

Canonical morphosyntactic features¹

Greville G. Corbett

Surrey Morphology Group

To appear in: Dunstan Brown, Greville G. Corbett & Marina Chumakina (eds) *Canonical Morphology and Syntax*. Oxford: Oxford University Press.

Draft of 5 July 2010, comments very welcome

Features like case, person and number are often taken for granted. They may be included in analyses without justification, as being too simple and uniform to require argumentation. In this, they contrast with lexical items, which can have highly idiosyncratic behaviour, and this contrast is part of the intuitive justification for features. And yet we should recognize that feature systems are more complex than that. Features vary: (a) according to how well founded they are, and (b) in how they distribute across the lexicon. To analyse this difficult area, I start from an idealized view, and then plot the deviations from that ideal. In other words, I take a ‘canonical’ approach.

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. The canonical instances are simply those that match the canon: they are the best, the clearest, the indisputable ones. Given that they have to match up to a logically determined standard, they are unlikely to be frequent. They are likely to be rare, and may even be non-existent. This is not an issue. The convergence of criteria fixes a canonical point from which the phenomena actually found can be calibrated. We may then go on to an investigation of the distribution of canonical and less canonical phenomena in terms of their frequency. The canonical approach has been applied to phenomena in syntax, notably agreement (Evans 2003, Corbett 2003, 2006, Comrie 2003, Polinsky 2003, Seifart 2005, Suthar 2006) and to phenomena in morphology, by Baerman, Brown & Corbett (2005: 27-35), Spencer (2005), Stump (2005, 2006), Corbett (2007), Nikolaeva & Spencer (2008), Stump & Finkel (2008) and Thornton (2008).² Within this approach, I have proposed (Corbett forthcoming) a specific set of converging criteria for canonical features and values, concentrating on the genuine morphosyntactic features. Various non-canonical behaviours have been identified. To take one example, we find that a lack of robust formal marking underlies non-autonomous case values. Rather

¹ Versions of this paper were read at: the University of Brighton (October 2007), the Australian Linguistic Institute (Sydney, July 2008), at the workshop “Creating Creating Infrastructure for Canonical Typology”, University of Surrey (January 2009) and at the Department of Linguistics, SUNY Stony Brook (March 2009). I am grateful to all who made comments and suggestions, and especially to Matthew Baerman, Dunstan Brown, Marina Chumakina, Richard Larson and Leon Poladian.

² A working bibliography of this growing body of research can be found at the Canonical Typology page:

<http://www.surrey.ac.uk/LIS/SMG/CanonicalTypology/index.htm>

than being restricted to one feature, like case, parallel behaviour recurs in different features, sometimes with different labels.

This line of research proves fruitful, but there appears to be a flaw: fully canonical morphosyntactic features are idealizations which seem indistinguishable one from another. To solve this problem, that is to differentiate canonical morphosyntactic features, we examine their interaction with parts of speech. In the canonical situation, the interaction follows four criteria: exclusiveness, exhaustiveness, closed class membership, and compositionality, each of which we discuss. The weakenings of these criteria define a space in which we can locate many problematic phenomena of morphosyntax.

A natural response is to appeal to semantics: after all, gender, person and number mean different things. Normally they do, and that gives us the basis for naming the features (just as we name particular values according to their common functions). But this is not a good strategy for the problem of identical canonical features. First, because there are tricky instances where semantics would be misleading: there are examples in English dialects (Siemund 2007) where ‘count’ is a criterion for gender assignment, while it “ought” to be concerned with number. Similarly ‘respect’ in languages like French is a criterion for the use of plural number. Second, this approach can lead to the illegitimate mixing of features, which we wish to avoid. And third, the whole spirit of the enterprise has been to look for logical criteria, and we should therefore attempt to do so here as well. Otherwise we could simply relabel any inconvenient examples of features or values.

One potential logical criterion is this: controller/governor and target/governee both carry the particular feature (or do not both carry the feature). This distinguishes the features as follows:

True for: gender, number, person, respect

False for: case

The point is that a target (say the verb) may be plural because the subject is plural; in this respect their feature specifications match. But if an object stands in the genitive because the verb governs the genitive, the verb is not itself genitive. Thus this potential criterion distinguishes agreement features from governed features (compare Corbett 2006: 7-8). Clearly we need to go further to separate out the features, and the approach we take will eventually encompass the distinction we have just noted. The strategy which proves fruitful is to examine the interaction with canonical parts of speech.

1.1 Canonical parts of speech

A word on terminology may be helpful. The term ‘part of speech’ is widely used, notably in HPSG. The term ‘word class’ is also found. LFG uses ‘lexical category’; we shall avoid this term, and similarly ‘lexical class’, because of the ambiguity of ‘lexical’. While for some linguists ‘lexical category’ is a category of lexemes, hence equivalent to ‘part of speech’, for others the term ‘category’ may be found as equivalent to ‘part of speech’, with a distinction drawn between ‘lexical category’ (part of speech with lexical meaning, for instance, noun, verb and adjective) and

‘functional category’ (part of speech without lexical meaning, such as auxiliary); this latter use of ‘lexical category’ is found particularly within Minimalism and its precursors.

The literature on parts of speech is extensive; Schachter (1985) was an important reference, more recently, Trask (1999), Anward (2000, 2001) and Schachter & Shopen (2007) are helpful, and Baker (2003: 21) gives an account of the treatment in Chomskyan generative grammar. It is occasionally suggested that categorization into parts of speech is not feasible: see Evans (2006) for a rebuttal of this position.

The key point for our purposes is summarized in Spencer (2005: 102); the following table, distilled from different sources including Stump (1998) and Croft (2000), is based on Spencer’s account:

Table 1: Canonical parts of speech

Canonical semantic functions	
Noun	denotes referential entities (‘things’)
Verb	denotes eventualities (process, telic or otherwise, and also states)
Adjective	denotes properties
Canonical syntactic functions	
Noun	argument (especially Subject) of a Verb; names an entity
Verb	relation between arguments; lexical head of a (temporarily situated) predication
Adjective	intransitive predicate; attributive modifier
Canonical inflection	
Noun	number, person, case, possession, gender, ...
Verb	tense-mood-aspect, voice, agreement, ³ ...
Adjective	agreement, comparison, ...

These are lists of likely characteristics, and there are default links between these characterizations. In the canonical situation, what Spencer calls the ‘morphologically coherent lexicon’, the members of the classes defined by these three specifications align perfectly (Spencer 2005: 102):

all syntactic classes correspond to uniquely characterized morphological classes and vice versa.

all semantic classes correspond to uniquely characterized morphological classes and vice versa.

all syntactic classes correspond to uniquely characterized semantic classes and vice versa.

³ The effect of Spencer’s including ‘agreement’ here is to reflect the fact that verbs and adjectives may show not only inherent inflection (for tense-aspect-mood, voice or comparison) but also contextual inflection, by agreement.

All that is necessary for our purposes is the notion of a canonical part of speech, which we define as one in which the semantic, syntactic and morphological behaviours line up perfectly.

1.2 Canonical parts of speech and canonical morphosyntactic features

We now return to the question of morphosyntactic features. Suppose we have some canonical or almost canonical morphosyntactic features, how do we differentiate them? Canonical morphosyntactic features can be described in terms of two overarching principles (covering ten converging criteria). According to Corbett (2008: 6, 10) a canonical morphosyntactic feature:

- I. has robust formal marking;
- II. is constrained by simple rules of syntax.

Various non-canonical behaviours have been identified. One type of deviation from Principle I is that a lack of robust formal marking can give rise to non-autonomous case values (Zaliznjak 1973). This is not something that we find only with case. We find parallel deviations with other morphosyntactic features, though sometimes there are different terms which conceal the parallelism (Corbett forthcoming). There seemed to be considerable progress along this line of research, until the following problem was noted: fully canonical morphosyntactic features are idealizations which appear indistinguishable one from another, since the structural properties that distinguish the features would all be non-canonical. We now tackle that issue. Recall that a canonical part of speech has a perfect alignment of semantics, syntax and morphology. For instance, a canonical noun would denote an entity, head a nominal phrase, and take the inflectional morphology appropriate in the given language. Given such canonical parts of speech, the canonical interaction with morphosyntactic features is diagrammed in Figure 5:

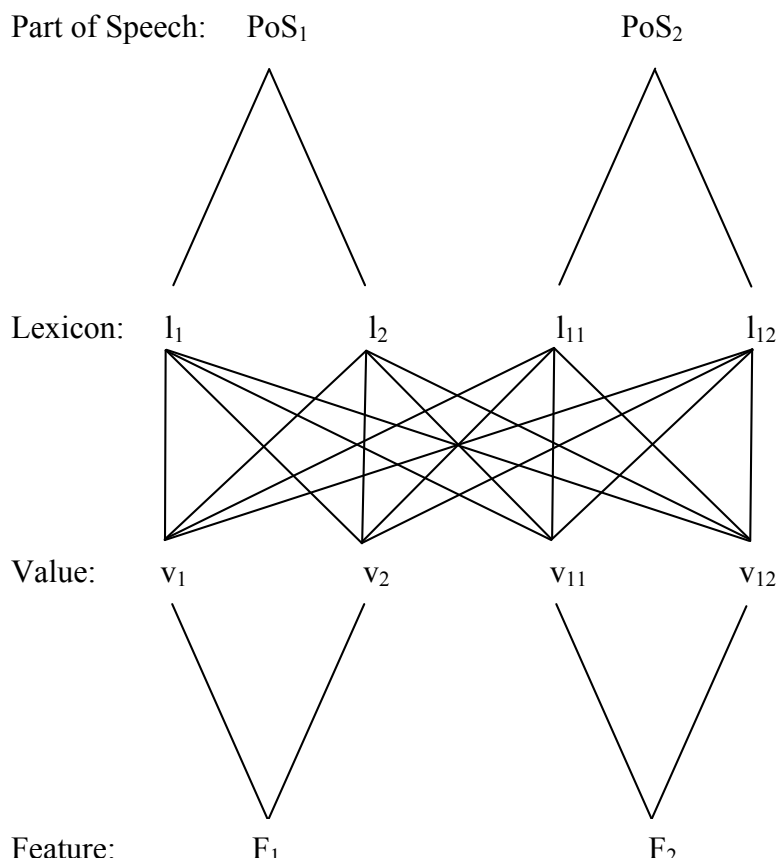


Figure 1: Canonical parts of speech and canonical morphosyntactic features

This schema rests on some important idealizations (canonical criteria), which need to be made explicit: it is the deviations from these criteria which give us the means to differentiate the features. We shall discuss general weakenings as we go through the criteria, since these are of interest in themselves, then return in §X.3 to consider how we can differentiate the features.

Criterion 1: **exclusiveness:**

A lexical item belongs to just one part of speech;
a value belongs to just one feature.

This criterion is relatively straightforward. Let us consider what happens if either part of it is weakened. If we allow a lexical item (abbreviated “l” in Figure 1) to belong to more than one part of speech, we find the well-known mixed categories such as participles. Participles may be analysed as belonging to the parts of speech ‘verb’ and ‘adjective’ (for discussion of a range of such mixed categories see Spencer 2005). Weakening of the second part would be more serious; if we allowed values to belong to more than one feature, an inventory of possible features and their values would become an elusive goal. There are, however, published analyses which do treat particular values as belonging to an unexpected feature, thus allowing them to belong to two features. Given the implications of this move, we should not allow this type of weakening if at all possible. I believe that the suggested analyses of this type are not optimal: there are better analyses for the data and these follow the requirement that a

value should belong to just one feature. For instance, in the Cushitic language Bayso, ‘plural’ had been treated as a value of gender; the system is reanalysed, preserving orthogonal features, in Corbett & Hayward (1987). Thus we wish to maintain the requirement of Criterion 1 for morphosyntactic feature values.

Criterion 2: **exhaustiveness:**

Every lexical item of every part of speech has available all values of all features.

(alternatively: every feature value applies to all lexical items)

In the clearest (canonical) instance, it is evident why we postulate a morphosyntactic feature, since it generalizes across a large number of items. The point needs making, since part of speech categories and features can in principle be ‘traded’ in a description: see Pullum & Tiede (forthcoming). If we did not postulate the feature, in a situation which was canonical according to Criterion 2, we would have to double the number of lexical items for each two-valued feature, and would miss evident generalizations. For instance, if a language has a fully regular number system, in which each noun and each adjective had singular and plural, we could of course avoid postulating a number feature by having additional lexical items; we would miss evident generalizations, and would require additional syntactic rules to give the effects normally attributed to agreement.

The weakenings of this criterion are of some interest. The first weakening is that particular features may not be available to all parts of speech. This is a fairly ‘clean’, weakening, since it applies to the higher-level elements involved (complete features and complete parts of speech).⁴ Thus, for instance, in Russian the feature number (both values) is not available to the part of speech preposition (all lexical items belonging to that part of speech). We might be tempted to think that semantic considerations should be invoked here. However, the fact that number *is* available to the verb, but not to the preposition, in Russian is not justifiable on semantic grounds: if a verb is plural, it does not denote more than one event, it shows agreement with a plural noun phrase. Hence in a particular language we need to specify which parts of speech have access to which features, and this is a weakening of the canonical situation. (We return shortly to the limiting instance for this criterion.)

A second, less neat weakening is that in which, within a part of speech, not all lexical items have access to all values of all features. The simplest instance is that in which some lexical items have access to a particular feature and some do not: the latter lie outside the choice (and may take a default value). Consider a language with three number values, singular, dual and plural: there may be nominals which are not number-differentiable. They have no number choice available (and may for instance appear in a default form equivalent to the singular). The distinction between number-differentiable nouns and those which are not number-differentiable is likely to be a principled one, constrained by the Animacy Hierarchy (see Corbett 2000: 54-75).

⁴ The lists of likely characteristics in Table 1 are of this type, giving typical features for particular parts of speech (and hence typical weakenings from the fully canonical situation).

The weakenings may go further in two directions. Instead of an ‘all-or-none’ type of access to feature values, some lexical items may have access to some but not all values. For instances of this in the number systems of Arapesh, Maori, Yimas and Manam see Corbett (2000: 90-93). And second, the lexical items with restricted access may be idiosyncratically labelled rather than being part of a principled division of the part of speech: we find unexpected gaps in the availability of feature values to particular lexical items, right through to defectiveness (where there is no acceptable form for a given context, see Baerman & Corbett 2010).

There are further weakenings of this criterion which will prove important when we come to the issue of differentiating the features in §X.3. At this point we should consider further the first weakening we discussed, in particular the limiting case where a feature applies to one part of speech only (see Figure 2).

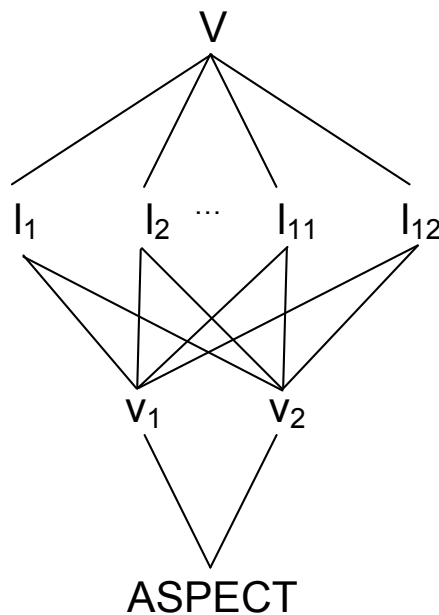


Figure 2: A feature available to just one part of speech

As indicated in our figure, a possible interpretation of this configuration is an aspect feature for verbs. In the simple situation where there are just two values of aspect, and each verb has access to both, this state of affairs could equally be represented as in Figure 3.

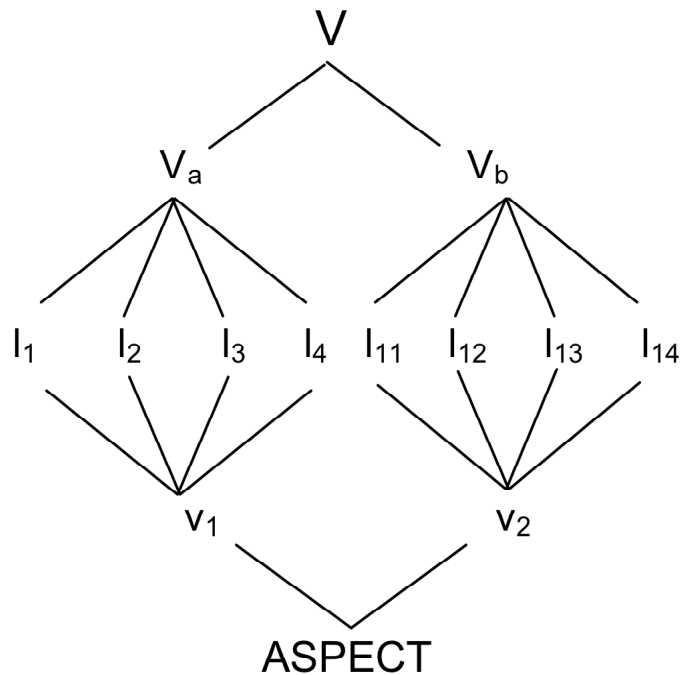


Figure 3: Alternative representation of aspect

In this alternative representation, we have two sub-categories of verb, and the members of each have access to only one value of the feature aspect. In the simplest situation, there is no obvious way to choose between the two representations. (Of course, as soon as the picture is less balanced, classical arguments such as regularity of formation and regularity of meaning will come into play.) For the present argument, the point is that a feature accessed by a single part of speech has special properties, and is to be treated separately. For instance, a feature available for one part of speech only may well show interactions with the lexical semantics of the individual lexical items, giving both idiosyncratic instances of lexical meaning and idiosyncratic gaps in the lexicon. It is this configuration with the attendant special properties which has led us to separate off such features as ‘morphosemantic’.

Having used our canonical scheme to locate morphosemantic features like aspect,⁵ we return to the possible weakenings, which will allow us eventually to situate the different morphosyntactic features.

Criterion 3: open and closed classes:

All classes are closed, except the class of lexical items.

This is the idealization which makes sense of the distinction between part of speech and morphosyntactic feature, in the regular correspondences between an open set of lexical items and closed sets of features and values. The canonical parts of speech have open membership (sometimes the term ‘lexical category’ is restricted to these). Their members (lexical items) are called ‘full words’ or ‘content words’.

⁵ We are concerned with differentiating the morphosyntactic features. Our approach separates off the morphosemantic features, but does not differentiate between them.

There are important weakenings here. First, some parts of speech (or often subcategories of parts of speech) can have closed membership. Besides the usual suspects (pronouns, conjunctions and adpositions) we find languages with a closed class of adjectives, for example. Conversely, features may have large inventories of values, with a membership that is hard to define. Note, however, that the values which cause the uncertainty (for example, spatial cases in large case systems) are typically morphosemantic in nature.

Criterion 4: compositionality:

Given the lexical semantics of a lexical item and a specification of its feature values, the meaning of the whole is fully predictable.

This criterion makes intuitive sense. A key motivation for separating out features and their values is the regularity involved. If we had to specify the lexical semantics of *tables* quite separately from *table*, and with no relation to pairings like *chairs* and *chair*, the attraction of the feature-based analysis would be considerably reduced. However, we certainly find weakenings of this criterion. Staying with number, we find principled weakenings: the items at the very top of the Animacy Hierarchy regularly show associative effects (*we* is typically ‘I plus associate(s)’), while items at the bottom of the Animacy Hierarchy often allow plurals only when recategorized (Corbett 2000: 83-87). We do also find quite idiosyncratic interactions; these are more usual, however, with morphosemantic features, as discussed above.

The weakenings of the four criteria define a space in which we can locate many problematic phenomena of morphosyntax. Here we concentrate on the key issue of differentiating features according to this typology.

1.3 Canonicity and common morphosyntactic features

We have defined a theoretical space, giving a canonical morphosyntactic feature in its canonical relations with a canonical part of speech. We have looked at the possible weakenings of this set of idealizations. We can now ask which is the canonical morphosyntactic feature. Let us return to our canonical schema in Figure 1, where each lexical item has access to every feature value. The morphosyntactic feature that comes closest to this idealization is number. We find instances where two (or more) parts of speech have all the number values, and indeed that all of their members do. True, it is common for the nouns involved to be constrained by the Animacy Hierarchy (Corbett 2000: 54-75) but equally there are languages where all, or almost all, of the members of different parts of speech (pronouns, nouns and verbs) have all the available numbers. Three examples, from typologically very different families, are: Central Alaskan Yup’ik (Eskimo-Aleut), Ojibway (Algonquian) and Miya (Chadic). These are discussed in Corbett (2000: 87); see that source for references. From our perspective, such languages are of particular interest.

There has been considerable debate on the relative status of features, including, among others, Greenberg (1963), Bybee (1985), Noyer (1997), Brown (1998), Baerman, Brown & Corbett (2005: 115-124). Particularly relevant for our discussion is the place of number in the work of the Set-theoretical School; there number is treated as first in the deductive order for noun features (see van Helden 1993: 949-955).

If we continue looking at deviations from the canonical schema in Figure 1 (where a deviation is a *weakening* of one of the criteria or a *combination* of such weakenings), we can locate the other morphosyntactic features.

1.3.1 Deviation 1

The exhaustiveness criterion (Criterion 2) states that in the canonical situation every lexical item of every part of speech should have available all values of all features. What if, however, the part of speech as a whole, rather than each lexical item, has access to all values of a feature? This deviation from the canonical is represented in Figure 4.

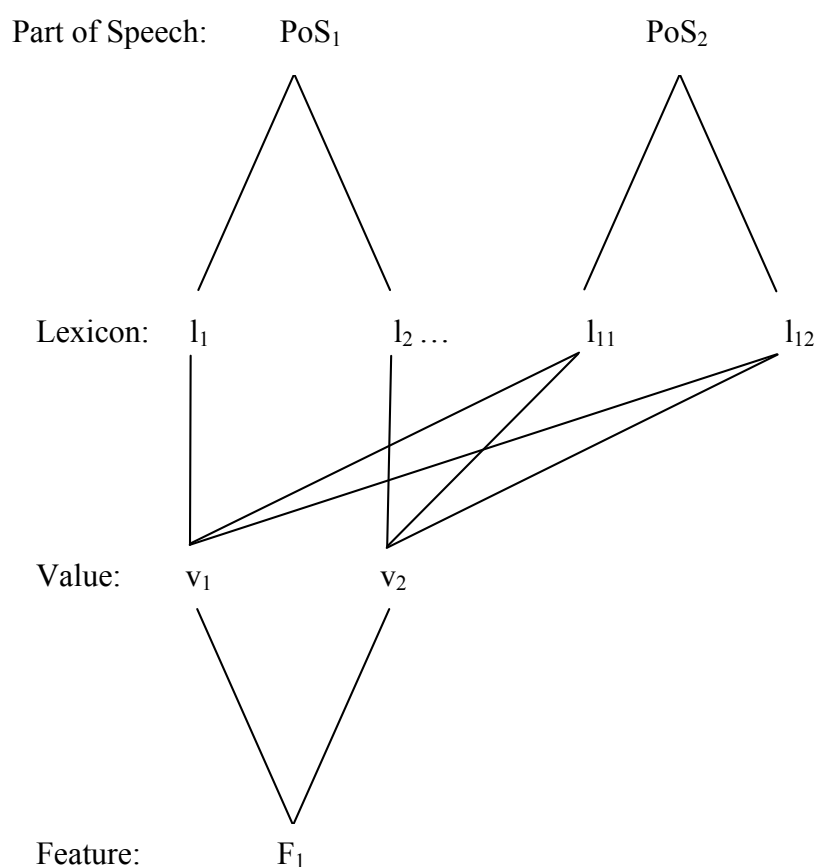


Figure 4: Canonical schema: deviation 1

We are concerned with the sole feature included here. It is relevant to two parts of speech, and so appears to qualify as a morphosyntactic feature. The deviation from canonicity involves its relation to the first part of speech (PoS₁ in Figure 4). While both of its values are available to that part of speech, as a whole, we do not find exhaustiveness. Rather we find that lexical entries in that part of speech select for just one of the values. This gives a clear asymmetry between the parts of speech involved.

Let us assume that there are no further deviations; for example, there are plenty of lexical items in the part of speech involved. We can see this as a representation of a

gender system: a typical situation would be that the first part of speech (PoS₁ in Figure 4) is noun, and the second part of speech (PoS₂ in Figure 4) is adjective.

1.3.2 Deviation 2

Let us maintain the schema recorded for Deviation 1, and add a second deviation, namely that the number of lexical items in the first part of speech is severely restricted. Do we find such instances in real languages? Certainly, if we think of the limited part of speech being the pronouns: we have few members of the class, and each has a fixed value for **person**. (The fit is particularly strong for those who treat the third person as a default, found with items which are not first or second person.) The fact that there are so few lexical items involved, with a straight matching to the person feature, is of course what leads to the ready availability of pronoun dropping (though note that we do not imply a processual account: ‘unirepresentation’ would be a better term, though tradition dictates ‘dropping’, Corbett 2006: 106).

A much less common feature, which is similar at this level of generality, is respect (in the few instances where it is a genuine morphosyntactic feature, rather than a condition on the use of another feature, such as number, as it is in languages like French and Russian). Here again the pronoun is the part of speech which has the restricted number of lexical items with access to the feature, and these require a particular value of the politeness feature.

1.3.3 Deviation 3

Let us return to our first type of deviation, and consider different parts of speech:

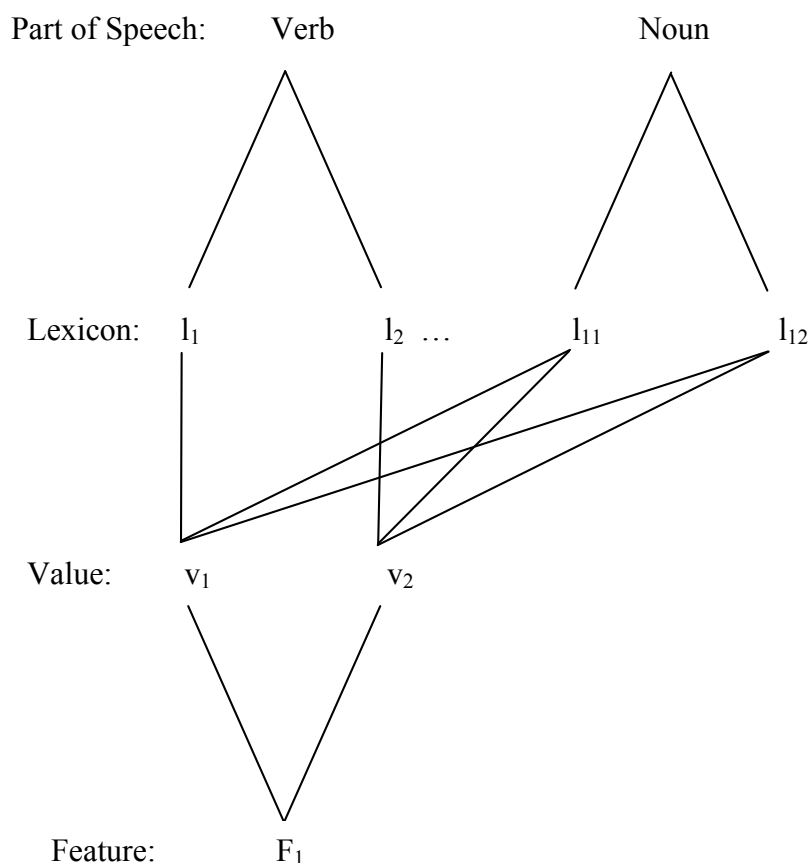


Figure 5: Canonical schema: Deviation 3 (based on Deviation 1)

With this labelling we have a representation of case. From one perspective, it is a mirror-image of gender, which seems right, up to a point. Gender is a classification of nouns, which determines the form of targets (such as verbs), while case requires a classification of governors (like verbs), which determines the form of their governees.

This parallelism is illuminating, but does not tell the full story. There is typically a patchier relation between the part of speech and the case requirement than we find with gender; different items require different numbers of governees, including no governees. This reflects deeper syntactic differences, which may be modelled using subcategorization.

More fundamentally, though, we need to examine the use of the notion ‘available’ in the exhaustiveness criterion. It could be argued that a verb having a case value available is not the same as a noun having a gender value available. Current syntactic frameworks, however, tend to treat such requirements and their satisfaction in comparable ways. If we believe there is a distinction to be maintained, it goes back to that raised in §X.0, namely that when we say a noun is feminine, we mean that it is in some sense feminine itself (it carries the relevant feature specification), and it controls feminine agreement; when we say that a verb takes the dative, we mean that it governs the dative but it is not itself dative (it carries a requirement for the dative). This distinction separates the two notions of ‘available’; we return to this point in §X.4 below.

Thus our canonical approach shows how the morphosyntactic features are related, and how they differ in some common manifestations. According to the type of deviation, we may order them according to decreasing canonicity:

number – gender – person – (respect) – case – (definiteness)

Number is closest to canonical, gender has one deviation, person has the same deviation plus a further one. The rare feature of respect fits next, being a less canonical version of person. Case is next, because of various non-canonical characteristics we have just discussed. Finally, if we need to recognize a morphosyntactic feature of definiteness (for those few instances where there is a plausible argument for it being a genuine morphosyntactic feature, see Corbett 2006: 135-137 for some discussion), we should treat it as somewhat like case. In addition to the deviations of case, it has the additional deviation of having few controllers (so not meeting Criterion 3).

1.4 Revisiting the criteria

We should now look back to our four criteria, as in §X.2, and check on what each is contributing. As Dunstan Brown points out (personal communication), if we take them at face value and consider the simplest situation where each can be satisfied or not, the four criteria would allow us to model 16 different situations. This is considerably more than the deviations we have discussed, and so it might appear that we have appealed to an excessive amount of theoretical machinery. As we shall see, this is not in fact the case; the criteria come close to giving us just the right inventory.

Criterion 1: **exclusiveness:**

A lexical item belongs to just one part of speech;
a value belongs to just one feature.

We allow for lexical items to belong to more than one part of speech, which gives rise to mixed categories like gerunds and participles, which are not our concern here. We suggested that weakening of the second part would be more serious, and decided not to allow values to belong to more than one feature. If we can maintain this limitation, then the number of types of feature we could describe with our criteria is reduced by half, to eight.

Criterion 2: **exhaustiveness:**

Every lexical item of every part of speech has available all values
of all features.
(alternatively: every feature value applies to all lexical items)

This criterion is key. As we saw, the only feature that approximates to it is number. This criterion therefore splits the possibilities into number on the one hand, and the rest on the other – the possibilities permitted by our two remaining criteria. This means that the possibilities are reduced to five. Gender is the feature that fails the exhaustiveness criterion but can, in principle, meet the remaining two.

Criterion 3: open and closed classes:

All classes are closed, except the class of lexical items.

This is the criterion which person cannot meet. Furthermore, if respect is to be included, it should probably fit within person.

Criterion 4: compositionality:

Given the lexical semantics of a lexical item and a specification of its feature values, the meaning of the whole is fully predictable.

The logic of the way in which the criteria fit to the features we find leads us to think more closely about Criterion 4. We noted earlier that morphosyntactic number, for instance, involves a matching between controller and target, and its relation to both controller and target is in principle compositional. This is not so for case. We may say of a noun phrase that the lexical meaning of the elements and the grammatical meaning of the case yield a predictable outcome. But this is not so for a verb and its requirement for governing a particular case. Thus Criterion 4, which is unlikely to be fully met by any feature in practice, is in principle not one that case could meet. It is thus Criterion 4 which covers the issue raised earlier, about the different nature of case.

We should ask, finally, if there is anything corresponding to the situation in which none of the criteria is met (except for the first, which is a requirement of features). One answer would be that such an object is simply too far from canonical to be a feature. An alternative would be that such an object is like case, but with the additional non-canonical behaviour of being lexically restricted (failing Criterion 3). If there is such a morphosyntactic feature it would be definiteness (recall that the status of definiteness needs further work).

A summary of this discussion is given in Table 2.

Table 2: The contribution of the criteria in determining possible morphosyntactic features

Criterion1 exclusiveness	Criterion2 exhaustiveness	Criterion3 open class	Criterion4 compositionality	
+	+	+	+	number
+	-	+	+	gender
+	-	-	+	person
+	-	+	-	case
+	-	-	-	?definiteness

If we recall that Criterion 1 must be met, and bear in mind the remaining discussion above, we observe that the criteria proposed reflect closely the morphosyntactic features that we actually find.

1.5 Further non-canonical feature values

We now move from features to their values. We shall see that the means of representation worked out above is helpful for getting to grips with some other non-canonical systems. Consider again the gender system in Figure 4. Let us go on to the limiting case, termed a ‘non-lexical value’, in which we have a gender value with *no* nouns in it.

Here in turn there are 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 analysing the Surselvan dialect of Romansh (Haiman 1974: 130-134), we would postulate two genders: 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:

- (1) 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 (1) cannot occur with a noun phrase headed by a noun as controller; we term this target gender the ‘neutral’ gender. We represent that as in Figure 6:

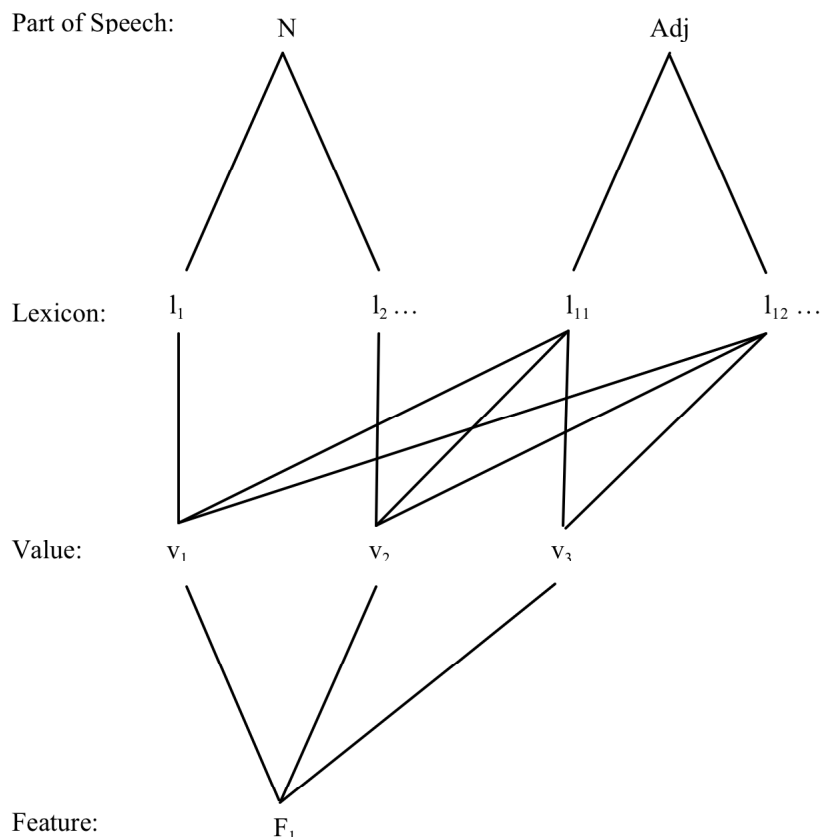


Figure 6: A non-lexical gender (Surselvan Romansh)

Here we see that the adjective has access to all three values of the gender feature (v_1 , v_2 and v_3), while the noun has access to two only (v_1 and v_2).

A more complex example is found in Walman, a Torricelli language of Papua New Guinea; thanks to Lea Brown and Matthew Dryer for the data (now in Brown & Dryer ms.) and for discussion of their significance (note that they do not agree that the diminutive is a gender):

Walman (Brown & Dryer ms.)

(2) Pelen n-aykiri.
 dog M.SG-bark
 'The male dog is barking.'

(3) Pelen w-aykiri.
 dog F.SG-bark
 'The female dog is barking.'

(4) 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. We represent the interesting situation of those nouns that can belong to one gender (masculine or feminine) and to the diminutive (v_3 in Figure 7) as follows:

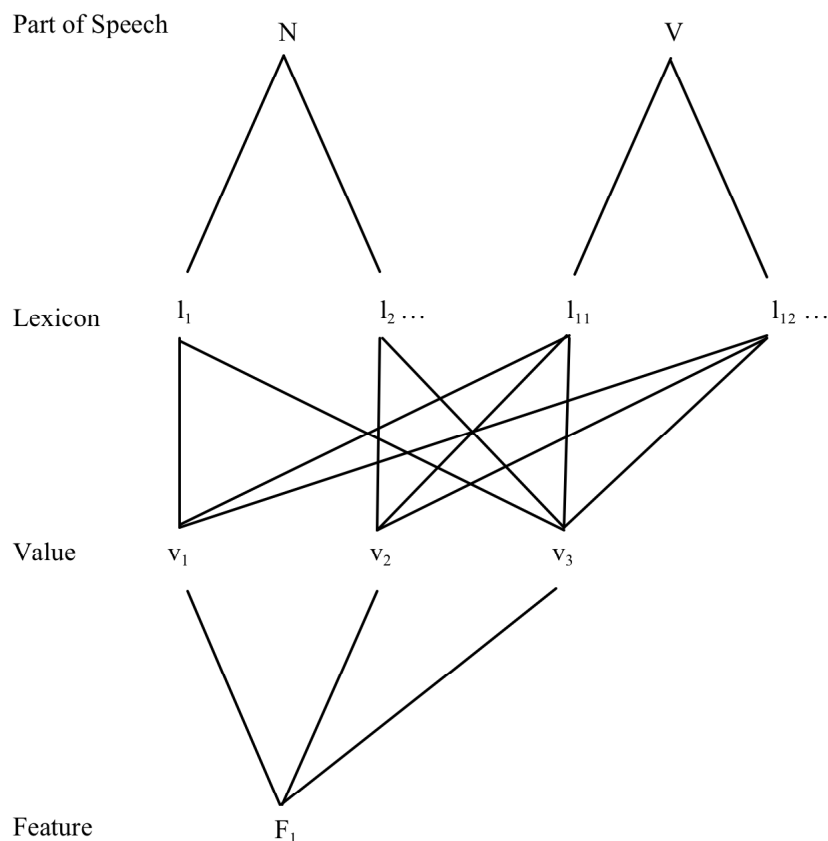


Figure 7: Minor target gender (Walman)

There are various types of non-canonical behaviour (see Corbett forthcoming for details). These two examples serve just to illustrate the point that the investigation of how we differentiate canonical features also clarifies some further aspects of morphosyntactic features.

1.6 Conclusion

We started from the problem of canonical morphosyntactic features. Earlier research had shown that there were various types of non-canonical behaviour, and that these were parallel, in a surprisingly regular fashion, across the different morphosyntactic features. But this left the puzzle that the morphosyntactic features that avoided these characteristics and so were fully canonical would then be indistinguishable. The way to distinguish them is through their interaction with canonical parts of speech. The canonical interaction was defined through four criteria, which proved interesting since the deviations from them helped to characterize several problems involved with feature systems. The main result,

however, is that number can in certain languages come very close to being canonical. Gender, person and case were defined in turn as showing deviations from the fully canonical situation. The approach also provided the means to discuss more deviant featural behaviour, illustrated with two further examples.

Morphosyntactic features are a difficult challenge for the typologist. The progress made is valuable for our understanding of features, and illustrates the usefulness of the canonical approach.

References

- Anward, Jan. 2000. A dynamic model of part-of-speech differentiation. In: Petra M. Vogel & Bernard Comrie (eds) *Approaches to the typology of word classes*, 3-45. Berlin: Mouton de Gruyter.
- Anward, Jan. 2001. Parts of speech. In: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds) *Language Typology and Language Universals I*, 726-35. Berlin: Walter de Gruyter.
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The Syntax-Morphology Interface: A study of syncretism*. Cambridge: Cambridge University Press.
- Baerman, Matthew & Greville G. Corbett. 2010. Introduction: defectiveness: typology and diachrony. In: Matthew Baerman, Greville G. Corbett & Dunstan Brown *Defective paradigms: missing forms and what they tell us* (Proceedings of the British Academy, 163), 1-18. Oxford: British Academy and Oxford University Press.
- Baker, Mark C. 2003. *Lexical Categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.
- 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 [accessed 19.1.2009.]
- Brown, Dunstan. 1998. Defining 'subgender': Virile and devirilized nouns in Polish. *Lingua* 104.187-233.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- 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.
- 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. 2007. Canonical typology, suppletion and possible words. *Language* 83.8-42.

- 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. Canonical features and their interaction with Part of Speech categories. Invited paper at the Workshop on Markedness and Underspecification in the Morphology and Semantics of Agreement (MUMSA), Harvard University, 29 February - 2 March 2008. [To appear in *Markedness in the Morphosemantics of ϕ -Features* (special issue of *Morphology*), edited by Jonathan Bobaljik, Uli Sauerland & Andrew Nevins.]
- Corbett, Greville G. & Richard J. Hayward. 1987. Gender and number in Bayso. *Lingua* 73.1-28.
- Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In: Petra M. Vogel & Bernard Comrie (eds) *Approaches to the typology of word classes*, 65-102. Berlin: Mouton de Gruyter.
- 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. 2006. Relationality and word class typology. Paper read at the Conference on Word Class Typology, Amsterdam, June 2006.
- 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.
- 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.
- Nikolaeva, Irina & Andrew Spencer. 2008. Nouns as Adjectives and Adjectives as Nouns. Ms.
- Noyer, Rolf. 1997. *Features, Positions and Affixes in Autonomous Morphological Structure*. New York: Garland. [Revised version of 1992 MIT Doctoral Dissertation.
- 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.
- Schachter, Paul. 1985. Parts-of-speech systems. In Timothy Shopen (ed.) *Language Typology and Syntactic Description, Vol. 1: Clause Structure*, 3-61. Cambridge: Cambridge University Press. [revised as Schachter & Shopen 2007.]
- Schachter, Paul & Timothy Shopen 2007. Parts-of-speech systems. In: Timothy Shopen (ed.) *Language Typology and Syntactic Description: Second edition: Volume I, Clause Structure*, 1-60. Cambridge: Cambridge University Press.
- Seifart, Frank. 2005. *The Structure and Use of Shape-based Noun Classes in Miraña (North West Amazon)*. PhD thesis, Radboud University, Nijmegen.

- Siemund, Peter. 2007. *Pronominal gender in English: A study of English varieties from a cross-linguistic perspective*. London: Routledge.
- Spencer, Andrew. 2005. Towards a typology of 'mixed categories'. In: C. Orhan Orgun & Peter Sells (eds) *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe* (Stanford Studies in Morphology and the Lexicon), 95-138. Stanford: CSLI Publications.
- Stump, Gregory T. 1998. Inflection. In: Andrew Spencer & Arnold M. Zwicky (eds) *The Handbook of Morphology*, 13-43. Oxford: Blackwell.
- 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. Heteroclisys 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.
- Trask, R. L. 1999. Parts of speech. In: Keith Brown & Jim Miller (eds) *Concise Encyclopedia of Grammatical Categories*, 278-284. Amsterdam: Elsevier.
- 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*, 613-647. Moscow: Jazyki slavjanskoj kul'tury.]