sup>	gaur	gamtul	gamhat	gam
3	-i/on/ái	diar	ditul	dihat	di

Here again we see clusivity running through the first person, in this instance through four values of number. For those who treat clusivity differently, the sub-values of person are rather to be looked for in the various proposed ‘fourth persons’, which may be analysed as sub-values of the third person.

To sum up, problems with criterion 2 lead to sub-values, restricted across other feature values or within their own feature. Given the number of values available, sub-values are prevalent when case is involved, since this gives more possibilities for discrepancies between proportions of the paradigm in which different feature values are realized distinctly. The clearest instances are those where there is a clear contrast between a distinction drawn right through a paradigm and another where there is only a minimal difference. Of course, there are many instances between the two extremes. In the canonical situation, the value is uniquely distinguished across all other logically compatible features and their values; in the least canonical situation, the value is distinguished minimally, and we recognize increasing non-canonical possibilities between the two extremes.

**4.3 Criterion 3: Canonical features and their values are distinguished consistently across relevant parts of speech**

As we saw earlier, systems which are non-canonical because of a problem with criterion 3, where only one part of speech has a particular distinction, give **limited** systems. We now look at more interesting examples where the distinctions are not consistent across parts of speech. (We mean that the distinctions available are different, not that ‘inconsistent’ choices can be made in particular instances, as with hybrid nouns, for example.)

In the easy instances, one part of speech has values which represent a collapsing of the values available to the other. In the least problematic instances, the distribution is principled, for instance, it is according to the Animacy Hierarchy. Thus in the number feature in Yimas, pronouns have singular/dual/paucal/plural, but nouns lack the

---

<sup>14</sup> á is used to indicate schwa; this is the preferred form according to Hutchisson (personal communication), rather than 'a, as in Hutchisson (1986: 20fn7). Other changes from the 1986 paper, like *-hat* for *-at* in the greater paucal, are based on personal communications.

paucal (William Foley 1986: 74, 86-87, 132-133; 1991 216-225, and personal communications, reported in Corbett 2000: 92, 120-121). Since personal pronouns come higher on the Animacy Hierarchy, Yimas conforms to a general pattern Yimas conforms to a general pattern, namely that there is greater differentiation higher on the hierarchy.

We also find collapsing of values in a less principled way; the Russian second locative discussed above is a distinction available to the noun but not to the adjective (where the two locatives are marked identically). There are much more challenging instances, such as Bayso (Cushitic) where nouns and verbs both mark nominal number, but where the mapping of values between the two is anything but straightforward (Corbett 2000: 127-129, 181-183).

Particularly interesting are the systems in which the combination of possibilities on different parts of speech gives rise to additional values. We begin with *gender*, where such systems have been called **combined** gender systems (Corbett 1991: 184). Few examples have been found; consider these data from Mba. Languages of the Mba group (Ubangian branch of Niger-Kordofanian; data from Tucker & Bryan 1966: 110, 114-23, 131-40; Pasch 1985: 69-71; 1986) have a combination of gender systems similar to the Bantu type (somewhat reduced), and a second system distinguishing up to four members: male human, female human, animal and inanimate. Most interesting of the group is Mba itself, discussed in detail in Corbett (1991: 185-188). The key data are presented in (33); the situation is somewhat more complex but the essential points are evident here.

(33) Consistent agreement patterns in Mba

attributive agreement		pronoun / optional agreement	gender	combined gender
singular agreement	plural agreement			
w	y	ndé	I	1/2 male personal
w	y	ɓi'	II	1/2 animate
w	y	∅	III	1/2 inanimate
l	s	∅	IV	3/4 inanimate
k	z	ɓi'	V	5/6 animate
k	z	∅	VI	5/6 inanimate
g	y	ndé	VII	7/2 male personal
g	y	ɓi'	VIII	7/2 animate
g	y	∅	IX	7/2 inanimate
ny	z	∅	X	9/6 inanimate
m	y	∅	XI	11/2 inanimate

The first two columns give the attributive agreement markers (the Bantu type), from which we set up agreement classes 1/2, 3/4, 5/6, 7/2, 9/6, 11/2. Note that not all the nouns in these proposed agreement classes behave identically, since there is the

further set of distinctions found in the pronoun (which also serves optionally as an agreement marker), shown in column three. Now the fact that different agreement targets behave differently is not in itself so unusual. Normally, however, when there is a difference, the distinctions made on one target are simply a subset of those made on the other (and there can therefore be a simple rule of agreement); this would be the ‘collapsing’ noted above. Mba is different, since the attributive modifier and the pronoun/optional agreement marker seem to be operating according to different systems.

At this point there are two ways forward. One is the way consistent with the approach of the Set-theoretical School. There are two contexts (attributive agreement and pronoun / optional agreement), and only those nouns which control the same set of agreements in all contexts have the same gender value. There are eleven possibilities, hence eleven gender values, as given in column four. (The combinations on which they are based are given in column five.) The alternative is to say that the items in column five are not just useful mnemonics for the eleven genders, rather they reflect the existence of two co-existing systems.

There are three considerations pointing towards the first alternative (combining the gender values as a single system, as in column four). First, we do not get all the possibilities from the two types of agreement; instead of eighteen theoretical possibilities we actually find eleven. Second, cross-linguistically we find one set of values for each morphosyntactic feature; relaxing that constraint would allow for a wide range of possibilities which appear not to occur. And third, Principle II requires morphosyntactic features to be manipulated or constrained by simple rules of syntax; if we require two different agreement rules, both to handle gender and number agreement, this is not simple syntax.

Whichever way we analyse the Mba data, by complicating the gender specification of nouns or by complicating the rules of agreement, it is clear that we do not have a canonical morphosyntactic gender feature, since its values are not distributed consistently across the relevant parts of speech. We should then ask whether there is anything similar to be found in the other morphosyntactic features. Consider these data on *number* from Mele-Fila (Corbett 2000: 35, 69). Mele-Fila is an Eastern Oceanic language spoken on Vanuatu. The data are from Ross Clark (personal communications). The article makes a three-way number distinction which, taken on its own, would give a singular-paucal-plural system. The forms for the noun *nuaane* ‘old man’ are given in (34). The underlying form of the plural article is /a/, but before nouns of more than two morae, the form is zero.



## (34) Number contrasts in Mele-Fila (data from Ross Clark)

article distinctions	SINGULAR	PAUCAL		PLURAL	
article plus noun (‘old man’)	t-nuaane	ru nuaane		nuaane	
pronoun distinctions	SINGULAR	DUAL	PLURAL		GREATER PLURAL
pronoun	aia	raaua	raateu		reafa
‘constructed’ number	SINGULAR	DUAL	PAUCAL	PLURAL	GREATER PLURAL

The pronoun makes four distinctions rather than three, and these do not map simply onto those of the article. The pronoun distinguishes singular and dual forms, and then the remaining space is divided between a plural and a greater (global) plural. The dual pronoun is appropriate only for some cases where the paucal article would be used. On the other hand, the pronoun *raateu* covers the remaining area of the paucal article, but splits the range of the plural article, the part left over being covered by *reafa*. If we follow the set-theoretical approach, we combine the two systems and we have five number distinctions, as shown in the last line of the table. The alternative is to have more complex syntactic rules. In either case the system is non-canonical, in a way analogous to the issue in the gender system of Mba. Other systems where different elements combine to give the full range of number distinctions include Hopi and Zuni (Corbett 2000: 169-171), see also Bliss (2005) on Hopi. Note however that sometimes only a part of the system works this way; thus in Hopi only the pronouns are affected in combination with the verb, while the noun has full marking. Systems like that of Mele-Fila were termed ‘constructed number’ systems in Corbett (2000). Once their affinity to the combined gender systems is recognized they could better be termed ‘combined number systems’.

Let us move on to *case*. Consider the Pama-Nyungan language Guugu Yimidhirr. The data are from Haviland (1979: 47–51, 66–67), discussed in Baerman, Brown & Corbett (2005: 42-45); the table has been reoriented to show the analogy with those for gender and number above.

## (35) Guugu Yimidhirr

	ERG	NOM-ABS	ACC
PRONOUN 1SG	ngayu	ngayu	nganhi
NOUN ‘girl’	gabiirngun	gabiir	gabiir

Given just the pronoun, we would think that Guugu Yimidhirr had a nominative-accusative system. With just the evidence of nouns we would say that it was ergative-absolutive. We can consider the two together, in a way that recalls the non-autonomous values discussed earlier, though with the major difference that they involved forms of a single paradigm, while here we are considering different lexical items of different part of speech. The consequence is that these data give evidence for

three distinct case values (and Guugu Yimidhirr has several additional simpler cases). This makes good sense, and might be called a ‘combined case system’. Such an analysis is argued for in Goddard (1982).

Such systems are often termed systems of ‘split ergativity’; the focus is then on which elements work according to an ergative-absolutive system (those lower on the Animacy Hierarchy) and which follow a nominative-accusative system (those higher on the Hierarchy). That analysis rests on the assumption that different parts of speech can vary in ways which make the syntax-morphology interface anything but simple.<sup>15</sup> Our Principle II, requiring morphosyntactic features be manipulated or constrained by simple rules of syntax, prevents us from concealing non-canonical behaviour by assuming complex syntax. By accepting the notion of combined case systems, analogous to combined gender and combined number systems, we avoid such complexity.

We now have analogous non-canonical behaviour for three morphosyntactic features, in respect of their not being consistent across the parts of speech. We should then ask if there is anything similar in *person*. Consider these data from Maybrat, (a West Papuan language spoken in the central area of the Bird’s Head, Irian Jaya), pointed out to me by Matthew Baerman (also discussed in Cysouw 2005: 84):

- (36) Maybrat (Dol 2007: 65)  
anu p-kias ania  
2PL 1PL-tell RECIP  
‘We (inclusive) tell each other.’

The inclusive is expressed by a combination of the second plural free pronoun and the first person marker on the verb (Dol 2007: 64). We thus have a combined system within person. Daniel Harbour points out (personal communication) that Kiowa has an analogous combination, though here it a first person free pronoun (which does not distinguish number) combined with a second person plural verb form. These combine to give a first person inclusive (Harbour 2007: 82-83).

The deviations concerning the third criterion have shown particularly clearly how the different morphosyntactic features can be analyzed in parallel with each other.<sup>16</sup>

---

<sup>15</sup> For the complex situation in Anatolian languages, see Patri (2007).

<sup>16</sup> Data like those from Guugu Yimidhirr recall the examples cited for non-autonomous values. In those instances where the combined system represents the complete picture (recall that there are instances, like Hopi, whose combined number system applies only for some items), it is the case that at least one value is non-autonomous. The point is that establishing the value in a combined system requires reference to different parts of speech, and not just to other different feature values.

**4.4 Criterion 4: Canonical features and their values are distinguished consistently across lexemes within relevant parts of speech**

The basic deviation gives us a **minor** value: the key point is that the limitation is on the number of items involved, rather than being a featural one. There are several types of deviation here. Let us start from instances of *number*.

In Maltese most nouns distinguish singular from plural. Now consider *uqija* ‘ounce’:

(37) Example of the Maltese dual

SINGULAR	DUAL	PLURAL
uqija	uqitejn	uqijiet

This noun has a dual in addition; there are around 30 nouns which have the dual. This is therefore a minor number value (Corbett 2000: 96). To complete the picture, only eight of these nouns, according to Fenech (1996), require the use of the dual (*uqija* ‘ounce’ is not one of them: for ‘two ounces’ one can use either the dual *uqitejn* or the form with the numeral: *żewġ uqijiet*). In terms of canonical morphology, such nouns are ‘overdifferentiated’ (Bloomfield 1933: 223-224); they have a distinct form for the dual in their paradigm while almost all nouns do not. Other examples include the paucal in Avar (data and source in Corbett 2000: 96-97), and as we now see, the paucal in Bezhta (like Avar, a Daghestanian language). The paucal is presented in Xalilov (1985), and the minority of nouns which have the paucal are indicated in the Bezhta dictionary (Xalilov 1995). The examples in Xalilov (1985) are mainly inanimates (concrete), with a few nouns denoting animals.

(38) Example of the minor paucal in Bezhta (Xalilov 1985)

SINGULAR	PAUCAL	PLURAL
sik ‘wineskin’	sika	sikla

According to Madžid Xalilov (personal communication), the use of the paucal, when available, is obligatory for reference to a small number; there is no difference for agreement between plural and paucal.

*Case* behaves similarly. The Russian second locative was discussed above (§4.2) as being non-canonical in that it shows a minimal difference within the paradigm. As mentioned there, it also non-canonical because there is a distinct form only for a relatively small number of nouns out of many thousands:

- (39) Nouns with the second locative: Ilola & Mustajoki (1989: 42-43)<sup>17</sup> from Zaliznjak (1977)<sup>18</sup>

inflectional class	example	nouns with second locative available	of these, second locative optional
I	<i>na beregú</i> ‘on the bank’	128	33
III	<i>v stepí</i> ‘in the steppe’	31	8

Clearly we have to distinguish case values like these from those which apply right across the noun lexicon. The issues of variability between speakers, acquisition by children, and observable change over a relatively small time-span indicate that we should take seriously the difference in status between features values that are close to canonical and those like the second locative which have marginal status.

*Gender* is the feature for which minor values have been most discussed; there are special terms, and for good reason. Let us start from an idealized (canonical) situation, in which each gender has a substantial number of nouns (controllers) and is reflected through agreement in a large number of targets, according to our four criteria. The deviations in terms of criterion 4 may affect the controller or the target. If we have a target, with a ‘problem’ (it has insufficient controllers), we label this a ‘minor target gender value’, though generally in the literature the simple ‘minor gender’ is used; see Corbett (1991: 159-160) for discussion of terms and earlier uses.

There is interesting variety in the possibilities here. Consider first the limiting case, termed a **non-lexical value**, in which we have a gender with *no* nouns in it.

These in turn are of two types; the first, the ‘neutral’ gender, is a target gender form which cannot normally have a canonical noun phrase headed by a noun or pronoun as its controller. When analyzing the Surselvan dialect of Romansh (Haiman 1974: 130-134), we would postulate two genders (following the agreement class approach); the masculine marked *-s* on agreeing targets, sometimes with a change of root vowel, and the feminine in *-a*. However, there is a third form of gender agreement: this is found with controllers which are not specified for gender and number: sentential subjects, the demonstrative pronoun *quei* ‘that’ and the impersonal pronoun *igl*. For example:

<sup>17</sup> Brown (2007) reports similar but slightly lower figures, and adds interesting information on frequency. The form is in overall decline and so published figures tend to be overestimates of its current use.

<sup>18</sup> Zaliznjak also includes *v zabyt’í* ‘in a state of unconsciousness’ and *v poluzabyt’í* ‘in semiconsciousness’ (Zaliznjak 1967/2002: 287), and does not mark these as optional. However, Plungian (2002) states that the distinction is largely lost for these two nouns, with one or other form (in *-e* or in *-í*) being generalized for all uses. This view is adopted in the 2003 revision of Zaliznjak’s dictionary.

- (40) Surselvan Romansh: (Haiman 1974: 130–132, Corbett 1991: 215)  
 Igl ei sesalzau in urezi  
 there is arisen[NEUTRAL] a storm  
 ‘a storm arose’

The form which is found in (40) cannot occur with a noun phrase headed by a noun as controller; we term this target gender the ‘neutral’ gender.

The second type has no nouns uniquely in it. A nice example is found in Walman, a Torricelli language of Papua New Guinea; my thanks to Lea Brown and Matthew Dryer for the data, now available in Brown & Dryer (ms.) and for discussion of their significance (our interpretations of the data differ):

Walman: minor target gender

- (41) Pelen n-aykiri.  
 dog M.SG-bark  
 ‘The male dog is barking.’

- (42) Pelen w-aykiri.  
 dog F.SG-bark  
 ‘The female dog is barking.’

- (43) Pelen l-aykiri.  
 dog DIMIN.SG-bark  
 ‘The puppy is barking.’

While *pelen* ‘dog’ can be masculine or feminine, there are numerous nouns that belong to the masculine or feminine gender. However, there are no nouns that belong uniquely to the diminutive gender.

Those were the two types of non-lexical values. Then there is the straightforward type of minor gender value, one which simply has few nouns in it (but is otherwise canonical). Here Lelemi (a Kwa language of south-eastern Ghana) provides an example; in addition to five clear gender values, there are two gender values, indicated with a unique agreement marker, but which appear to have extremely few nouns in them (Corbett 1991: 173-175, following Heine 1968).

There is a term proposed specifically for a combination of non-canonical behaviours, namely **inquorate** gender value. Inquorate genders are agreement classes which comprise a small number of nouns, but whose agreements can be readily specified as an unusual combination of forms available for agreement with nouns in the normal genders. The origin is that an inquorate meeting is one at which there are insufficient appropriate persons present to take decisions; hence an inquorate gender is an agreement class with insufficient nouns to deserve being labelled a gender. But note that the number of members is not the only criterion (now criterion 4); there is also the question as to whether the agreements can be characterized as an unusual

combination of forms available for agreement with nouns in the normal genders (criterion 1). Consider the following examples:

(44) Inquorate genders in Serbian/Croatian/Bosnian

SINGULAR		PLURAL		gloss
akt	MASCULINE	akta	NEUTER	document
oko	NEUTER	oči	FEMININE	eye
mače	NEUTER	mačići	MASCULINE	kitten

There are three genders in Serbian/Croatian/Bosnian, with robust agreement and a sizable number of nouns in each. There are much smaller groups of nouns, represented by those in (44), which do not fit into the three-gender system. However, they simply take an irregular combination of genders in the singular and plural.<sup>19</sup> They are non-canonical in being non-autonomous and in having few members. We can treat these as lexical exceptions, rather than postulating additional gender values. It is worth checking on the motivation here. Unlike a language like Lelemi, where the nouns in question control gender agreements with unique forms, the examples in (44) take agreement forms which are found elsewhere in the system. What is special about them is just the unusual combination of gender agreements according to whether the noun is singular or plural. In that latter respect they resemble Romanian; and yet, the situation is very different from that of Romanian in another respect: the non-autonomous neuter gender of Romanian has many hundreds of nouns (and gains more through borrowings), while those in (44) have relatively few members.

Finally for non-canonical behaviour in respect of criterion 4, we consider *person*. Here we need look no further than English, where the person values distinguished by *be* differ from those of regular verbs, with the modals then showing no distinctions at all.

## 5 Combinations of non-canonical behaviour

We have tried to separate out the non-canonical behaviours associated with feature values. It is important to note that these can coexist, as we have just seen when considering inquorate genders. A telling example of the coexistence of non-canonical behaviour is provided by the Russian second genitive (see Zaliznjak 1973, Worth 1984, and Comrie 1986). Contrast these forms of *kisel* 'kissel' (a thickened fruit drink) and *čaj* 'tea'. Both are members of inflectional class I (they vary somewhat from the forms in (1), but in predictable ways), and as expected both have the normal (first) genitive:

---

<sup>19</sup> Nouns which can have an unusual combination of gender features are the gender equivalents of heteroclites in terms of inflectional patterns. The nouns given are all heteroclites, and the different gender values are predicted from the inflectional patterns, as is a normal pattern of gender assignment.

- (45) vkus                      kiselj-a  
 flavour[SG.NOM] kissel-SG.GEN  
 ‘the flavour of kissel’
- (46) vkus                      čaj-a  
 flavour[SG.NOM] tea-SG.GEN  
 ‘the flavour of tea’

We find a contrast, however, in certain partitive expressions:

- (47) stakan                    kiselj-a  
 glass[SG.NOM] kissel-SG.GEN  
 ‘a glass of kissel’
- (48) stakan                    čaj-u  
 glass[SG.NOM] tea-SG.GEN2  
 ‘a glass of tea’

In contemporary Russian, in the active use of the speakers I have consulted, *kisel’* ‘kissel’ is an example of a regular noun, while *čaj* ‘tea’ is one of the subclass which has a separate second genitive.

According to Criterion 1, canonical features and their values have dedicated forms (are ‘autonomous’). With this in mind, consider these forms (within partial paradigms):

- (49) Russian partial singular paradigms

NOMINATIVE	kisel’	čaj
GENITIVE	kiselja	čaja
GENITIVE 2	(as genitive)	čaju
DATIVE	kiselju	čaju

Nouns like *kisel’* ‘kissel’ have no separate second genitive; in environments like (47) the normal genitive is used. For nouns like *čaj* ‘tea’, which do have a second genitive, it always has the form of the dative: it is non-autonomous.

Criterion 2 requires that canonical features and their values are uniquely distinguished across other logically compatible features and their values. The logically compatible feature is number. In fact the second genitive is found only in the singular, and has no distinct form in the plural. It is therefore non-canonical in respect of Criterion 2.

Criterion 3 states that canonical features and their values are distinguished consistently across relevant parts of speech. This is not straightforward, since in the modern language – at least for some speakers – the inclusion of an agreeing modifier disfavors the use of the second genitive; instead the ordinary genitive is more likely.

However, in those instances where the noun stands in the less likely second genitive in an expression similar to (48) genitive agreement is still required. Thus *krepkogo čaju* ‘strong tea’ is possible as a second genitive. This distributional test shows that we have to distinguish the second genitive from other forms. However, it clearly is non-canonical with respect to Criterion 3, since only the noun makes the distinction, while adjectives do not.

Finally, Criterion 4 requires that canonical features and their values are distinguished consistently across lexemes within relevant parts of speech. In fact, the number of nouns with this second genitive is restricted and declining.<sup>20</sup> They are all members of the inflectional class I. Of the nouns which have a second genitive, for some the second genitive is normally used in partitive expressions, for the others the second genitive is a possibility, but in competition with the ordinary genitive; for data on this see Panov (1968: 180), Graudina, Ickovič & Katlinskaja (1976: 121-125), Comrie, Stone & Polinsky (1996: 124-125), and especially Paus (1994). Thus the second genitive varies from being normally used, to being optional, to unusual – according to the particular lexical item.

The Russian second genitive is non-canonical in respect of each of our four criteria. Given its uncertain, marginal status, it is therefore somewhat ironic that it was one of the case values incorporated, alongside the robust case values, in Jakobson’s famous cube (1958).

## **6 Typology of feature values and non-canonical behaviour**

We should take stock. We set out to demonstrate the existence of a penumbra of non-canonical behaviour for the different features and their values. We looked only at the strictly morphosyntactic features, those with a role both in syntax and in morphology. For lack of data, since they are found only rarely as genuine morphosyntactic features, we omitted respect and definiteness (Corbett 2006: 135-138). We considered the criteria in order, starting with whichever feature provided clear or familiar data.

---

<sup>20</sup> Ilola & Mustajoki (1989: 41-41) reporting on Zaliznjak (1977), identify 396. However, the form has been in decline, continuing since the publication of Zaliznjak’s dictionary; see Brown, Tiberius & Corbett (2007: 522) for recent statistics. Our example helps show this: *kisel’* ‘kissel’ is given by Zaliznjak as having a second genitive, but consultants do not offer this form. A Google search gave over 200 examples of *stakan kiselja* ‘glass of kissel’ and none of *stakan kiselju*; however, other web searches reveal occasional sporadic instances of *kiselju* as a second genitive. A Google search on 29.5.08 produced four examples of *stakan kiselju* compared with around 1070 examples of *stakan kiselja*. All this shows that this particular noun has almost lost its second genitive, but not quite. For further discussion of the second genitive see Wierzbicka (1983: 249-252) and Uspenskij (2004: 11-26) and for a recent analysis of the second genitive in the Russian National Corpus see Brehmer (2009).



It is now time for a systematic summary of what we have found. In (50), the examples of non-canonical behaviour for each of the four criteria are included, feature by feature, together with the section where they were introduced.

(50) Non-canonical feature values with respect to the four criteria

	Criterion 1: autonomous	Criterion 2: distinguished across features/values	Criterion 3: distinguished consistently across parts of speech <sup>21</sup>	Criterion 4: distinguished consistently across lexemes
number		Biak §4.2	Mele Fila §4.3	Maltese §4.4
gender	Romanian §4.1	Russian §3.2	Mba §4.3	Lelemi §4.4
person	Old Nubian §4.1	Belhare §4.2	Maybrat §4.3	English §4.4
case	Classical Armenian §3.1	Russian §4.2	Guugu Yimidhirr §4.3	Russian §5

It is important to keep in mind our original aim, namely to demonstrate the types of non-canonical behaviour which are found, across the different features. This did not therefore require a systematic survey, though many languages were analysed. Given this, the picture that emerges in (50) is remarkably complete. We would not have expected such full coverage. It suggests that whatever the semantic and syntactic differences between the morphosyntactic features, there are strong similarities in their morphology. And this has implications for the status of the morphological component.

There is one gap in (50): we have not found any examples of non-autonomous number. We should be cautious, given the relative rarity of the examples we have overall: and yet, the fact that there is no example in the data collected for the Surrey Syncretism Database (Baerman, Brown & Corbett 2002) may be indicative. It is worth checking what an example would look like:

(51) Hypothetical non-autonomous number value

	SINGULAR	DUAL	PLURAL
NOMINATIVE	a	a	b
ACCUSATIVE	c	d	d

<sup>21</sup> For Criterion 3 the straightforward instances of non-canonicity would include Yimas (number), and gender and case in English; since we need to show just the existence of a particular type of non-canonical behaviour – and so one example is sufficient – we have rather plotted the more interesting instances of combined values in the table.

This hypothetical situation appears unlikely.<sup>22</sup> Let us suppose, while continuing to check the evidence, that there is a genuine gap in the pattern. We should ask why that would be. There is a suggestive remark in Carstairs-McCarthy (1994: 771). Discussing the relation of gender and number, he claims that the pattern of Romanian (as in Figure 4 above) must be of that type. Non-autonomous gender values (what he calls gender mixtures) can be based only on number. I suspect he is right, and this observation, though only partly relevant, is a useful hint: number is a good base for non-autonomous values of other features. There are various indicators for the primacy of number among the morphosyntactic features. I believe that the key one which makes it a good ‘base’ (for non-autonomous values of other features) is that lexemes with just one number value (singularia and pluralia tantum) are relatively common, while those with, say, one case value are not. Since number is a good base for non-autonomous values of the other features, this may be the reason why it cannot itself have non-autonomous values based on other features.<sup>23</sup> However, before speculating further, we should remind ourselves that the degree of coverage in (50) is already remarkable and further work will be needed to establish whether the gap in the data is indeed significant.

## **7 Conclusion**

Feature systems are frequently much less clean and neat than is generally supposed. In order to tackle the systems which have proved difficult to analyse, we adopted the canonical approach, previously applied specifically to case. While this type of detailed work has typically been carried out on the individual features separately, we noted intriguing similarities in the non-canonical behaviours of each. We then found a surprisingly complete distribution of the types of non-canonical behaviour over the different features. Thus the differences between the morphosyntactic features in their semantics and syntax do not determine similar differences in morphology, which illustrates again the autonomy of morphology.

---

<sup>22</sup> Note, however, that the conditions where it could arise are readily found. All that is required in principle is a number feature with two autonomous values, intersecting with one other two-valued feature (gender, case or person). That, in the right configuration, could give a non-autonomous number value, as in hypothetical (51).

<sup>23</sup> Jonathan Bobaljik offers an explanation for the gap (personal communication). He suggests that the examples of non-autonomous values that we find are combinations of otherwise attested syncretisms. Thus we find examples of first person being contrasted with syncretic second and third persons, and elsewhere first and second syncretic versus third person (see Baerman, Brown & Corbett 2005: 59-81 for examples). His suggestion is that the gap may be connected to a gap in possible syncretisms, if no language has a systematic singular and dual syncretism (opposed to plural). However, the Otomanguean language Pame has extensive singular-dual syncretism in the noun system, also affecting verb agreement (see Gibson & Bartholemew (1979), discussed in Corbett (2000: 121-122)). Hence we cannot account for the gap, if it proves robust, along the lines of Bobaljik’s suggestion.

There are three main implications following from this research. First, in theoretical syntax we tend to be concerned with the distribution of feature values; we need to bear in mind that these values may have radically different status, one from another. Thus there is no one simple numerical answer to a question like: ‘how many case values has Russian?’ Rather there are some case values which occupy such core positions that without them Russian morphosyntax would fall apart, and others with varying types of marginal status, each requiring careful analysis.<sup>24</sup> Second, a similar point holds for grammar writing. When we learn that, for instance, a language has a dual, we need clear information whether its dual is like that of Sanskrit, available across the relevant parts of speech and used obligatorily, or like that of Maltese. And third, the less clear areas of feature systems, their penumbra, are important for diachrony, since they show potential routes through which feature values can arise and die out.

## References

- Baerman, Matthew. 2002a. Armenian. In: Surrey Syncretisms Database. Available at: <http://www.smg.surrey.ac.uk/>
- Baerman, Matthew. 2002b. *Surrey Person Syncretism Database*. Available at: <http://www.smg.surrey.ac.uk/>
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2002. *The Surrey Syncretism Database*. Available at: <http://www.smg.surrey.ac.uk/>
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The Syntax-Morphology Interface: A study of syncretism*. Cambridge: Cambridge University Press.
- Bechhaus-Gerst, Marianne. 1996. *Sprachwandel durch Sprachkontakt am Beispiel des Nubischen im Niltal: Möglichkeiten und Grenzen einer diachronen Soziolinguistik*. Cologne: Rüdiger Köppe.
- Bickel, Balthasar & Johanna Nichols. 2005. Inclusive-exclusive as person vs. number categories worldwide. In: Elena Filimonova (ed.) *Clusivity: Typology and case studies of the inclusive–exclusive distinction* (Typological Studies in Language 63), 49-72. Amsterdam: John Benjamins.
- Blake, Barry J. 1994. *Case*. Cambridge: Cambridge University Press.
- Bliss, Heather. 2005. Constructing dual number in Hopi. In: Marie-Odile Junker, Martha McGinnis & Yves Roberge (eds) *Proceedings of the 2004 Canadian Linguistics Association Annual Conference*. [Available at: <http://www.carleton.ca/~mojunker/ACL-CLA/pdf/Bliss-CLA-2004.pdf> ]
- Bloomfield, Leonard 1933. *Language*. New York: Holt, Rinehart and Winston.
- Bobaljik, Jonathan D. 2008. Missing persons: a case study in morphological universals. *The Linguistic Review* 25.203-230.
- Brehmer, Bernhard. 2009. Changes and persistencies in the use of Russian masculine genitives in *-u*: New evidence from corpus linguistics. In: Sandra Birzer, Miriam Finkelstein & Imke Mendoza (eds) *Proceedings of the Second International*

---

<sup>24</sup> See Corbett (2008) for the detail.

- Perspectives on Slavistics *Conference: Regensburg 2006* (=Die Welt der Slaven special volume 36), 53-67. Munich: Otto Sagner.
- Brown, Dunstan. 1998a. From the General to the Exceptional: A Network Morphology Account of Russian Nominal Inflection. PhD thesis, University of Surrey.
- Brown, Dunstan. 1998b. Defining 'subgender': Virile and devirilized nouns in Polish. *Lingua* 104.187-233.
- Brown, Dunstan. 2007. Peripheral functions and overdifferentiation: the Russian second locative. *Russian Linguistics* 31.61-76.
- Brown, Dunstan, Carole Tiberius & Greville G. Corbett. 2007. The alignment of form and function: corpus-based evidence. *International Journal of Corpus Linguistics* 12.511-534.
- Brown, Lea & Matthew S. Dryer. Ms. Diminutive as an inflectional category in Walman. Ms. University at Buffalo. Available at: [linguistics.buffalo.edu/people/faculty/dryer/dryer/BrownDryerWalmanDimin.pdf](http://linguistics.buffalo.edu/people/faculty/dryer/dryer/BrownDryerWalmanDimin.pdf) [accessed 19.1.2009.]
- Browne, Gerald M. 2002. *Old Nubian Grammar* (Languages of the World/Materials 330). Munich: Lincom Europa.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70.737-788.
- Chumakina, Marina, Anna Kibort & Greville G. Corbett. 2007. Determining a language's feature inventory: person in Archi. In: Peter K. Austin & Andrew Simpson (eds) *Endangered Languages* (special issue of *Linguistische Berichte*, number 14), 143-172. Hamburg: Helmut Buske.
- Comrie, Bernard. 1986. On delimiting cases. In: Richard D. Brecht & James Levine (eds) *Case in Slavic*, 86-106. Columbus, OH.: Slavica.
- Comrie, Bernard. 1991. Form and function in identifying cases. In: Frans Plank (ed.) *Paradigms: The Economy of Inflection* (Empirical Approaches to Language Typology 9), 41-55. Berlin: Mouton de Gruyter.
- Comrie, Bernard. 2003. When agreement gets trigger-happy. In: Dunstan Brown, Greville Corbett & Carole Tiberius (eds) *Agreement: a typological perspective*. Special issue of *Transactions of the Philological Society* 101 no. 2, 313-37. Oxford: Blackwell.
- Comrie, Bernard, Gerald Stone & Maria Polinsky. 1996. *The Russian Language in the Twentieth Century*. Oxford: Clarendon Press.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2003. Agreement: Canonical instances and the extent of the phenomenon. In: Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds) *Topics in Morphology: Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001)*, 109-128. Barcelona: Universitat Pompeu Fabra
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007a. Canonical typology, suppletion and possible words. *Language* 83.8-42.

- Corbett, Greville G. 2007b. Deponency, syncretism and what lies between. In: Matthew Baerman, Greville G. Corbett, Dunstan Brown & Andrew Hippisley (eds) *Deponency and Morphological Mismatches* (Proceedings of the British Academy, 145), 21-43. Oxford: British Academy and Oxford University Press.
- Corbett, Greville G. 2007c. Gender and noun classes. In: Timothy Shopen (ed.) *Language Typology and Syntactic Description: III: Grammatical categories and the lexicon*, second edition, 241-279. 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: papers in honor of Bernard Comrie*, 1-34. Amsterdam: John Benjamins.
- Corbett, Greville G. Forthcoming a. Higher order exceptionality in inflectional morphology. To appear in: Horst J. Simon & Heike Wiese (eds) *Expecting the unexpected: Exceptions in grammar*. Berlin: Mouton de Gruyter. [With comments by Stephen Anderson.]
- Corbett, Greville G. Forthcoming b. Features: some key concepts. To appear in: Anna Kibort & Greville G. Corbett (eds) *Features: Perspectives on a Key Notion in Linguistics*. Oxford: Oxford University Press.
- Corbett, Greville G. Forthcoming c. Canonical morphosyntactic features. To appear in: Dunstan Brown, Greville G. Corbett & Marina Chumakina (eds) *Linguistic Diversity and the Universal: canonical approaches*.
- Corbett, Greville G. & Matthew Baerman. 2006. Prolegomena to a typology of morphological features. *Morphology* 16.231-246.
- 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.
- Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
- Cysouw, Michael. 2005. Syncretisms involving clusivity. In: Elena Filimonova (ed.) *Clusivity: Typology and case studies of the inclusive-exclusive distinction* (Typological Studies in Language 63), 73-111. Amsterdam: John Benjamins.
- Daniel, Michael. 2005. Understanding inclusives. In: Elena Filimonova (ed.) *Clusivity: Typology and case studies of the inclusive-exclusive distinction* (Typological Studies in Language 63), 3-48. Amsterdam: John Benjamins.
- Dol, Philomena. 2007. *A Grammar of Maybrat: a language of the Bird's Head Peninsula, Papua Province, Indonesia* (Pacific Linguistics, 586). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Evans, Nicholas. 2003. Typologies of agreement: some problems from Kayardild. In: Dunstan Brown, Greville G. Corbett & Carole Tiberius (eds) *Agreement: a typological perspective* (Special issue of *Transactions of the Philological Society* 101, no. 2) 203-34. Oxford: Blackwell.

- Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 2002. The semantics of gender in Mayali: Partially parallel systems and formal implementation. *Language* 78.111-155.
- Fankhauser, Franz. 1911. *Das Patois von Val d'Illiez (Unterwallis)*. Halle (Saale): Ehrhardt Karras.
- Fenech, Edward. 1996. Functions of the dual suffix in Maltese. *Rivista di Linguistica* 8.89-99.
- Foley, William A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William A. 1991. *The Yimas Language of New Guinea*. Stanford: Stanford University Press.
- Friedman, Victor A. 1993. Macedonian. In: Bernard Comrie & Greville G. Corbett (eds) *The Slavonic Languages*, 249-305. London: Routledge.
- Gibson, Lorna & Doris Bartholomew. 1979. Pame noun inflection. *International Journal of American Linguistics* 45.309-322.
- Gladkij, A. V. 2007. O točnyx i matematičeskix metodax v lingvistike i drugix gumanitarnyx naukax. *Voprosy jazykoznanija* no. 5.22-38.
- Goddard, Cliff. 1982. Case systems and case marking in Australian languages: a new interpretation. *Australian Journal of Linguistics* 2.167-196.
- Graudina, L. K., V. A. Ickovič. & L. P. Katlinskaja. 1976. *Grammatičeskaja pravil'nost' russkoj reči: opyt častotno-stilističeskogo slovarja variantov*. Moscow: Nauka.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In: Joseph H. Greenberg (ed) *Universals of Language*, 73-113. Cambridge, MA: MIT Press.
- Haiman, John. 1974. *Targets and Syntactic Change*. The Hague: Mouton.
- Harbour, Daniel. 2007. *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features* (Studies in Natural Language and Linguistic Theory 69). Dordrecht: Springer.
- Haviland, John. 1979. Guugu Yimidhirr. In: R. M. W. Dixon & Barry J. Blake (eds), *Handbook of Australian languages*, 27-180. Canberra: Australian National University Press.
- Heine, Bernd 1968. *Die Verbreitung und Gliederung der Togorestsprachen* (Kölner Beiträge zur Afrikanistik 1). Berlin: Reimer.
- Heine, Bernd. 1982. African noun class systems. In: Hansjakob Seiler & Christian Lehmann (eds) *Apprehension: Das sprachliche Erfassen von Gegenständen: I: Bereich und Ordnung der Phänomene*, 189-216. Tübingen: Narr.
- 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.
- Heuvel, Wilco van den. 2006. Biak: Description of an Austronesian language of Papua. PhD thesis, Vrije Universiteit Amsterdam [Available at: <http://www.lotpublications.nl/index3.html>]
- Hutchisson, Don. 1986. Sursurunga pronouns and the special uses of quadral number. In: Ursula Wiesemann (ed.) *Pronominal Systems* (Continuum 5), 217-255. Tübingen: Narr.

- Igartua, Iván. 2006. *Genus alternans* in Indo-European. *Indogermanische Forschungen* 111. 56-70.
- Iggesen, Oliver A. 2005. *Case-Asymmetry: A world-wide typological study on lexeme-class-dependent deviations in morphological case inventories*. Munich: Lincom Europa.
- Ilola, Eeva & Arto Mustajoki. 1989. *Report on Russian Morphology as it appears in Zaliznyak's Grammatical Dictionary* (=Slavica Helsingiensia 7). Helsinki: Department of Slavonic Languages, University of Helsinki.
- Jakobson, Roman O. 1958/1971. Morfoložičeskie nabljudenija nad slavjanskim sklonenijem (sostav russkix padežnyx form). In: *American Contributions to the Fourth International Congress of Slavists, Moscow, September 1958*. The Hague: Mouton, 127-156. [Reprinted in: Roman Jakobson (1971) *Selected Writings II*. The Hague: Mouton, 154-183. Translated 1984 as: Morphological observations on Slavic declension (the structure of Russian case forms). In: Linda R. Waugh & Morris Halle (eds) *Roman Jakobson. Russian and Slavic grammar: Studies 1931-1981*. Berlin: Mouton de Gruyter, 105-133.]
- Mel'čuk, Igor. 1986. Toward a definition of case. In: Richard D. Brecht & James Levine (eds) *Case in Slavic*, 35-85. Columbus, OH.: Slavica. [revised version in: Igor Mel'čuk. 2006. *Aspects of the Theory of Morphology*, edited by David Beck. Berlin: Mouton de Gruyter, 110-179.]
- 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.
- Nikolaeva, Irina & Andrew Spencer. 2008. Nouns as Adjectives and Adjectives as Nouns. Ms.
- Panov, M. V. (ed.) 1968. *Morfologija i sintaksis sovremennogo russkogo literaturnogo jazyka* (Russkij jazyk i sovetskoe obščestvo: Sociologolingvističeskoe issledovanie: III). Moscow: Nauka.
- Pasch, Helma. 1985. Possession and possessive classifiers in 'Dongo-ko. *Afrika und Übersee* 68.69-85.
- Pasch, Helma. 1986. *Die Mba-Sprachen: Die Nominalklassensysteme und die genetische Gliederung einer Gruppe von Ubangi-Sprachen* (Sprache und Geschichte in Afrika 6). Hamburg: Buske.
- Patri, Sylvain. 2007. *L'alignement syntaxique dans les langues indo-européennes d'Anatolie* (Studien zu den Boğazköy-Texten: herausgegeben von der Kommission für den Alten Orient der Akademie der Wissenschaften und der Literatur, Mainz, 49). Wiesbaden: Harrassowitz.
- Paus, Charles. 1994. Social and pragmatic conditioning in the decline of the Russian partitive case. *Russian Linguistics* 18: 249-266.
- Plungian [Plungjan], Vladimir A. 2002. K semantike russkogo lokativa. *Semiotika i informatika* 37: 229-254.
- Polinsky, Maria. 2003. Non-canonical agreement is canonical. In: Dunstan Brown, Greville Corbett & Carole Tiberius (eds) *Agreement: a typological perspective*. Special issue of *Transactions of the Philological Society* 101 no. 2, 279-312. Oxford: Blackwell.

- Pullum, Geoffrey K. & Hans-Jörg Tiede. forthcoming. Inessential features and expressive power of descriptive metalanguages. To appear in: Anna Kibort & Greville G. Corbett (eds) *Features: Perspectives on a key notion in linguistics*. Oxford: Oxford University Press.
- Rapold, Christian J. 2007. From demonstratives to verb agreement in Benchnon: A diachronic perspective. In: Azeb Amha, Maarten Mous & Graziano Savà (eds) *Omotiic and Cushitic Language Studies: Papers from the Fourth Cushitic Omotiic Conference, Leiden, 10-12 April 2003*, 69-88. Cologne: Rüdiger Köppe Verlag.
- Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia* (Pacific Linguistics, series C, no 98). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Seifart, Frank. 2005. The Structure and Use of Shape-based Noun Classes in Miraña (North West Amazon). PhD thesis, Radboud University, Nijmegen.
- Sparing-Chávez, Margarethe W. 1998. Amahuaca (Panoan). In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds) *Handbook of Amazonian Languages: IV*, 441-485. Berlin: Mouton de Gruyter.
- Spencer, Andrew. 2005. Extending deponency. In: Matthew Baerman, Greville G. Corbett, Dunstan Brown & Andrew Hippisley (eds) *Deponency and Morphological Mismatches* (Proceedings of the British Academy, 145), 45-70. Oxford: British Academy and Oxford University Press.
- Steinhauer, Hein. 1985. Number in Biak: Counterevidence to two alleged language universals. *Bijdragen tot de Taal-, Land- en Volkenkunde* 141.462-485.
- Stump, Gregory T. 2005. A non-canonical pattern of deponency and its implications. In: Matthew Baerman, Greville G. Corbett, Dunstan Brown & Andrew Hippisley (eds) *Deponency and Morphological Mismatches* (Proceedings of the British Academy, 145), 71-95. Oxford: British Academy and Oxford University Press.
- Stump, Gregory T. 2006. Heteroclis and paradigm linkage. *Language* 82.279-322.
- Stump, Gregory T. & Raphael Finkel. 2008. Stem alternations and principal parts in French verb inflection. Paper presented at Décembrettes 6: Colloque International de Morphologie, «Morphologie et classes flexionnelles», December 4-5, 2008, Université de Bordeaux, France.
- Suthar, Babubhai Kohyabhai. 2006. *Agreement in Gujarati*. PhD Dissertation, University of Pennsylvania
- Thornton, Anna. 2008. A non-canonical phenomenon in Italian verb morphology: double forms realizing the same cell. Paper read at the First Oxford Workshop On Romance Verb Morphology, 27-28 August, Oxford.
- Tucker, A. N. & M. A. Bryan. 1966. *Linguistic Analyses: The Non-Bantu Languages of North-Eastern Africa* (With a supplement on the Ethiopic Languages by Wolf Leslau). London: Oxford University Press.
- Uspenskij, Boris A. 2004. *Čast' i celoe v russkoj grammatike*. Moscow: Jazyki slavjanskoj kul'tury.
- Wierzbicka, Anna. 1983. The semantics of case marking. *Studies in Language* 7.247-275.
- Worth, Dean S. 1984. Russian gen<sup>2</sup>, loc<sup>2</sup> revisited. In: Joost J. van Baak (ed.) *Signs of friendship: to honour A. G. F. van Holk, Slavist, Linguist, Semiotician*. Amsterdam: Rodopi, 295-306.



- Xalilov, Madžid Š. 1985. Ob ograničennom množstvennom čisle suščestvitel'nyx bežtinskogo jazyka. In: K. Š. Mikailov (ed.) *Kategorija čisla v dagestanskix jazykax: sbornik statej*, 136-143. Makhachkala: Dagestanskij filial AN SSSR.
- Xalilov, Madžid Š. 1995. *Bežtinsko-russkij slovar'*. Makhachkala: Dagestanskij naučnyj centr Rossijskoj akademii nauk.
- Zaliznjak, Andrej A. 1967. *Russkoe imennoe slovoizmenenie* Moscow: Nauka. [Reprinted in: Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*. Moscow: Jazyki slavjanskoj kul'tury. 1-370]
- Zaliznjak, Andrej A. 1973. O ponimanii termina 'padež' v lingvističeskix opisanijax. In: Andrej A. Zaliznjak (ed.) *Problemy grammatičeskogo modelirovanija*. Moscow: Nauka, 53-87. [Reprinted in: Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*. Moscow: Jazyki slavjanskoj kul'tury. 613-47.]
- Zaliznjak, Andrej A. 1977. *Grammatičeskij slovar' russkogo jazyka: slovoizmenenie*. Moscow: Russkij jazyk. [A fourth, corrected edition appeared in 2003, Moscow: Russkie slovari.]
- Zwicky, Arnold. 1996. Syntax and phonology. In: Keith Brown & Jim Miller (eds) *Concise Encyclopedia of Syntactic Theories*, 300-305. Oxford: Elsevier Science.