

## UNIVERSALS AND FEATURES

Greville G. Corbett  
Surrey Morphology Group

Key terms: features, universals, typology, morphosyntax, standardization,

Abstract: Greenberg's paper on universals (1963) contains an interesting set of generalizations relating to features. It is a good time to review the issues involved in establishing universals of features. These verge on the philosophical at one extreme, while at the other they concern the practical question of how we present and gloss examples. Various initiatives concerned with standardization, taken broadly, are under way, and it is important that they should be fully informed by the linguistic issues. There are two main areas to discuss: the *Analysis problem* and the *Correspondence problem*.

*The Analysis problem*: for a given language, we need to be able to justify the postulation of any feature (such as number or case). Equally, for each feature in the language we need to be able to justify the set of values postulated (for example: singular, dual, paucal and plural; nominative, accusative and genitive). For some languages the analysis is trivially simple, in others it is exceptionally complex (for some there have been long-running debates). In this context, it is worth reviewing the work of the Set-theoretical School, given its undoubted relevance for typology. The difficulties posed by hybrids will be discussed; this leads naturally to typological hierarchies and the 'Canonical' approach in modern typology.

*The Correspondence problem*: as typologists we need to be able to justify treating features and their values as comparable across languages. This is not straightforward, and yet a good deal of typology, including enterprises such as the *World Atlas of Language Structures*, depends upon it. The problem has a second, more subtle version. Even within a single language, features and their values do not necessarily line up consistently. In Bayso, the number system of nouns and verbs interact in a complex way. In Romanian, the genders of nouns and adjectives differ, and there are many more such examples. Here a typological perspective can inform the analysis of a single language and, of course, a typology which ignored these languages would be considerably impoverished.

Features are an area where the concerns of the typologist meet those of computational linguists, formal linguists, fieldworkers, in fact linguists in many different guises. As we put increasing theoretical weight on features, it is important to review our assumptions and check our progress in understanding them.

# Universals and Features<sup>1</sup>

Greville G. Corbett  
Surrey Morphology Group

Greenberg (1963) includes several interesting generalizations on features. Universals 30-45 are relevant – in particular to morphosyntactic features. Since that time there has been substantial typological research into features. Equally in formal syntax, features have taken on an ever increasing significance. Indeed, Miller and Sag (1997: 579) call feature structures ‘the fundamental construct used to model linguistic entities’. It is now an appropriate time to review the issues involved. Some are profound, and will always be the subject of debate, while others are highly practical, concerning standardization and the presentation and glossing of examples. As is becoming generally accepted, we shall use the term ‘feature’ for gender, number, and so on, and ‘value’ for feminine, neuter ... and singular, dual, plural ... (compare Ramat 1998, Corbett 2006b). We shall concentrate on morphosyntactic features, taking the term strictly to imply features that are relevant to morphology and syntax. We are not here concerned with purely morphological features (for which see Corbett and Baerman 2007). We also distinguish morphosyntactic features from morphosemantic features, which are not relevant to syntax: see Stump (2005) and Corbett (forthcoming) for discussion; an example of a morphosemantic feature would be tense in the numerous languages where tense is morphologically distinguished on the verb but where this has no impact on syntax.

## 1 Why features?

It is worth reminding ourselves why we use features. There are different motivations which have converged, so that features are now shared across a very wide range of linguistic work, from the most theoretical to the highly applied.

### 1.1 *An abbreviatory device*

In one sense, features come ‘free’, since they do not increase the expressive power of a grammar. We can interpret a symbol like NP<sub>pl</sub> as a single ornate symbol (Halle 1969, Gazdar, Klein, Pullum and Sag 1985: 20-21). At this level of analysis, features are a useful abbreviatory device.

### 1.2 *A way of making generalizations*

The other side of the coin is that features allow us to make generalizations. They allow us to say, for example, that within a given language the same distinctions of number occur across different constructions (agreement within the noun phrase as

opposed to within the clause) and yet are realized differently across lexemes (thus *this : these :: runs : run*).

### 1.3 *The basis for typology*

Having isolated the distinctions which we model using features, it is natural to typologize across them, as in Greenberg (1963). As with all typology, we need to consider carefully whether we are comparing like with like, an issue to which we return in §6.

## 2 **Usefulness of features and issues of standardization**

Features are central to various initiatives concerned with standardization, taken broadly, some of which are currently under way. It is important that such initiatives should be fully informed by the linguistic issues.

### 2.1 **EAGLES**

The report on morphosyntactic annotation (Leech and Wilson 1996) is an early attempt to grapple with the issues. It was restricted to languages of the European Union, and does not fully distinguish part of speech and semantic subcategories from morphosyntax. Tags suggested for particular languages were included rather than their being rigorously compared with the general set established for a wider range of languages.

### 2.2 **Lexical markup framework (LMF):**

The International Organization for Standardization (ISO), in particular Technical Committee ISO/TC 37, *Terminology and other language resources*, Subcommittee SC 4, *Language resource management*, is working on ISO 24613 ‘Language resource management – Lexical markup framework’. (Revision 14 was circulated in mid July 2007.) The goals, as stated in the introduction (p. 5) are as follows:

Lexical Markup Framework (LMF) is an abstract metamodel that provides a common, standardized framework for the construction of computational lexicons. LMF ensures the encoding of linguistic information in a way that enables reusability in different applications and for different tasks. LMF provides a common, shared representation of lexical objects, including morphological, syntactic, and semantic aspects.

The goals of LMF are to provide a common model for the creation and use of electronic lexical resources ranging from small to large in scale, to manage the exchange of data between and among these resources, and to facilitate the merging of large numbers of different individual electronic resources to form extensive global electronic resources. The ultimate goal of LMF is to create a modular structure that will facilitate true content interoperability across all aspects of electronic lexical resources.

A previous draft was quite disappointing from a linguist's standpoint; however, the Committee has taken on board comments from linguists, and the latest draft is considerably improved.

### **2.3 *E-MELD (Electronic Metastructure for Endangered Languages Data) and GOLD (General Ontology for Linguistic Description)***

E-MELD had two primary objectives: contributing to preserving data on endangered languages, and helping to develop the infrastructure for effective collaboration between electronic archives (Aristar Dry 2002). The first objective was focused on best practice, in a variety of areas. So far as it concerned morphosyntactic markup, the direction was not so much to suggest a standard, as to ensure that non-significant differences in annotation should not hamper further understanding and analysis. This was consonant with the second objective, and led to initial work on an ontology of linguistic concepts (Farrar and Langendoen 2003). Work continues in this direction, taking account of the notion of canonicity (discussed in §5.2).

### **2.4 *The Leipzig Glossing Rules***

The Leipzig Glossing Rules (Comrie, Haspelmath and Bickel 2004), which build on earlier work particularly by Lehmann (1983), represent a bottom-up approach to standardization. They are available at:

<http://www.eva.mpg.de/lingua/files/morpheme.html>

At the simplest level it is eminently sensible that we should use the same symbols (e.g. '=' for clitic boundary) and the same abbreviations. We should be certain whether a colleague wishes to indicate perfect tense or perfective aspect.

While the proposals in the Leipzig Glossing Rules may seem low-level and relatively uncontroversial, this perspective on them may change quite rapidly if one tries to apply them consistently for a large and diverse set of examples (as in Corbett 2006a). The first observation is that glossing to the level of detail required is a demanding undertaking. And second, quite substantial issues come to the surface surprisingly quickly. The reason is that there are numerous problems with the analysis of features (which will come to shortly) and these come to the fore in glossing. Before considering those further, let us ask what is, or at least might be, universal.

## **3 Can there be universals in this area?**

Given the genuine difficulties of analysis, it is worth asking what we may hope to identify as universal. A simple suggestion can be found in Zwicky (1986: 988), namely that 'universal grammar should permit only a finite number of attributes and values – indeed [...] universal grammar should provide finite lists of the attributes and values available for service in a particular grammar'. Zwicky points out the difficulty with the approach, as put to him by Gerald Gazdar:<sup>2</sup>

Gazdar's challenge (in personal communication) cuts deeper. He observes that there is a serious correspondence problem involved in talking about 'the illative case' in two different languages: what allows us to identify the two grammatical cases? Similarly for other agreement properties, other head properties, and foot properties as well.

This is not the place to mount a full response, but I believe it is possible to require that every property on the lists have semantic concomitants. I am not maintaining here that these properties are to be IDENTIFIED with semantic features; grammatical categories are virtually always arbitrarily distributed (from the semantic point of view) in the lexicon to some extent. I am suggesting that a head or foot property is never a FULLY arbitrary and language-particular categorization of words and phrases: it has a semantic core that runs across languages.

Zwicky (1986: 988-989)

Zwicky's suggestion, then, is that morphosyntactic features always have a semantic core, and it is this core which allows comparison. We return to this issue in §6 below.

#### **4 The analysis problem: features**

For a given language, we need to be able to justify the postulation of any feature. Since as we have seen features are an abbreviatory device, we have to ask persistently whether each is needed. This is essential for the typologist, since there is the danger of always finding the features we expect, especially if we take functions as the starting point. An interesting example of an argument that an accepted feature is not actually required is Spencer's (forthcoming) analysis of Hungarian, in which he argues that there is no need to recognize a case feature for that language.

##### **4.1 Phonological form**

We might reasonably assume that in order to postulate a morphosyntactic feature, and its various values, we would require that for each there would be an inflected form (unique in its phonology) which could be explained only in terms of the particular feature and value. The discussion is usually for justifying particular values, but it is necessary at the feature level too. The existence of a unique form may seem an obvious requirement, but in fact it is too strong. There are situations where a feature is justifiable even though there is no dedicated form to support it. Thus Chumakina, Kibort and Corbett (2007) argue that the feature person is required in the grammar of Archi, on the basis that it is required if the resolution rules of the language are to be stated simply and in ways which are plausible in cross-linguistic terms.

##### **4.2 Conditions versus features**

It is important for typologists to distinguish clearly between morphosyntactic features and conditions. As a brief example, consider these data on agreement with

conjoined noun phrases in Russian:

(1) Predicate agreement with conjoined noun phrases in Russian

subject type word order	animate		inanimate	
	N	%PL	N	%PL
subject-predicate	115	100	67	85
predicate-subject	89	84	114	28

The data are from a corpus of literary works (Corbett 1983: 106, 128, 130). They indicate clearly that the word order has a major influence on the agreement form selected. The plural, the semantically agreeing form, is more likely to be found than the singular if the subject precedes the predicate. We would not add word order to the list of morphosyntactic features. Rather we say that it is a possible condition of the use of a morphosyntactic feature (number in this instance). The data show the effect of a second condition, namely animacy. If the subject is semantically animate, plural agreement is more likely than if it is inanimate. (Russian does have animacy as a subgender, but it is rather semantic animacy which influences the agreement choice we are considering.) For valid typological comparisons we must distinguish between morphosyntactic features and conditions on their use. Thus respect is often a condition on the use of a feature (often number, sometimes person) and yet it may also be a feature in its own right, with a dedicated form (§4.1). Conditions have interesting properties (for instance, they have consistent effects cross-linguistically); for discussion and key examples see Corbett (2006a: 176-205).

## 5 The analysis problem: values

Equally, for each feature in the language we need to be able to justify the set of values postulated (for example: singular, dual, paucal and plural; nominative, accusative and genitive). For some languages the analysis is simple, in others it is exceptionally complex (as demonstrated by discussions in the literature that have persisted over decades).

### 5.1 *Set-theoretical approaches*

In this context, it is worth reviewing the work of the Set-theoretical School, given its undoubted relevance for typology (and the fact that fifty years have just passed since the first meeting of the famous seminar on mathematical linguistics in Moscow). The famous mathematician Andrej Kolmogorov posed the following questions (van Helden 1993: 138):

“What exactly do we mean when we say that two words are in the same case?”

“How many cases does the Russian language possess?”

There was a flowering of interesting work on such questions, including particularly relevant work by Zaliznjak (1973). It is carefully surveyed by van Helden (1993) and a good introduction is Meyer (1994). In brief, Zaliznjak and others worked out careful and consistent methods for determining the feature and value inventory of a language (and a good deal of substandard work in typology could have been avoided if their legacy were better known). Typically the expected features and values are established, but less clear instances often emerge too, that is, the formal approach highlights interesting data and challenges. A phenomenon recognized and documented within this approach is the instances of non-autonomous case values (Zaliznjak 1973: 69-74). Here there is no form uniquely associated with a particular value, but the value is justified on the basis of syncretic forms. There are comparable non-autonomous gender values (Corbett 1991: 150-154); an alternative term is *genus alternans*; see Igartua (2006) and references there for discussion of the development of such instances in Indo-European.

Set-theoretical accounts are ‘brittle’, by which I mean that finding a single additional lexeme or context may be sufficient to invalidate an analysis. This is exactly what is required in terms of falsifiability – it is clear what constitutes a counter-example. And yet morphosyntactic feature systems often seem to be less rigid than such analyses allow. We consider an instance of this in the next section.

## 5.2 Hybrids

A significant problem for set-theoretical approaches is hybrids, that is, controllers whose feature specification varies according to the target. A familiar example is *committee* nouns in various varieties of English.

(2) *Committee* nouns in spoken American English and British English (Levin 2001: 109)

	verb		relative pronoun		personal pronoun	
	N	% plural	N	% plural	N	% plural
LSAC	524	9	43	74	239	94
BNC	2086	32	277	58	607	72

Note: LSAC = Longman Spoken American Corpus  
 BNC = British National Corpus (section on spoken language)

We see that the number value for *committee* and similar nouns varies according to the target. It is not straightforwardly singular or plural. Early researchers in the Set-theoretical School were aware of the issue. One obvious approach was to treat hybrids as having a different feature value. This would work if all hybrids behaved alike. However, this turns out not to be the case. Evidence demonstrating this for gender is reported in Corbett (1991: 183-184), and for number in Corbett (2006a:

213). Since each hybrid can be different, each would require a different feature value, and the number of values would be hugely extended.

The alternative is to restrict the number of feature values, essentially to those required by non-hybrid nouns (already the notion of ‘canonicity is coming into play, to be discussed further below). The problem of hybrids is then dealt by two interrelated means. First by typological hierarchies; in the example in question, this would be the Agreement Hierarchy:

- (3) Agreement Hierarchy (Corbett 2006a: 207)
- attributive > predicate > relative pronoun > personal pronoun

This hierarchy allows us to constrain possible agreement patterns as follows:

For any controller that permits alternative agreements, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement with greater semantic justification will increase monotonically.

Thus the variability in the morphosyntactic feature specification of hybrids is constrained, rather than varying freely. The second part of the analysis is conditions (as discussed in §4.2); these may involve semantic information, down to the detail of particular lexical items.

This general approach is consonant with the Canonical approach in modern typology. Here we set up clear definitions and take them to the logical end point, defining a theoretical space before asking where particular examples fit into it. In the problem just discussed, our definitions (specifically our definitions of feature values) will be based on controllers which take consistent agreements. Hybrids are then non-canonical. Furthermore, particular feature values in a given language may be more or less canonical. While in this way we avoid the explosion in the number of feature values, we leave open the possibility that a particular feature might have some certain values and further values of less certain status (a classic instance is the Russian case system; see, among others, Zaliznjak 1973, Comrie 1986, Mel’čuk 1986/2006, Corbett forthcoming).

## 6 The correspondence problem: cross-linguistic

As typologists we need to be able to justify treating features and their values as comparable across languages. This is not straightforward, and yet a good deal of typology, including enterprises such as the *World Atlas of Language Structures*, depends upon it. At the level of features, provided we are concentrating on morphosyntactic features, there is rarely a problem. That is, we know whether we are comparing case across languages, as opposed to gender or person. At the level of values, however, which is as Gazdar stated the correspondence problem, things are more difficult. There are two ways forward. The first is to avoid the problem by

lowering our sights to comparing systems only in terms of size (that is, the number of distinct values). We can make generalizations about the maximal and minimal systems (as in §8.3 below). And, as Greenberg did (1963), we can make claims about interactions between feature values; however, even this requires us to be able to compare at least some feature values cross-linguistically.

The second way forward, the one Zwicky suggests, is that ‘it is possible to require that every property on the lists have semantic concomitants’ (1986: 988). At the level of features, this is plausible. Thus gender always has a semantic core (Aksenov 1984, Corbett 1991: 8-69) and we could extend that to every morphosyntactic feature. There is the issue of case, which Zwicky (1992) treated as indirect, that is, as not ‘associated directly with prototypical, or default, semantics’ (1992: 378). Nevertheless, the argument for a correspondence with semantics, taken broadly, can be made. This is also the stance of Svenonius (2007). However, Gazdar’s point concerns values, and here the issue is more difficult.<sup>3</sup> For some values, cross-linguistic comparison is straightforward: feminine gender is the value which includes nouns denoting females, and the interesting typological considerations are what other nouns may be included in this gender value, how the assignment rules overlap or are distinguished from others, and whether or not the feminine is the default gender value. We need to define the core meanings and functions: we call a gender value the feminine if it includes nouns denoting females, whether or not it also includes diminutives. Similarly we call a case value the dative if used for recipients, whether or not it can also be governed by prepositions. But it does not follow that all values can be compared in this way. While gender features always have a semantic core, it is not clear that all gender values have a semantic core. The issue needs to be resolved first at the level of the individual languages. If it proves to be the case that some values have no semantic core, then we should compare features first in terms of the semantic core, and then in terms of the possible remaining values which fall outside that core.

## **7 The correspondence problem: intra-linguistic**

While Gazdar stated the correspondence problem in cross-linguistic terms, there is an analogous, more subtle intra-linguistic version. Even within a single language, features and their values do not necessarily line up consistently.

An instance of a feature which does not correspond within a given language is number, of the nominal and verbal types. Nominal number is concerned with the number of entities; it may appear on targets by agreement. Thus in *Mary runs*, number is of the nominal type, reflecting the fact that *Mary* is one individual, and not that there is a single running event. Verbal number indicates the number of events, or the number of participants in events. The two work rather differently, as documented in Corbett (2000: 243-264). However, they can appear together, as illustrated by Georgian (Corbett 2000: 254-255). Only nominal number is a morphosyntactic feature.

Turning to the intra-linguistic lack of correspondence of values, a well-studied instance is the gender system of Romanian. Here nouns are assigned to three genders, while agreeing targets distinguish only two. In other words, there are three controller genders and two target genders (Corbett 1991: 150-154). There are various other examples, and we take a less familiar one for illustration, namely the number system in the Cushitic language Bayso (Hayward 1979, Corbett and Hayward 1987, Corbett 2006a: 172-174). In Bayso, the number systems of nouns and verbs interact in a complex way. Nouns mark four numbers (general, singular, paucal and plural), while verbs show singular agreement (and gender agreement) for general and singular, plural agreement with the paucal, and masculine singular agreement with the plural.

We should note that both for Romanian and for Bayso we are not considering small numbers of irregular lexical items. We are looking at the normal system which involves substantial proportions of the lexicon. This section shows that for some languages there may be no straightforward response to the questions 'how many gender values?' and 'how many number values?' Here a typological perspective can inform the analysis of individual languages and, of course, a typology which ignored these languages would be considerably impoverished.

## **8 What is universal?**

Given the care that must be taken over issues of correspondence, we may wonder what we can hope for when looking for universals of features. The strategy advocated here is to opt for the simplest outcome, and give that up only if it can be demonstrated to be unattainable.

### ***8.1 The simplest possibility***

We should start from the simplest possibility, which would be a Zwicky-type list. If again we restrict ourselves to clearly morphosyntactic features, it is clear that the well-established agreement features (phi-features) all qualify. These are gender, number and person. In addition, case is clearly relevant to syntax.

There are two further, less obvious morphosyntactic features. Respect is often conveyed by the use of other features, thus it is often a condition on the use of number or person (§4.2). However, it may also, if rarely, appear as a morphosyntactic feature.<sup>4</sup> This is shown by Muna (an Austronesian language spoken on Muna, off the southeast coast of Sulawesi).

- (4) Number and politeness markers in Muna (van den Berg 1989: 51, 82)

'go' (second person)	singular	plural
neutral	o-kala	o-kala-amu
polite	to-kala	to-kala-amu

These equivalents of '(you) go' vary according to number and politeness; *to-* marks polite address, irrespective of number.

Definiteness too can occur as a morphosyntactic feature, as these Norwegian data show:

Norwegian (Bokmål, Torodd Kinn and Tore Nessel, personal communications)

- (5) det            ny-e            hus-et            mitt  
 DEF.N.SG    new.DEF.N.SG    house(N)-DEF.N.SG    my.N.SG  
 'my new house'

- (6) mitt            ny-e            hus  
 my.N.SG    new-DEF.N.SG    house(N)[INDEF]  
 'my new house'

Clearly definiteness marking is sensitive to the syntactic environment, and so appears to qualify as a morphosyntactic feature, in a small number of languages. For more on definiteness marking in Scandinavian languages see Delsing (1993: 113-184), and for recent discussion see Hankamer and Mikkelsen (2002).

## 8.2 A possible need to extend: Kayardild

The possibility of maintaining a relatively small list of morphosyntactic features (in the strict sense, excluding morphosemantic features) is severely challenged by data from Kayardild. As well as ordinary cases, Kayardild has various verbalizing cases (Evans 2003). Consider the verbalizing dative (V\_DAT), which is used for beneficiaries. Its marker, which repeats through the noun phrase, is *-maru-*. The surprising thing is that this marker takes regular verbal inflections:

Kayardild (Evans 2003: 215)

- (7) ngada    waa-jarra    wangarr-ina    ngijin-maru-tharra    thabuju-maru-tharra  
 1SG.NOM    sing-PST    song-MOD\_ABL    my-V\_DAT-PST    brother-V\_DAT-PST  
 'I sang a song for my brother.'

In example (7) we see tense marked on different elements of the noun phrase; Evans gives comparable examples for aspect, mood and polarity. Thus while these features are often morphosemantic (since they need not be referred to by rules of syntax), this is not evident in Kayardild. We may analyse such examples in different ways; see Evans (2003) and Corbett (2006a: 138-140) for discussion. If one believes that

tense, aspect, mood and polarity are features of the clause, then marking of these features on more than one item is symmetrical marking, and hence not (canonical) agreement. (In dependency approaches this would be agreement.) Whatever our analysis, the Kayardild data show that we cannot limit the list of morphosyntactic features to the obvious core instances without careful argumentation.

### 8.3 *Minimal and maximal systems?*

It may not be possible to achieve typologies which specify the possible configurations of features (for instance, we might have liked to claim that a language cannot have a paucal unless it has a dual, but Bayso appears to be a counter-example). However, we may be able to specify the smallest and largest systems.

For most morphosyntactic features, the **smallest system** is the logically possible smallest system, that with two values. Thus many languages have two genders only (see Corbett 2005). Similarly number systems with just two values are commonplace.

One feature that might seem problematic here is person. Greenberg's universal number 42 (1963: 113) states that: 'All languages have pronominal categories involving at least three person and two numbers.' However, we would not necessarily treat pronominal distinctions as morphosyntactic, and so we might find two-valued systems without conflict with Greenberg's universal. And indeed, it has been claimed that the Daghestanian language Archi makes a binary morphosyntactic distinction between first and second person on the one hand and the third person on the other (Chumakina, Kibort and Corbett 2007). There are distinct pronominal forms, but no morphosyntactic evidence to split first and second persons. The relevant contrasts are shown in this paradigm:

(8) Person agreement in Archi

PERSON	NUMBER	
	SG	PL
1	gender agreement	Ø-
2	gender agreement	Ø-
3	gender agreement	gender agreement

In the plural, the marker for first and second persons is Ø- (the bare stem is used). As mentioned above (§4.1), there is no dedicated form for person agreement, since this marker is also that for genders III and IV in the third person. Yet we need a morphosyntactic person feature both to account for forms in the paradigm, and for the resolution rules. Other small systems are discussed in Cysouw (2003: 127-139); note that Cysouw discusses individual paradigms, so that a paradigm with only two person values may apply only to some items and not necessarily be indicative of the features values available in the language as a whole.

When we come to specify the **largest system**, this is naturally harder. See, for instance, the discussion of large case systems in Comrie and Polinsky (1998). However, for number, considerable progress has been made. It seems that the largest systems of number contain five values (see Corbett 2000: 39-42). The interesting point is that they appear in different configurations. Thus Mele-Fila has the values: singular, dual, paucal, plural, greater plural; while Sursurunga has singular, dual, paucal, greater paucal, plural.<sup>5</sup>

## 9 Conclusion

Features are an area where the concerns of the typologist meet those of computational linguists, formal linguists, fieldworkers, in fact linguists in many different guises. As we put increasing theoretical weight on features, it is important to review our assumptions and check our progress in understanding them. A reasonable strategy is to try for the simplest typology: fixed lists of features, of values, and of configurations of values. The latter two lists are different, since we know that the largest systems do not necessarily include all the attested values. Of course, our lists are open to challenge from every new piece of research, but we should be able to construct them with sufficient plausibility for us to wish to scrutinize claims for necessary extensions with a degree of scepticism.

## References

- Aksenov, A. T. (1984). K probleme èkstralingvistièeskoj motivacii grammatièeskoj kategorii roda. *Voprosy jazykoznanija* no. 1.14-25.
- Aristar Dry, Helen. (2002). E-MELD: Overview and Update. International Workshop on Resources and Tools in Field Linguistics, Las Palmas 26-27 May 2002. [available at: <http://linguistlist.org/emeld/documents/index.cfm> ]
- Berg, René van den (1989). *A Grammar of the Muna Language* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 139). Dordrecht: Foris.
- Chumakina, Marina, Anna Kibort and Greville G. Corbett. (2007). Determining a language's feature inventory: person in Archi. In *Endangered Languages* (special issue of *Linguistische Berichte*, number 14), ed. Peter K. Austin and Andrew Simpson, 143-172. Hamburg: Helmut Buske.
- Comrie, Bernard. (1975). Polite plurals and predicate agreement. *Language* 51.406-418.
- Comrie, Bernard. (1986). On delimiting cases. In *Case in Slavic*, ed. Richard D. Brecht and James Levine, 86-106. Columbus, OH.: Slavica.
- Comrie, Bernard, Martin Haspelmath and Balthasar Bickel. (2004). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. [available at: <http://www.eva.mpg.de/lingua/files/morpheme.html>]
- Comrie, Bernard and Polinsky, Maria. (1998). The great Daghestanian case hoax. In *Case, Typology and Grammar: In Honor of Barry J. Blake* (Typological Studies in Language 38), ed. Anna Siewierska and Jae Jung Song, 95-114. Amsterdam: John Benjamins
- Corbett, Greville G. (1983). *Hierarchies, Targets And Controllers: Agreement Patterns in Slavic*. London: Croom Helm.
- Corbett, Greville G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. (2000). *Number*. Cambridge: Cambridge University Press.

- Corbett, Greville G. (2005). The number of genders (chapter and map). In *The World Atlas of Language Structures*, ed. Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie, 126-129. Oxford: Oxford University Press.
- Corbett, Greville G. (2006a). *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. (2006b). Linguistic features. In *Encyclopedia of Language and Linguistics, 2<sup>nd</sup> Edition, Volume 7*, ed. Keith Brown, 193-194. Oxford: Elsevier.
- Corbett, Greville G. (forthcoming). Determining morphosyntactic feature values: the case of case. To appear in *Case and grammatical relations: papers in honour of Bernard Comrie*, ed. Greville G. Corbett & Michael Noonan. Amsterdam: John Benjamins.
- Corbett, Greville G. and Matthew Baerman. (2007). Prolegomena to a typology of morphological features. *Morphology* 16.231-246.
- Corbett, Greville G. and Richard J. Hayward. (1987). Gender and number in Bayso. *Lingua* 73.1-28.
- Cysouw, Michael. (2003). *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.
- Delsing, Lars-Olof. (1993). *The internal structure of noun phrases in the Scandinavian languages: a comparative study*. Lund: Department of Scandinavian Languages, University of Lund.
- Einarsson, Stefán. (1945). *Icelandic: grammar, texts, glossary*. Baltimore: Johns Hopkins Press.
- Evans, Nicholas. (2003). Typologies of agreement: some problems from Kayardild. In *Agreement: a typological perspective* (Special issue of *Transactions of the Philological Society* 101, no. 2), ed. Dunstan Brown, Greville G. Corbett and Carole Tiberius, 203-234. Oxford: Blackwell.
- Farrar, Scott and D. Terence Langendoen. (2003). A Linguistic Ontology for the Semantic Web. *GLOT International* 7, no. 3.97-100
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. (1985). *Generalized Phrase Structure Grammar*. Blackwell: Oxford.
- Gazdar, Gerald and Chris Mellish. (1989). *Natural Language Processing in Prolog: An Introduction to Computational Linguistics*. Wokingham: Addison-Wesley.
- Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, ed. Joseph H. Greenberg, 73-113. Cambridge, MA: MIT Press. [Paperback edition published 1966; page references to this edition.]
- Guðmundsson, Helgi (1972). *The Pronominal Dual in Icelandic* (University of Iceland Publications in Linguistics 2). Reykjavík: Institute of Nordic Linguistics.
- Halle, Morris. (1969). How not to measure length of lexical representations and other matters. *Journal of Linguistics* 5.305-308.
- Hankamer, Jorge and Line Mikkelsen. (2002). A morphological analysis of definite nouns in Danish. *Journal of Germanic Linguistics* 14.137-175.
- Hayward, Richard J. (1979). Bayso revisited: some preliminary linguistic observations - II. *Bulletin of the School of Oriental and African Studies, University of London* 42.101-32.
- van Helden, W. Andries. (1993). *Case and gender: Concept formation between morphology and syntax (II volumes)* (Studies in Slavic and General Linguistics 20). Amsterdam: Rodopi.
- Igartua, Iván. (2006). *Genus alternans* in Indo-European. *Indogermanische Forschungen* 111. 56-70.
- Leech, Geoffrey and Andrew Wilson (main authors). (1996). *EAGLES: Recommendations for the Morphosyntactic Annotation of Corpora (EAGLES Document EAG-TCWG-MAC/R)*.

- [available at: [www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/annotate.ps.gz](http://www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/annotate.ps.gz)]
- Lehmann, Christian. (1983). Directions for interlinear morphemic translations. *Folia Linguistica* 16.193-224.
- Levin, Magnus. (2001). *Agreement with Collective Nouns in English*. Almqvist & Wiksell.
- Mel'čuk, Igor. (1986). Toward a definition of case. In *Case in Slavic*. ed. Richard D. Brecht and James Levine, 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 ...'. *Russian Linguistics* 18.341-377.
- Miller, Philip H. and Ivan A. Sag. (1997). French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15.573-639.
- Ramat, Paolo. (1998). On categories and categorizations. In *Language Change and Typological Variation: In Honor of Winfred P. Lehmann on the Occasion of His 83rd Birthday: II: Grammatical Universals and Typology* [=Journal of Indo-European Studies, Monograph 31], ed. Carol F. Justus and Edgar C. Polomé, 330-347. Washington DC: Institute for the Study of Man.
- Spencer, Andrew. (forthcoming). Does Hungarian have a case system? To appear in *Case and grammatical relations: papers in honour of Bernard Comrie*, ed. Greville G. Corbett & Michael Noonan. Amsterdam: John Benjamins.
- Stump, Gregory T. (2005). Word-formation and inflectional morphology. In *Handbook of Word-formation* (Studies in Natural Language and Linguistic Theory 64), ed. Pavol Štekauer and Rochelle Lieber, 49-71. Dordrecht: Springer.
- Svenonius, Peter. (2007). Interpreting uninterpretable features. *Linguistic Analysis* 33.375-413.
- Zaliznjak, Andrej A. (1973). O ponimanii termina 'padež' v lingvističeskix opisanijax. In *Problemy grammatičeskogo modelirovanija*, ed. Andrej A. Zaliznjak, 53-87. Moscow: Nauka.
- Zwicky, Arnold M. (1986). German adjective agreement in GPSG. *Linguistics* 24.957-990.

---

<sup>11</sup> The support of the ESRC under grants RES-051-27-0122 and RES-062-23-0696, and of the AHRC under grant B/RG/AN4375/APN18306 is gratefully acknowledged. I thank Anna Kibort and two anonymous referees for helpful comments.

<sup>2</sup> Zwicky (1986: 988) also reports a challenge by Geoffrey Pullum, suggesting that a fixed list is not plausible, given the remarkable variety which is already known. Like Zwicky, I am not deterred by this.

<sup>3</sup> This is a point where we have to be explicit about whether we are discussing features or values; discussion often flits between the two, and the intention is sometimes not clear.

<sup>4</sup> For interesting discussion of the development of the plural pronoun as a politeness marker in Icelandic, and comparison with various other languages, see Guðmundsson (1972). However, in Icelandic, only pronoun choice is involved: the original dual pronouns, which became plurals, and the plural pronouns, which became polite pronouns, took plural verb agreement. The predicate adjective may be singular or plural for polite plural pronouns (Comrie 1975: 409, citing Einarsson 1945: 134), but again we are dealing with values of number, and not a distinct morphosyntactic feature.

<sup>5</sup> Corbett (2000: 29) is correct here for Sursurunga; there is a typo on page 42 of that publication.