



HAL
open science

Canonical gender

Greville G. Corbett, Sebastian Fedden

► **To cite this version:**

Greville G. Corbett, Sebastian Fedden. Canonical gender. *Journal of Linguistics*, 2016, 52, pp.495-531.
10.1017/S0022226715000195 . hal-03622822

HAL Id: hal-03622822

<https://hal.science/hal-03622822>

Submitted on 29 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Canonical gender¹

GREVILLE G. CORBETT

University of Surrey

SEBASTIAN FEDDEN

University of Surrey & The University of Sydney

(Received 10 July 2014; revised 24 April 2015)

Nominal classification remains a fascinating topic but in order to make further progress we need greater clarity of definition and analysis. Taking a Canonical Typology approach, we use canonical gender as an ideal against which we can measure the actual gender systems we find in the languages of the world. Building on previous work on canonical morphosyntactic features, particularly on how they intersect with canonical parts of speech, we establish the distinctiveness of gender, reflected in the Canonical Gender Principle: IN A CANONICAL GENDER SYSTEM, EACH NOUN HAS A SINGLE GENDER VALUE. We develop three criteria associated with this principle, which together ensure that canonically a noun has exactly one gender value; we give examples of non-canoncity for each criterion, thus gradually building the typology. This is the essential groundwork for a comprehensive typology of nominal classification: the Canonical Typological approach allows us to tease apart clusterings of properties and to characterize individual properties with respect to a canonical ideal, rather than requiring us to treat the entire system as belonging to a single type. This approach is designed to facilitate comparisons of different noun classification systems across languages.

1. PURPOSE OF THE PAPER

The broad topic of nominal classification has long fascinated linguists across the range of the discipline, from psycholinguists to typologists. Further progress in this area requires greater clarity of definition and analysis, and we offer a way forward, concentrating on the specific question of gender. A major motivation for this definitional work is to lay the foundation for a full and explicit typology

[1] The support of the AHRC (grant: Combining Gender and Classifiers in Natural Language, grant AH/K003194/1) and of the European Research Council (grant ERC-2008-AdG-230268 *MORPHOLOGY*) is gratefully acknowledged. We wish to thank Jenny Audring, Matthew Baerman, Dunstan Brown, Hans-Olav Enger, Alexandra Grandison, Tania Paciaroni, Tatjana Pišković, Serge Sagna, and Anna Thornton for reading and commenting on an earlier version of this article. We also thank three anonymous *JL* referees for their careful reading and insightful suggestions, which have led to considerable improvements in the article. We are grateful to Penny Everson and Lisa Mack for help in preparing the manuscript. We follow the Leipzig Glossing Rules.

of nominal classification and to facilitate research on some specially interesting languages, namely those which have more than one system of classification. To see how we have come to this point, consider a significant paper by Dixon (1982).² In that paper, gender systems (noun classes in his terms) and classifiers are opposed, according to varied sets of criteria. There is a neat typological correlation: we find gender in languages that are agglutinating or inflecting, and classifiers in isolating languages. Dixon points out some problems, but the paper stands as a helpful summary of where we had reached.

Since then, the field has moved forward, both in terms of data and analysis, as can be seen in Audring's (2011) bibliography of work on gender. Classifiers prove to be more varied than once thought (see Aikhenvald 2000, Grinevald 2000 and Kilarski 2013 for examples). Significantly too, several languages have been identified with both a gender system and a classifier system, which undermines a straightforward typological generalization relating systems of classification to language type. An important contribution is Seifart's (2005) thesis on Bora-Miraña, and this for two reasons. First, Bora-Miraña stands mid-way between classical gender and classifiers: one could cite data to argue that Bora-Miraña has a gender system and other data to show that it has classifiers. Second, Seifart tackles the difficult issues head on, adopting a Canonical Typological approach.

Essentially then, a simple opposition between gender and classifiers no longer makes sense. There is a way in which converging criteria define a canonical point, a canonical gender system, which is comparable in useful ways to the real systems we find. The same cannot be said for classifier systems: they differ substantially, and we do not find systems which could be usefully analysed as genuine opposites of gender systems. This is for two reasons: first, many items labelled 'classifiers' share significant properties with gender; and second, the phenomena treated as classifiers do not form a coherent grouping. Our goal in this paper is to further the definition and description of gender systems, by providing a canonical ideal of gender. We need progress here if we are to create the possibility for meaningful comparisons across languages of different families and types. Part of the problem is that there are familiar gender systems which show peculiar combinations of elements. To understand and compare such systems we need to dissociate those elements, since the particular combinations we find in familiar languages are not necessary combinations, nor should they be definitional of gender systems. Indeed some of the most familiar gender languages, French and German, are particularly challenging. A full and explicit typology of nominal classification remains a goal

[2] It may help the reader to state briefly how we have come to write the paper. Corbett (1991: 136–137) found Dixon's (1982) paper particularly helpful. Since that point, seeing languages with more than one classification system has made him look for better typological definitions. Fedden's (2011) work on Mian meant that he was confronted with a language with two systems, a clear gender system and a (possible type of) classifier system. The convergence of these interests led to the current research. We are joint authors and the order of names is not significant.

for further research but we believe that the criteria for canonical gender which we develop here can be used to integrate many (if not all) types of classifier into the large possibility space projected by the canonical criteria we specify. Ultimately, then, we envisage a typology in which (traditional) gender and the various types of classifier can be treated together.

In [Section 2](#) we give an introduction to Canonical Typology, particularly its application to morphosyntactic features. Given this essential basis, in [Section 3](#) we home in on the feature gender and discuss what is special about it (drawing on earlier work by Corbett 2013): gender is a non-canonical morphosyntactic feature in that nouns are restricted to a single value. We then set out the various distinctive properties of gender systems. In order to account for their degree of (non-)canonicity, we use both existing principles for canonical agreement and canonical morphosyntactic features and an additional principle for canonical gender in particular. In [Section 4](#) and [Section 5](#) we discuss canonical and non-canonical agreement controllers and agreement targets, respectively. [Section 6](#) deals with the interaction of gender with number, particularly cases of mismatch between controller and target genders. In [Section 7](#) we move on to agreement domains and the non-canonical behaviour of lexical hybrids. A key part of understanding gender is how it is assigned. We tackle gender assignment in [Section 8](#), especially the non-canonical cases of hybrids and recategorization. In [Section 9](#) we present our conclusions.

2. CANONICAL TYPOLOGY

Typologists are naturally attracted to clusterings of properties. When we find phenomena which are problematic, such as difficult instances of the gender feature, these are often problematic in more than one way. We should investigate whether these clusterings of problems are meaningful, or are merely coincidental. One path towards that goal is to draw out the theoretical space so that the clusterings can be pulled apart: then we find that the space is rather larger than generally imagined. We can anchor this space by starting with the indisputable instances. We use these to establish the properties of a ‘canonical’ or ideal gender feature and its values, and then take this as the baseline from which we measure the actual examples we find.

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 fully this space of possibilities is populated with real instances. Canonical instances are those that match the canon: they are the clearest. Since they have to reach a logically determined standard, they are unlikely to be frequent. On the contrary, they are 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.

The canonical approach is justified by its usefulness. Naturally we want to argue that the proposals we make are principled and valid, but the real justification

of the canonical approach is in how useful it proves to be; the goals are to enable insightful typological work and to facilitate understanding across theoretical divides. The key ideas have been laid out in various places. Most recently, Brown & Chumakina (2013) offer an outline of Canonical Typology, followed by a varied set of applications of the approach by different researchers, mainly in the areas of morphology and syntax. A working bibliography of this body of research is available at www.smg.surrey.ac.uk/approaches/canonical-typology/bibliography/.

The canonical approach offers a practical point for our current concern, namely defining gender systems. Since the examples nearest to canonical are those which are the clearest, this means that when defining a canonical use of a term we should be able to assume that it covers the canonical core. In the ideal scenario, differences in use of terms reduce to the question 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. (Thus the Evening Star is certainly not a canonical star.) Within previous work in Canonical Typology there is relevant research, which will give us a good start; this research is on canonical agreement (Section 2.1) and canonical morphosyntactic features (Section 2.2).

2.1 *Canonical agreement*

It is generally accepted that gender is an agreement feature, since we can demonstrate the existence of gender in a language only by the evidence of agreement. Therefore work on canonical agreement is relevant here. Three overarching principles have been proposed (Corbett 2006: 8–27):

Canonical agreement – Principle I (information content)

Canonical agreement is redundant rather than informative.

Canonical agreement – Principle II (syntax)

Canonical agreement is syntactically simple.

Canonical agreement – Principle III (morphological realization)

The closer the expression of agreement is to canonical (i.e. affixal) inflectional morphology, the more canonical it is as agreement.

The detailed working out of these criteria can be found in Corbett (2006: 10–26). We give just a brief summary of canonical agreement here:

Criteria of canonical agreement

- controller: is present, has overt expression of features, and is consistent in the agreements it takes, its part of speech is not relevant
- target: has bound expression of agreement, obligatory marking, doubling the marking of the noun, marking is regular, alliterative, productive; the target has a single controller and its part of speech is not relevant
- domain: agreement is asymmetric (e.g. the gender of the adjective depends on that of the noun), local (i.e. within the NP), and there are multiple domains
- features: lexical, matching values, not offering any choice in values
- conditions: no conditions

The degree of canonicity of a gender system depends in part on the degree of canonicity of the agreement system through which it is expressed. In other words, a gender system is not fully canonical as long as its agreement system is not fully canonical. Canonical gender has redundant, overt and local agreement, but these are not directly properties of the morphosyntactic feature gender but rather of canonical agreement. Together with what we have for canonical agreement, we should now consider what would be canonical for morphosyntactic features generally.

2.2 Canonical morphosyntactic features

The motivation for postulating any morphosyntactic feature is that we wish to model the cross-cutting of lexical and grammatical meaning. Thus for English we propose a feature number, with the values singular and plural, on the basis of patterns like crocodile : crocodiles :: alligator : alligators. That is, our account of the difference between singular and plural is equally true for each of these two nouns. And for each one, if we can describe the lexical semantics of the noun independently of number, we can predict the meaning of its singular and its plural. In a canonical account, we push such a distinction to its logical end point. Thus the more evidence we find indicating that a morphosyntactic feature is indeed orthogonal to lexical meaning, the more canonical the morphosyntactic feature is. This means that a system in which number is marked on nouns, verbs and adjectives is more canonical than one in which it is marked only on nouns and verbs. And the more fully compositional the combination of lexical and grammatical meaning, the more canonical the feature. Observations of variability in given languages suggest our criteria, which we push to their logical extremes.

We now consider the abstract ideals which such an approach induces. To a great extent, canonicity of morphosyntactic features applies equally to gender, number, person and case. Since gender is a morphosyntactic feature, it inherits

the characteristics of a canonical morphosyntactic feature, some of which are overwritten by characteristics more appropriate to gender. We consider the shared characteristics first. Then we consider what is special about gender. For each of the features we need to be clear about whether we are considering the feature as a whole, for instance, number, or whether we are examining particular values (such as singular, plural or dual). However, the more canonical each value is, the more canonical the system of the given feature as a whole.

Canonical morphosyntactic features in general can be described in terms of two overarching principles; these principles cover ten converging criteria. Following Corbett (2012: 156–199), a canonical morphosyntactic feature can be defined according to these principles:

Canonical morphosyntactic features – Principle I (evidence from form)

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).

Canonical morphosyntactic features – Principle II (syntax)

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

Besides these two, for the realization of morphosyntactic features there is an additional general principle of inflection (Corbett 2007):

Canonical morphosyntactic features – Principle III (morphological realization)

Canonical morphosyntactic features and their values are expressed by canonical inflectional morphology.

It is evident – and encouraging – that Principle III is the same for both canonical agreement and canonical morphosyntactic features, and that Principle II is comparable (though in fact the criteria under Principle II differ somewhat in the two instances). Both include a syntactic criterion of obligatoriness (for morphosyntactic features Criterion 5 states that the use of canonical morphosyntactic features and their values is obligatory (Corbett 2012: 191–192), while for agreement Criterion 6 states that canonical agreement is obligatory (Corbett 2006: 14–15)).

3. GENDER AS A FEATURE

So far we seem to be making good progress towards defining gender as a canonical morphosyntactic feature. However, imagine that all morphosyntactic features were fully canonical. They would be formally well distinguished, subject to simple rules of syntax and realized by canonical resources of morphology. They would be elegant, but identical: there would be no means of telling them apart, a

situation that does not fulfil our typological needs.³ We need therefore to consider how canonical gender is a NON-canonical morphosyntactic feature to establish how it differs from other morphosyntactic features.

3.1 *Differentiating the morphosyntactic features*

The strategy which proves successful is to examine the interaction of morphosyntactic features with canonical parts of speech. We can deal with canonical parts of speech briefly here; see Corbett (2013) and Spencer (2005: 102) for more detail. Canonical parts of speech are those for which the semantics, syntax and morphology align. 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 way in which they interact, canonically, with morphosyntactic features is diagrammed in Figure 1. For simplicity we consider a small system, comprising two parts of speech, with just two lexemes included for each, and having two features each with two values. This schema rests on canonical criteria, which we must make explicit. It is the deviations from these criteria which give us the means to differentiate the features, and in particular the gender feature. The basic idea is that morphosyntactic features and parts of speech cross-cut, and the more fully they do so the more canonical their relation:

Canonical morphosyntactic features – Principle IV (interaction with parts of speech)

Canonical morphosyntactic features and canonical parts of speech are fully orthogonal.

This is a general principle; we spell out the implications in four criteria:

Canonical morphosyntactic features – Principle IV: 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.

Canonical morphosyntactic features – Principle IV: Criterion 2: EXHAUSTIVENESS

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

[3] While the features would be realized by different formal means, there is no general way of predicting WHICH feature is being realized, just from knowing the forms. Or else the features could have different semantics; however, this too is not a foolproof predictor, since we find, for instance, information about countability being used in gender assignment rules (Corbett 2013: 50). The spirit of our approach is to seek logical criteria, and the strategy below conforms to this.

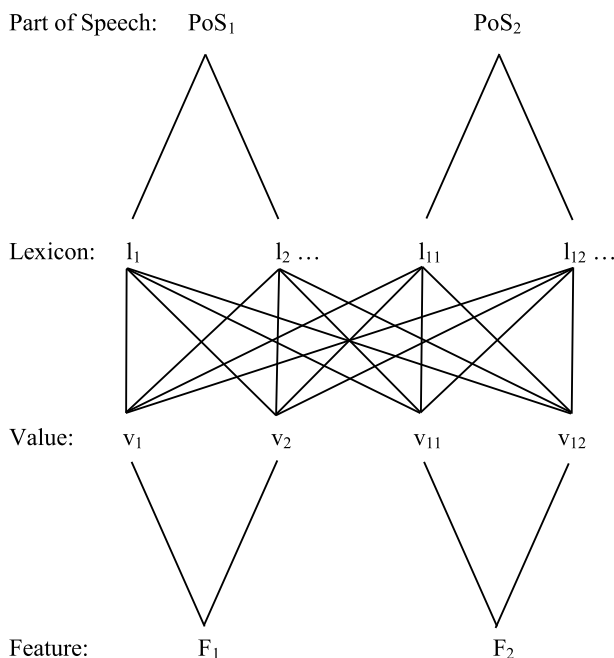


Figure 1

Canonical parts of speech and canonical morphosyntactic features.

In the canonical instance, it is evident why we postulate a morphosyntactic feature, since it generalizes across a large number of items. 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. At the logical extreme, if we have a number feature with the values singular and plural, then in the canonical situation every lexeme would have singular and plural available to it. The notion of exhaustiveness would apply across the board to all features (and their values) and to all parts of speech (and the lexemes belonging to them).

Canonical morphosyntactic features – Principle IV:

Criterion 3: OPEN AND CLOSED CLASSES

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

This idealization makes sense of the distinction between part of speech and morphosyntactic feature. We have four classes under discussion: features, values, parts of speech and lexical items. Of these, in the canonical instance, all are closed classes, except the class of lexical items. This criterion falls under Principle IV, because the fact that the sets of features and values are closed makes it more

likely – owing to the finite number of combinations – that we can find and demonstrate orthogonality, as is required in the canonical case.

Canonical morphosyntactic features – Principle IV:

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.

A key motivation for separating out features and their values is the regularity involved. If we had to specify the lexical semantics of *crocodiles* quite separately from *crocodile*, and with no relation to pairings like *alligators* and *alligator*, the attraction of the feature-based analysis would be considerably reduced. We now turn to the key issue of differentiating features, particularly gender, according to this typology.

3.2 Gender as a non-canonical morphosyntactic feature

According to the exhaustiveness criterion (Criterion 2), in the canonical situation every lexical item of every part of speech should have available all values of all features. What if, however, a given part of speech has access to all values of a feature, but the individual lexical items within that part of speech do not? This deviation, we suggest, is exactly what is found in a gender system, where canonically the part of speech NOUN as a whole has access to all values but where each noun has access to a single value of the feature only. This deviation from the canonical is represented in [Figure 2](#).

As we are concerned with only one feature (F_1), we grey out F_2 and all links between its values and lexical items. The deviation from canonicity involves its relation to the first part of speech (PoS_1). 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 numerous lexical items in the relevant part of speech. We can see this as a representation of a gender system: a typical situation would be that the first part of speech (PoS_1 in [Figure 2](#)) is noun, and the second part of speech (PoS_2) is adjective.

Thus gender is a non-canonical morphosyntactic feature in having one value available per controller (in canonical instances). This leads us to the Canonical Gender Principle:

Canonical Gender Principle

In a canonical gender system, each noun has a single gender value.

This point is made in various places, notably in Dixon (1982: 166), but bear in mind that Dixon was concerned with actual systems rather than canonical systems. The Principle as stated requires that each noun has a value (leaving

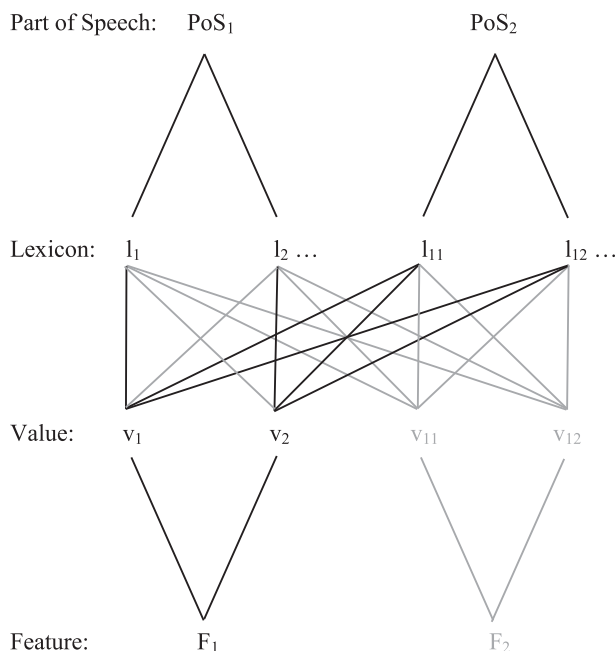


Figure 2
Canonical schema: key deviation.

no noun outside the system) and that each has only one value. There is an interesting difference when we compare gender with number. For gender, nouns may exceptionally have more than one value (and we have special terms for them, discussed below in [Section 4.1](#)). For number the situation is the opposite, namely that we have special terms for nouns with only one number value, such as *pluralia tantum*.

The point about nouns having a single gender value links to the observation that gender relies on canonical agreement, where the canonical feature type is lexical (Corbett 2006: 23–24). In this article, we take a ‘lexical feature’ to be tied to the lexical entry and to be invariant (though not necessarily arbitrary). Such a feature is considered canonical for agreement because it demonstrates the existence of agreement: only by agreement is its realization on other elements to be explained. If the values of a feature could be derived entirely by semantic information (as is imaginable for a fully regular number system), then it is harder to demonstrate the need for a syntactic rule of agreement. This is the case for example in Tamil, where gender assignment is strictly semantic, i.e. male humans and gods are masculine, female humans and goddesses are feminine and everything else is neuter (Asher 1985: 136–137; Corbett 1991: 8–10). In such a case, it might be possible to argue that repeated accessing of the properties of the referent would be an alternative analysis. Since in the case of canonical agreement we are dealing with a lexical

feature and its values, this leads to the prediction that there will be one value per controller (from a given lexical entry, just one feature value is computed). This gives the nice conclusion that gender is not a canonical morphosyntactic feature, since controllers do not have all values available, but gender is the canonical agreement feature, for the same reason, since it is a lexical feature. (We take this further in [Section 8.1](#) below.) We now move from the discussion of gender as a non-canonical morphosyntactic feature to an account of canonical gender.

3.3 *Establishing gender values: The agreement class approach*

There is a recognized analytical technique for establishing the number of gender values in a given language. It is based on AGREEMENT CLASSES, which are set up on syntactic evidence. The approach can be traced back to Zaliznjak (1964); the account here is from Corbett (2012: 80–85); recent discussion can be found in Mel'čuk (2013). Note that agreement classes are the first step to determining potential gender values, but further analysis is required to determine which agreement classes are to be recognized as gender values. In the canonical world, that is, in our idealized, fully canonical language, the following criterion holds:

Canonical Gender – Criterion 1

Canonical gender values match agreement classes.

Any mismatch between gender values and agreement classes is a source of non-canonicity. The notion is that each gender value has an associated set of agreements, which match up to ensure that each noun has a single gender value. We define an agreement class as follows:

Definition of agreement class

A set of nouns is an agreement class if and only if all members of that set have the property that whenever

(i) they have the same morphosyntactic specification

and

(ii) they occur in the same agreement domain

and

(iii) they have the same lexical item as agreement target

then their agreement targets have the same morphological realization.⁴

The basic idea of the definition is that nouns are in the same agreement class provided that given the same conditions they will control the same agreement form. That is, however we alter their environment, nouns of the same agreement

[4] Note that there can be orthogonal shape conditions. Thus vowel-initial feminine nouns in French consistently select a form of the possessive pronoun homophonous with the masculine, provided the two are adjacent. Compare *sa mère* 'his mother(F)' and *son père* 'his father(M)' with *son enfance* 'his childhood(F)'; see, for instance, Pullum & Zwicky (1988: 263) for these conditions.

class will induce the same agreements. The three clauses of the definition spell this out.

The effect of clause (i) above, having ‘the same morphosyntactic specification’, is that the nouns have the same values for all relevant morphosyntactic features. The features usually involved are case and number; gender is excluded, since that is what we are defining. We rely on these notions being given, since they are simpler notions, though not trivial (interestingly, Meyer 1994: 360 argues that to model a particular feature in a given language we need full knowledge of all the others). Identity of morphosyntactic specification does not imply morphological identity. Nouns may have the same morphosyntactic specification and yet differ morphologically (they may belong to different inflection classes). Conversely, different morphosyntactic specifications may have a single morphological realization; this is the situation known as syncretism. The intuitive content of clause (i) is simply that the nouns start out ‘on level terms’, in the relevant respect.

Clause (ii) requires that the nouns occur in the same agreement domain. This means that the construction in which agreement applies must be identical in each instance: it might be the agreement of modifiers with the head of a noun phrase, subject–verb agreement, and so on. Put simply, the two nouns must be put in the same syntactic setting.

Clause (iii) requires that the lexical item which functions as the target must be the same. This is because lexical items can differ in their agreement possibilities. Some may not realize agreement at all, and others may distinguish different numbers of gender values, and so we need to hold this variable constant.

In all instances we are interested in those agreement domains and those lexical items which allow the largest number of forms. We ensure this effect by specifying that identity must be found ‘whenever’ the conditions listed are met. That is, we are looking for the domain which is most favourable to gender agreement and for the most differentiated agreement target. Given these conditions, if the same result follows, the nouns must be in the same agreement class.

A key point is that establishing the agreement classes is the first step in the analysis of gender systems. Typically we do not recognize all the agreement classes as gender values. In the canonical world, however, each agreement class is a gender value; the agreement class analysis would be the full account. One part of the reason is that all agreement domains have exactly the same possibilities in a canonical system, so clause (ii) has no differentiating effect. Moreover all lexical items have identical agreement possibilities, hence there is no differentiation according to targets, so clause (iii) has no effect (see Criterion 4 in Section 5.2 below). Once again the phenomena which linguists have chosen to name, such as common nouns, hybrid nouns and sub-genders, can be understood as deviations from the canonical ideal. Our interest is precisely in the differences between real world gender systems and the canonical ideal. We discuss these in turn, starting with controllers (Section 4), moving on to targets (Section 5) and the dependency of gender on other features (Section 6). We discuss domains next (Section 7). Then we return to the key question of how gender values are assigned to nouns (Section 8).

4. CONTROLLERS

The distinctive behaviour of gender among the morphosyntactic features is that the controller has a single value available. This is encapsulated in the Canonical Gender Principle, repeated from [Section 3.3](#):

Canonical Gender Principle

In a canonical gender system, each noun has a single gender value.

The closer that a system approximates to this situation, the more canonical it is. There are two main types of non-canonical behaviour of particular controllers, which we discuss in turn.

4.1 *Specific featural problems*

There are controllers, typically nouns, which are obvious non-canonical types, and their special nature is reflected in the fact that we have terms for them. ‘Common gender’ is a label for items which are arguably single lexical items, yet which have more than one gender value. For instance, English *sibling* may, in appropriate circumstances, be masculine or feminine (since we accept the notion of languages with pronominal gender). The same is true of Mian *aban* ‘orphan’ (Fedden 2011: 170). Conversely, ‘epicene’ nouns are those where more than one gender value might have been expected (given the rest of the particular system) but only one is found. More specifically, there is one item, with a single gender value, where elsewhere we find paired lexical items.⁵ For instance, in Serbo-Croat we find *krava* ‘cow’ (feminine) and *bik* ‘bull’ (masculine), and in contrast there are epicene nouns *krokodil* ‘crocodile’, which is masculine, while *žirafa* ‘giraffe’ is feminine (Pišković 2011: 161). In Mian the contrast is only in agreement: *éil* ‘pig’ takes masculine agreement for the male and feminine for the female (Fedden 2011: 170), while the epicene *koból* ‘cassowary (a flightless bird)’ takes feminine agreement, irrespective of sex (Fedden 2011: 171).

4.2 *Consistency across other features*

One of the criteria for any canonical morphosyntactic feature concerns its relation to the other features. Just as grammatical meaning is orthogonal to lexical meaning so also, in the canonical instance, it is orthogonal to other grammatical meanings. This is true in general of the feature system (see [Section 6](#) below). But there are also non-canonical instances involving particular controllers. A clear case is Serbo-Croat *oko* ‘eye’, which is non-canonical in terms of its gender

[5] Genuine epicene nouns denoting humans are hard to find: typically those cited in the literature have different possible agreements, at least for the personal pronoun, and so are hybrids in our terms (see [Sections 7](#) and [8.2](#)).

specification. The basic gender system of Serbo-Croat has three gender values, controlling these agreements (Table 1).

GENDER	SINGULAR	PLURAL
MASCULINE	njezin	njezin-i
FEMININE	njezin-a	njezin-e
NEUTER	njezin-o	njezin-a

Table 1

Gender agreement in Serbo-Croat (the possessive *njezin* 'her', nominative forms only).

While this is the general system, *oko* 'eye' does not fit. Consider examples (1) and (2):

- (1) njezin-o ok-o
her-SG.N.NOM eye-SG.NOM
'her eye'
- (2) njezin-e oč-i
her-PL.F.NOM eye-PL.NOM
'her eyes'

Recall that in the agreement class approach, the requirements include the following: for nouns to be in the same agreement class, provided they have the same morphosyntactic specification (clause (i) above), then their agreement targets must have the same morphological realization. Applying this criterion puts *oko* 'eye' in a different agreement class from neuter nouns (since its agreements in the plural are not the same as those for neuter nouns). It is also in a different agreement class from feminines (since its agreements in the singular are different). There is just one other noun which behaves in the same way, namely *uho* 'ear'. We could set up an additional gender value, but it would be just for these two nouns. An alternative (and our preferred option) is to treat the agreement class as 'inquate' and label the two nouns as lexical exceptions, with neuter gender in the singular and feminine in the plural. This is a simpler solution because it minimizes the number of values (hence the gender system is simpler). We need the values neuter and feminine anyway for the large number of nouns which are of either of these genders. This analysis also shows clearly the non-canonical situation of the two nouns with respect to the Canonical Gender Principle: these are nouns which in a sense have two gender values. There is no impact on the rest of the system (the agreement forms are available). This locates the irregularity where it belongs, namely on the two lexical items. Furthermore, they have an irregular split in their

paradigm, which correlates with their gender irregularity (Corbett 2015a: 167)⁶. It is important to bear in mind that we are concerned here just with the non-canonical behaviour attributable to individual controllers. Instances where there is an issue concerning the feature system more generally are considered in Section 6 below.

5. TARGETS

There are three sets of issues: the question of how gender is realized (Section 5.1), the coverage of the system (Section 5.2), and the matching or mismatching of values between controller and target (Section 5.3).

5.1 Exponence

As we would expect, canonical gender is realized through agreement by canonical inflectional morphology, which is affixal. We have already met the relevant principle in Section 2.1. Typologically, most gender agreement is affixal, but there are rare cases of non-concatenative gender agreement, which are thereby non-canonical. An example is the Papuan language Marind, where some adjectives show ablaut agreement forms, e.g. *ak-k* ‘light’ has *akek* for gender I, *akuk* for II, *akak* for III, and *akik* for IV (Drabbe 1955: 18–20). Related to this, canonical gender is expressed by bound forms, but there is some evidence that gender systems can be non-canonical in that they allow the expression of agreement to be accomplished by a free form. The Daly languages of north Australia are potential examples of having free words as agreement markers (Corbett 2006: 13–14). For instance, Ngan’gityemmerri has 15 gender values (Reid 1997), most of which are realized through bound agreement, but six of which have optional generic nouns, which appear as free words. An example is given in (3) (Reid 1997: 177):

- (3) (syiri) magulfu (syiri) marrgu
 STRIKE cylindrical.fighting.stick STRIKE new
 dem-wurity-dim
 3SG.SBJ.AUX-make-3SG.SBJ.sit
 ‘He is making a new cylindrical fighting stick.’

The free form *syiri* can be employed with expressions referring to weapons used for striking. It is the second instance of *syiri* in (3) which is relevant: appearing in front of the modifier *marrgu* ‘new’ it looks like the expression of agreement. This is no doubt a borderline phenomenon: the forms are free and they are optional. Yet their function appears to parallel that of the bound agreement forms. While canonical gender relies on bound forms, there are rare examples, like the Daly

[6] See the same source for another instance of a controller irregularity which is a matter of an individual lexeme. The noun *muir* ‘sea’ in some Gaelic dialects has different gender values according to case.

languages, in which free words can conceivably be analysed as expressions of agreement.

5.2 *The coverage of the gender system*

There is a simple intuition that the more evidence there is for gender (other things being equal), the more canonical the system (see e.g. Audring 2014). That is, the more instances of agreement in gender, the more canonical the gender system. This intuition applies equally to other features, and is covered by two existing criteria for morphosyntactic features (Corbett 2012: 162–163). These are:

Canonical morphosyntactic features – Principle I: Criterion 3

Canonical features and their values are distinguished consistently across relevant parts of speech (word classes).

Canonical morphosyntactic features – Principle I: Criterion 4

Canonical features and their values are distinguished consistently across lexemes within relevant parts of speech.

Clearly the more targets that mark gender, the more canonical the system, and the two criteria given have that effect. They ensure that all nouns are allotted to a gender value. Recall too from Section 2.2 that the use of canonical morphosyntactic features and their values is obligatory, including in agreement. This ensures that the gender value of nouns is evident in use.

5.3 *Inequalities of values*

Controller genders are set up based on the agreements of a noun independent of other features. Target genders are based on the number of distinctions any given target makes. Target genders can vary depending on other features, typically number. In the canonical situation the number of controller and target genders is the same. The two criteria in Section 5.2 go part way towards this requirement. In the canonical world, if lexemes of one part of speech, say verbs, distinguish number, then so do lexemes of other parts of speech. Moreover, the number of values distinguished will be the same. However, there is a complication for gender; since in the canonical world nouns have only one gender value (compare the Canonical Gender Principle in Section 4 above), this means that ‘distinguished consistently’ is not so straightforward as for other features. First, at the feature level, we can say that for controllers to have gender at all there must be evidence from targets. Second, at the value level, since canonically nouns have a single gender value, it is the number of values distributed over nouns as a part of speech which is to be compared with values of targets. We need a new principle here, but it is one which turns out to have useful application beyond gender:

Canonical morphosyntactic features – Principle V

In a canonical system of morphosyntactic features the contextual values match the inherent values.

Put briefly, inherent features are those found where their grammatical meaning is relevant (for instance, number on nouns), while contextual features are those realized ‘in the wrong place’ as a result of agreement or government (for instance, nominal number on verbs); compare Booij (1996) and Corbett (2012: 66–68). Principle V requires that the systems match, and specifically for gender if there are a given number of gender values available to nouns (the inherent values), this is the same number as those available for agreement targets (the contextual values). This principle also applies to canonical number: the values of nouns should match those available to agreement targets. Systems like that in Modern Hebrew, where some nouns have a singular-dual-plural system but agreement targets treat dual and plural alike as plural, are non-canonical in this respect (see Corbett 2000: 95–96 for the data).

In terms of inequality of values there are three non-canonical possibilities here, which we discuss in turn. First, it is possible that the source of non-canonical behaviour is not part of the system but rather a lexical property of a small number of targets (Section 5.3.1). Second, the non-canonical behaviour can be systematic and all targets show more values under certain conditions (Section 5.3.2). Third, there can be inequalities when we compare the number of values of a controller with different targets rather than with a single target (Section 5.3.3).

5.3.1 *Lexical issues*

We begin with the situation where there is an inequality of values, and it arises in extremely few targets. A natural analysis would treat this as a question of the lexical properties of the target rather than as a question of the system. This is the parallel to what we have said for controllers in Section 4.2. Two types can be distinguished: (i) the target may distinguish more contextual values than the controller has inherent values, or (ii) the target may distinguish fewer contextual values than the controller has inherent values.

The first type can be found in the Dravidian language Kolami, in which typical targets distinguish two genders, male human versus the rest. However, low numerals (‘two’, ‘three’ and ‘four’) show an additional value, namely human female (Emeneau 1955: 56, cited in Corbett 1991: 168):

- (4) iddar ma:sur
two.M.HUMAN men
‘two men’
- (5) i:ral pillakul
two.F.HUMAN women
‘two women’

- (6) *indij sidl*
two.N buffaloes
'two buffaloes'

According to one analysis the low numerals are 'overdifferentiated', in that they split the regular gender value 'non-male human' into female human and neuter. In a canonical system the different agreement targets would behave alike; in this non-canonical system, only a small fraction of the targets make a three-way gender distinction, while the remainder of them show syncretism of female human and neuter. It is just the low numerals which, contrary to the general syncretism, are lexically specified as differentiating all three gender values.

The converse type is Russian, where the numeral *dva* 'two' makes a two-way distinction in a three-valued gender system (it has an unusual syncretism). *Dva* is used for masculine and neuter nouns, while the form *dve* is used for feminines. The idiosyncratic nature of these inequalities of values has led to Kolami being analysed as having two gender values (which is what most targets distinguish) and Russian as having three, despite the fact that in each language there is a small number of targets which distinguish more or fewer values than the majority of targets.

5.3.2 Systematic inequalities between controller and targets

We continue with inequalities of inherent and contextual values, now looking at instances where the inequality does not originate in the idiosyncratic behaviour of individual lexemes but affects the gender system more globally. There are languages where a certain type of target in certain constructions makes more distinctions than the typical target in the language. In Ukrainian there is a unique neutral form in predicative adjectives (Shevelov 1963: 128–133).

- (7) *V odnij simji nam žyty i lehk-o i*
in one family 1PL.DAT live.INF and easy-NEUTRAL and
prekrasn-o.
wonderful-NEUTRAL
'For us to live in one family is both easy and wonderful.'

The adjectives *lehko* 'easy' and *prekrasno* 'wonderful' show neutral agreement in *-o*, the form appropriate when there is no subject headed by a normal noun or pronoun. Agreement in neuter gender would be in *-e*. (The copula 'to be' has the null form in the present tense.)

It is worth mentioning why we are treating this type of inequality between controllers and targets here as a target issue. We cannot systematically have fewer genders in the targets than in controllers because there would be no evidence: there

could not be a controller gender without evidence in the targets.⁷ The reverse is possible: targets can make more distinctions than there are controller genders, as we have seen in Ukrainian. This is non-canonical, according to Principle V.

5.3.3 *Comparisons involving additional targets*

Suppose now that instead of taking the targets one by one, we combine controllers with more than one target in a given construction. Of course, in the canonical world, adding a second or further targets to the comparison would make no difference, since the values would be distinguished consistently across all targets, following the criteria discussed in Section 5.2. There are various non-canonical situations, which vary in interest according to how different the possibilities of the various targets are. Imagine a language in which verbal predicates distinguish three gender values, say masculine, feminine and neuter, but the adjective modifiers distinguish only two. Suppose further that the adjectives pick out the same neuter nouns as do the verbs, and the masculines and feminines fall together. Clearly we shall have three agreement classes, on the basis of the evidence from the verbs, the adjectives add nothing to this, and we have arguably an instance of systematic syncretism.

More interesting is the situation in which a small part of the system gives rise to an additional distinction. The often-discussed issue of sub-genders fits here (Corbett 2012: 86–88). Thus in Russian there is ample evidence for three genders, from verbs, adjectives and pronouns. However, just within targets which mark case, within the accusative case, there is evidence for a split between animates and inanimates. This animacy distinction fully cross-cuts the main genders (including the neuter). When we apply the agreement class approach, it is evident that we have six agreement classes. Since the animacy distinction is restricted to just one case value, it is often treated as a sub-gender, that is, as a less important distinction, subordinate to the main genders.

Sub-genders bring us naturally to the most interesting systems, namely those in which the targets are so different as to make us ask whether we have one system or two. There are few systems for which we have the level of information we would like. One of these is Mba (for sources and analysis see Corbett 2011: 465–466), and another is Michif (Bakker 1997: 106–107 discussed in Corbett 2006: 269–270). A fascinating example is Burmeso (Donohue 2001), a language of Western New Guinea. The verb distinguishes six gender values, and the adjective distinguishes six different values. If these two systems were independent of each other, we would expect 36 possibilities. However, only 16 combinations are attested, which shows clearly that it is not a matter of two systems combining freely. For discussion of this aspect of Burmeso, see Corbett (2012: 176–180).

[7] Some analyses of Bantu languages suggest that this could occur, for instance when classes such as 1a and 1b are distinguished. But these indicate different inflectional possibilities, not different gender values.

It is worth noting that in this section we have been dealing with the possibilities available to targets of different types, which is a matter of their inflectional morphology. The USE of these possibilities, controlled by particular nouns, is a matter of the domains of agreement, and is discussed in [Section 7](#).

6. FEATURES

In the canonical world, features cross-cut parts of speech (see [Section 3.1](#) above). And equally they cross-cut each other. That is to say, the distribution of gender is not dependent on, say, person, nor is number dependent on case, and so on. This is covered by the following criterion (Corbett 2012: 158):

Canonical morphosyntactic features – Principle I: Criterion 2

Canonical features and their values are uniquely distinguished across the other logically compatible features and their values.

Note that here we are considering features and their values as a whole, particularly gender in relation to other features, rather than particular problematic controllers (which we discussed in [Section 4.2](#) above). The interaction that is most often problematic for gender is that with number. In the canonical situation, each gender value in each number value matches the gender value in every other number value. In other words, gender could be described quite separately from number. In graphic terms, the gender system is then ‘parallel’ (Corbett 1991: 155, and compare Heine 1982: 196–197). A parallel system can be illustrated from Italian, using the forms of the adjective *nuovo* ‘new’ in [Figure 3](#).

SINGULAR		PLURAL
nuovo	————— M —————	nuovi
nuova	————— F —————	nuove

Figure 3
Gender in Italian.

There are, of course, various deviations from this canonical picture. A mild deviation is regular syncretism, as in German or Russian. German has three controller genders and also three target genders but the only evidence we have for them comes from the singular, since in the plural all gender contrasts are neutralized (Corbett 1991: 155). A more complex relationship between gender and number is found in Romanian. Here the effect of number gives a system with three controller genders but only two target genders. The controller genders are masculine, feminine and neuter or ‘ambigenic’, where there is syncretism with the masculine form in the singular but syncretism with the feminine form

in the plural. The neuter gender is non-autonomous (Zaliznjak 1973[2002]: 69–74); neuter nouns do not have any agreement forms which are used uniquely for them, and this leads to a mismatch, induced by number, between controller and target genders. For a more detailed account of the Romanian gender system, see Corbett (1991: 150–152), and for a recent analysis see Loporcaro (forthcoming). The alternating neuter gender in parts of Romance is a similar phenomenon (Loporcaro & Paciaroni 2011).

Below we discuss Archi and Mian, two interesting cases of deviation from Criterion 2 for canonical morphosyntactic features, which states that canonical features and their values are uniquely distinguished across the other logically compatible features and their values.

6.1 Archi

The gender and number agreement system of the Daghestanian language Archi may be represented as in Table 2 (*x*- is the prefixal form, and <*x*> the infixal form); the original source is Kibrik et al. (1977: 55–66). There is a four-way distinction of gender values in the singular, collapsing to two in the plural; this is a convergent system. It is clearly non-canonical in that the values of gender are not uniquely distinguished across number. The deviation is fairly minor, in that given the gender value in the singular, the plural agreement is predictable (with few exceptions). On the basis of the verbal forms in Table 2, it appears that there is a more serious issue, in the syncretism of forms; gender III singular matches I and II in the plural, while IV singular matches III and IV in the plural. This interesting syncretism holds for most parts of speech; however, adjectives have one shared form for the plural, distinct from the singular forms, thus separating the syncretic forms. When we turn to Mian below, this issue will return with no distinguishing forms.

GENDER (controller gender values and assignment)	NUMBER	
	SINGULAR	PLURAL
I (male human)	w-/<w>	b-/
II (female human)	d-/<r>	
III (some animates, all insects, some inanimates)	b-/	∅-/<∅>
IV (some animates, some inanimates, abstracts)	∅-/<∅>	

Table 2
Gender and number in Archi (verbal agreement).

Before leaving Archi it is worth considering the gender assignment system briefly. The assignment of gender values I and II is straightforward, being semantic in nature. For III and IV, while there are clear overlaps with semantic

categories, the assignment is not straightforward. Given an inanimate object of unremarkable size there is (as yet) no clear prediction of its gender. Typologically, genders I and II resemble those of a Dravidian language, while III and IV are more like those of an Indo-European language; this is a helpful reminder that gender systems do not necessarily come all of a piece.

6.2 *Mian*

The Ok language *Mian* (Trans New Guinea; Fedden 2011), spoken in Papua New Guinea, also shows an interesting interaction of gender with number. We establish four controller gender values for *Mian* (Table 3) based on the agreements found on the clitic article. The other targets, i.e. the demonstratives and the verb, show different agreement forms but follow exactly the same pattern. A rough characterization of the semantics underlying the assignment is also given in Table 3.

GENDER (controller gender values and assignment)	NUMBER	
	SINGULAR	PLURAL
MASCULINE (males)	= <i>e</i>	= <i>i</i>
FEMININE (females)	= <i>o</i>	= <i>i</i>
NEUTER 1 (inanimates)	= <i>e</i>	= <i>o</i>
NEUTER 2 (specific set of inanimates: locations, body decoration, weather phenomena, illnesses, abstract nouns, some tools and weapons)	= <i>o</i>	= <i>o</i>

Table 3
Controller genders in *Mian*.

We are justified in setting up these controller genders because, as in Romanian, there is a sizable number of nouns in each of them. However, note that the agreements show a striking pattern of syncretism. The target genders distinguish just =*e* and =*o* in the singular and =*i* and =*o* in the plural. Looking at the *Mian* system, it combines the non-canonical behaviour of Romanian and that of *Archi*. First, all *Mian* genders are non-autonomous values (Zaliznjak 1973[2002]: 69–74), they have no agreement forms which are unique to them. Second, there are interesting syncretisms of gender across number, but unlike in *Archi* there is no part of speech where these are differentiated. In fact they constitute partial polarity (see Baerman, Brown & Corbett 2005: 103–111), since stating these syncretisms requires reference to two features (gender and number). This is easier to see graphically; the relation between controller genders and target genders is illustrated in Figure 4. Thus *Mian* has a system with four controller genders

but only two target genders. This arises from the non-canonical interaction with number.

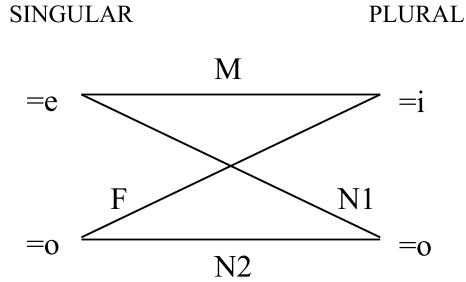


Figure 4
Controller and target genders in Mian.

7. DOMAINS

Agreement in gender, like agreement in general, can manifest itself over a range of domains (Corbett 2006: 21). The closest domain is within the phrase, for example when an article or an attributive adjective shows agreement with a head noun. Within the clause we find agreement of the predicate with one of its arguments. Relative pronouns can show agreement beyond the clause but within the sentence and, finally, free pronouns agree anaphorically within or across sentence boundaries.

Recall from Section 5.2 that the more evidence there is for gender (other things being equal), the more canonical the system, hence the more targets which agree in gender in a language the better. A similar point can be made about the number of domains: the more domains the better. However, there is a more important requirement of domains. In the canonical world, domains do not simply have the same possibilities available (as ensured by the criteria requiring that the different parts of speech have the same values). Given the gender values available to them, different domains must be consistent in their use for a given noun. We can state this as follows:

Canonical Gender – Criterion 2

In a canonical gender system the gender of a noun is constant across all domains in which a given language shows agreement.⁸

This means that there would be no change in gender agreement depending on the agreement domain. For example, the Italian noun *vescovo* ‘bishop’ is masculine

[8] To put it another way, the ‘consistent agreement patterns’ (Corbett 2007: 250–253) are the only agreement patterns in the canonical world.

and we find that the article, the predicate and the relative pronoun agree in masculine gender (from Maiden & Robustelli 2007: 134):

- (8) Il vescovo, in vece de-l quale io avevo parlato,
 ART.M bishop in stead of-ART.M whom.SG I had spoken
 era stato detenuto a Siena.
 was been.M detained.M in Siena
 ‘The bishop, in whose stead I had spoken, had been detained in Siena.’

Any anaphoric pronoun used later in the discourse and referring back to *il vescovo* ‘the bishop’ would also be masculine.

Some of the most interesting nouns for gender studies are those which behave non-canonically with respect to the domains criterion. These are lexical hybrids, that is, nouns which take different agreements depending on the domain. A hybrid (in terms of gender) is a noun where conflicting gender assignment has the effect that agreements differ in different domains, and some targets may allow alternative agreements even in the same domain. (This contrasts with the normal situation where a given noun has a particular gender value irrespective of syntactic domain.) For example, the German noun *Mädchen* ‘girl’ gives the appearance of being neuter, like all other German nouns ending in *-chen*. Targets inside the noun phrase, such as attributive adjectives, articles and demonstratives obligatorily show neuter agreement: *das Mädchen* [the.N girl]. However, *Mädchen* ‘girl’ is not a straightforward neuter noun: for the anaphoric pronoun either neuter or feminine agreement is possible. This is shown in (9):

- (9) Weiß dieses Mädchen überhaupt, was sie/es da getan hat?
 knows this.N girl at.all what she/it there done has
 ‘Does this girl know at all what she’s done there?’

Well-known examples of lexical hybrids in other languages are titles like French *Sainteté* ‘Holiness’ or Italian *Maestà* ‘Majesty’, which are grammatically feminine but allow masculine agreement in the free pronoun. For a recent survey of hybrids in Croatian, see Pišković (2011: 136–146). For examples outside Indo-European, see Corbett (1991: 229–231). Note that hybrids do not depend on additional target values; rather they are based on inconsistencies of agreement in different domains.

Hybrid nouns typically arise when the gender assignment rules of the language are in conflict. For German, the conflict would be that nouns denoting females should be feminine while nouns ending in *-chen* should be neuter (we take this further in Section 8.2). These nouns are clearly non-canonical with regard to the Canonical Gender Principle since they are allotted to more than one gender value. However, this non-canonical behaviour is constrained by the Agreement Hierarchy, given in Figure 5.

Constraint on agreement

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.

(A monotonic increase is one with no intervening decrease.) If we look back to (9) we see that semantic agreement is possible just in the position on the extreme right of the hierarchy, thus matching the constraint. For a summary of the evidence for the Agreement Hierarchy (including gender hybrids) and for recent references see Corbett (2012: 94–99).

attributive > predicate > relative pronoun > personal pronoun

Figure 5
The Agreement Hierarchy (Corbett 1979).

8. ASSIGNMENT

For any feature we need to know how it is assigned. This is especially relevant for gender because in the canonical situation gender is not context-dependent, nor is there speaker-choice. There is exactly one value per noun, as required by the Canonical Gender Principle.

8.1 *Canonical assignment*

The gender values of a language are formal classes, which are defined by sets of agreement markers (Corbett 1991: 145–150). This means that we can take different perspectives on gender values: they are both classes of nouns and values of a morphosyntactic feature. As Thornton (2009: 14) puts it, linguists talk of genders using the metaphor of containers (allotting nouns to genders), and taking the opposite perspective, assigning genders to nouns so that they function correctly in the syntax. Both perspectives are valid. Here we take the second perspective – given the lexical entry for a noun, we ask how its gender value can be assigned.

One key issue in gender assignment, or indeed in any linguistic categorization system, is what is being categorized: a linguistic entity (nouns in this instance) or the extra-linguistic reality (the referents of the phrases headed by these nouns). Some suggestions can be found in Senft (2000a: 27, 36–37; 2007: 689–690), though Senft is more concerned with the related issue of classifiers; see also Dahl (2000), Croft (2013), and Contini-Morava & Kilarski (2013) for discussion. It will often be the case that categorization could apply equally well to the semantics of

a noun or to a referent (otherwise the mechanism of reference could not work so well). It is here that we need also to bear in mind that agreement in gender appears in different domains, which affect the mechanism of assignment involved. Within the noun phrase, the noun is readily available, and assignment is almost always based straightforwardly on the lexical entry. At the other extreme of the Agreement Hierarchy there are anaphoric pronouns. These may be controlled by an antecedent noun, or they may be used deictically. And, most importantly, there are instances where it is not clear, to the hearer for instance, which it is. In other words, gender may be assigned on the basis of categorization of the referent, or of the lexical item, and it will not always be clear which is actually the case.

Canonical gender assignment is governed by the following criterion:

Canonical Gender – Criterion 3

In a canonical gender assignment system, the gender of a noun can be read unambiguously off its lexical entry.

The relevance of this criterion to the Canonical Gender Principle is evident: there ‘should’ be a single value read off the lexical entry, not open to other sources of gender assignment. This is also in line with the more general tenet of Canonical Typology that in canonical systems things line up; in this instance, the part-of-speech, the meaning and the form line up (see Spencer 2005: 102). From the perspective of Canonical Typology, the fact that the assignment of many nouns does not follow noun semantics but instead makes reference to the form (phonological or morphological) of the noun is not a problem. Such nouns are simply less canonical. There is a cline from languages with strictly semantic rules of assignment, like Tamil, to those where phonological or morphological factors have a major role. As an example of the latter, German has many inanimate nouns which are masculine, e.g. *Tisch* ‘table’, *Stuhl* ‘chair’. Their gender cannot be read off the semantics (lacking biological sex, they ‘should’ be neuter), but the lexical entry still helps because their form points towards masculine gender, given that most monosyllabic nouns in German are masculine (see Köpcke & Zubin 1984: 29), and the fuller paradigm is also a strong predictor. Nonetheless, these are less canonical examples of gender assignment because the phonological information points to masculine gender only statistically, the inflection class information only indicates that the noun is not feminine (it could be neuter though, like *Gebet* ‘prayer’, which belongs to the same inflection class as *Tisch* ‘table’), and semantics do not factor into gender assignment at all for these nouns.

It follows that a possible canonical system is one where the meaning of a noun is sufficient to allow for gender assignment. This goes against a widely held belief that gender ‘should’ be opaque. We contend that while an opaque gender system is easier to spot due to clear morphosyntactic evidence of gender, this is not the canonical situation. This might give the impression of a conflict with our position on canonical agreement in Section 2.1 above, but this is not the case. The point to

keep in mind is that made above, that in the canonical world the meaning and form of the controller should line up. Assignment is on the basis of lexical information, which in the canonical world is consistent between meaning and form. When meaning and form do not line up, for instance, if nouns denoting inanimate objects may have various inflectional types and thus be assigned to different genders, it is just easier to recognize a system of canonical agreement, since agreement is controlled by information that is unambiguously lexical.

It is worth recalling that systems of gender assignment always have a semantic core (Aksenov 1984: 17–18; Corbett 1991: 8). Thus there are consistent systems (strictly semantic), which are the canonical ones, and those which involve additional information on form (morphological and/or phonological), which are thereby less canonical.

8.2 *Conflicting assignment*

Having established canonical gender assignment we now turn to various cases of non-canonical assignment. We take up the more complex issue of recategorization in Section 8.3, where the recategorized noun can have a different gender available, which of course violates the Canonical Gender Principle.

The simplest case of conflicting assignment is illustrated by nouns like Russian *djadja* ‘uncle’, which should be assigned feminine gender according to their form (inflection class), but whose meaning indicates masculine. These are form–meaning mismatches which are straightforwardly determined in favour of the meaning. Hence, Russian *djadja* is masculine, as in *moj djadja* ‘my uncle’. Such nouns show consistent agreements across all targets.

Lexical hybrids are a slightly more involved case of a form–meaning mismatch. Recall that a hybrid is a noun which controls different agreements in different domains (Section 7). An example of a hybrid which will prove enlightening is German *Mädchen* ‘girl’. In contemporary German, *Mädchen* requires neuter agreement in the noun phrase but allows the neuter or the feminine form of the anaphoric pronoun (see example (9) above). This pattern of gender agreement, following the Agreement Hierarchy, was discussed in the section on domains (Section 7). Here we focus rather on the issue of form–meaning mismatch.

A common explanation for its hybrid status is that *Mädchen* ‘girl’ is neuter according to its diminutive morphology in *-chen*, but feminine according to its meaning. It is worth considering the issue of form a little further, since it is less clear that *Mädchen* ‘girl’ is an example of a form–meaning mismatch than is often assumed. This is because the word synchronically is no longer a transparent derivation. As opposed to transparent diminutives, such as *Häuschen* ‘little house’, there is no noun stem **Mad*, which would serve as the base to which the diminutive suffix *-chen* attaches. It seems that *Mädchen* ‘girl’ has gone down the path of lexicalization from the original transparent derivation *Mägdchen* ‘girl’ from *Magd* ‘maidservant’ (Birkenes, Chroni & Fleischer 2014: 4 fn. 4). In terms of derivational morphology, then, the status of *Mädchen* ‘girl’ is no

longer clear-cut. In inflectional terms, however, the picture is clearer. Its genitive singular (*Mädchen-s* ‘girl-SG.GEN’) shows it belongs to an inflection class which is unambiguously not feminine. It could in principle be masculine or neuter, though the phonological shape of the stem makes neuter assignment much the more likely. There is still, therefore, a form–meaning mismatch.

The meaning side of the equation is also interesting for this German noun. Braun & Haig (2010) show that for *Mädchen* ‘girl’, the age of the referent is an important factor in the choice between syntactic agreement and semantic agreement. In a questionnaire experiment, speakers used significantly more feminine pronouns when referring to an 18-year-old *Mädchen*, as opposed to a two- or 12-year-old one, where neuter pronouns were more frequent. These results show that speakers perceive biological sex as more important for adults than for children. Birkenes et al. (2014) find evidence for this in a corpus of German narrative texts written between the 17th and the 19th centuries. They report a high proportion of feminine agreements in the personal and possessive pronouns with the controller *Mädchen* ‘girl’ and suppose that the reason for this is the fact that the referents of *Mädchen* ‘girl’ in the examined texts were generally adults (Birkenes et al. 2014: 18). This single noun illustrates the complex interest of hybrids (see also Köpcke, Panther & Zubin 2010). However, there is essentially a form–meaning mismatch as the basis for its hybrid status.

We should also mention the comparable item here, namely German *Weib* ‘woman, wife’. This noun behaves like *Mädchen* ‘girl’, but according to Dahl (2000: 111), there is nothing in its phonology or morphology which would explain the neuter gender assignment. However, while there is no derivational indicator, *Weib* ‘woman, wife’, also has inflectional indication of a mismatch (its genitive singular is *Weib-s* and its nominative plural is *Weib-er*).

We conclude that a form–meaning mismatch is certainly not a sufficient condition for a noun to be a hybrid. Items like Russian *djadja* ‘uncle’ demonstrate this. Whether a form–meaning mismatch is a necessary condition is less clear: certainly the vast majority of hybrids identified show evidence of a mismatch, including the two we have just discussed. But it appears that there can be hybrids with no form–meaning mismatch. These have different possible semantic assignments, neither of them supported by formal properties, and yet they would still show differences in agreement according to the target. An example are female proper names in some west German and Swiss dialects and in Luxembourgish. These require neuter agreement within the noun phrase, but allow either a neuter or a feminine anaphoric pronoun. (Interestingly, some dialects have the converse, that is, feminine agreement in the noun phrase, with a choice between a neuter or a feminine anaphoric pronoun.) Neither of these assignments depends on the form, and these names are underived: specifically, they do not contain diminutive morphology, which would account for the neuter gender assignment (Nübling, Busley & Drenda 2013). We leave open the very challenging question of whether it might be possible to predict which nouns with a mismatch are hybrids and which are not (see Corbett 2015b for further discussion).

8.3 *Recategorization*

This is a fascinating, complex and still understudied area. Recategorization can be hard to spot, because there may be an overlap in the feature values between the original use and the recategorization. To take a familiar instance first, let us look at number (which is also affected by recategorization);⁹ specifically we look at mass nouns being recategorized as count nouns. The use of *coffee* as a mass noun is illustrated in (10):

(10) It's coffee that keeps us going till it's time to open the wine.

The following two examples show recategorization from a mass to count. *Coffee* appears as a count noun in the singular in (11) and in the plural in (12).

(11) I'd like a coffee.

(12) Two coffees and a tea please.

It is sometimes implied that only the plural is involved, but really it is recategorization from mass to count. This is generally easier to see in the plural, but the point is that mass nouns have one number value only, and when recategorized as count they have both values; one of these overlaps with the non-count usage. Thus English has singular for mass and singular-plural for count.

We can now turn to gender. A masculine noun might be recategorized (assigned by a different principle) so that we get masculine and feminine values available. But it would be clearest if the recategorization had only new values. We find this in the following Russian example (elicited 21 November 2013 from Marina Chumakina):

(13) Kak ja nenaviž-u èt-o čudovišč-e!
 how 1SG.NOM hate-1SG this-N.SG.NOM monster(N)-SG.NOM
 'How I hate this monster!'

(14) (a) Ona mne vsj-u žičn'
 3SG.F.NOM 1SG.DAT whole-F.SG.ACC life(F)[SG.ACC]
 isporti-l-a!
 spoil-PST-F.SG
 'She has spoiled my whole life!'

(b) On mne vsj-u žičn'
 3SG.M.NOM 1SG.DAT whole-F.SG.ACC life(F)[SG.ACC]
 isporti-l!
 spoil-PST[M.SG]
 'He has spoiled my whole life!'

[9] And this is an area where recategorization has received more attention, as shown by various papers in Massam (2012).

Čudovišče ‘monster’ denotes non-humans and is of neuter gender; this gender agreement is seen on the attributive modifier *ëto* ‘this’ in (13). However, it may be recategorized as human, in which case it is above the threshold of sex-differentiability, and hence takes masculine or feminine agreement of the pronoun (naturally to indicate a male or a female) in (14). The key point is that *čudovišče* ‘monster’ is normally neuter, but when recategorized it has quite different gender possibilities, namely feminine or masculine. Note too that this example has inconsistent agreements: the attributive remains neuter, while the personal pronoun is feminine or masculine, hence we are dealing with a hybrid. Such inconsistent agreement may occur with recategorization, as here, but it need not; some instances of recategorization have consistent agreement.

There are many phenomena which come under the broad umbrella of recategorization and the labels used are so disparate that the whole spread is rarely surveyed. Terms include ‘regular polysemy’, ‘meaning transfer’, ‘regular metonymy’ and more; see Ward (2004: 262–263) and Wechsler (2011: 1012 fn. 11) for lists of these terms.¹⁰ It is worth stressing that the phenomena vary along two dimensions (as noted in Nunberg 1996: 115–119).

First, instances vary as to the effect that morphosyntax has on them. This can affect what is possible and whether resulting agreements are consistent or not. We saw in (14) above that recategorization gave rise to a hybrid noun in Russian. In English, the most exotic transfers seem to be unhampered by morphosyntactic clashes, and have consistent agreement. There are languages where morphosyntactic clashes constrain the possibilities for such transfers (see, for instance, Kathol 1999: 246–247 on German). Where transfers are possible, which is a non-canonical situation, we can distinguish the outcomes as more or less canonical. The more canonical situation is naturally the one showing consistent agreement. However, this is frequently not what is found; the less canonical situation shows inconsistent agreement. Interestingly, these instances of inconsistent agreement, in gender and/or in number, are constrained by the Agreement Hierarchy.¹¹

[10] It is perhaps worth mentioning that some of these phenomena are often assumed to be problems of pragmatics, but this is far from obvious; see in particular Nunberg (1996) on meaning transfer.

[11] We should mention ‘pancake sentences’ here, as found in various languages, particularly in north-west Europe, and discussed recently in Enger (2013), Wechsler (2013) and Haugen & Enger (2014). These show recategorization from a specific item to a more general/abstract situation. To cite a poster-child example from Norwegian Nynorsk, Faarlund (1977: 251):

(i) Ein ny utanriksminister ville ikkje vere så dum-t.
a.M.SG new foreign.secretary(M) would not be so stupid-N.SG
‘A new foreign secretary would not be a bad idea.’

If the subject were a normal noun phrase, agreement would be *dum*, the masculine/feminine (traditionally ‘common’) gender form, and this gives an unkind reading. If there is recategorization, the reading is that HAVING a new foreign minister would not be stupid. While the predicate takes semantic agreement, note that the attributive modifier shows syntactic agreement (masculine).

Second, and cross-cutting the first dimension, instances vary from those which are so regular and apparently normal that we scarcely notice them to those which are highly context specific. There is variation both within and across languages. Apresjan (1974) drew attention to what he termed ‘regular polysemy’. Thus in Russian, as in English, the name of the animal can be used also for its meat, as in *gus* ‘goose’. At the other end of the scale we find the context specific ‘ham sandwich’ sentences, made famous by Nunberg (in various publications, including Nunberg 1993), for example in the restaurant context where the name of a dish can be used for the guest who has ordered it. Even the most innocuous examples vary across languages. While English shares with Russian the transfer from animal to meat, just noted, it is not so open to the transfer from bodily organ to a disease of that organ.¹² Thus in Russian one can say (15) (Apresjan 1974: 23–24; Nunberg 2008: 353), a transfer which is not readily available in English.

- (15) U nee počk-i.
 at 3SG.F kidney-PL.NOM
 ‘She has kidneys.’ (i.e. she has a disease of the kidneys)

A similar construction is possible in Italian (16) (Benni 1976[2011]: 27), but its context is restricted, being limited to athletes and their injuries.

- (16) Fedele ha il menisco.
 Fedele has the meniscus
 ‘Fedele has a meniscus injury.’ (of Adriano Fedele, an Italian footballer of the 1970s)

In gender systems, recategorization frequently involves changing the threshold of sex-differentiability, which means allowing items below the threshold to be reclassified as being above it, and thus to have additional gender possibilities. English is very liberal in allowing *he* and *she* for reference to personified animals and even inanimate objects in stories. Other languages are much less ready to allow recategorization of this type. The Daghestanian language Tsez has the following gender system: male humans – gender I, female humans – gender II, nouns denoting animals – gender III, but nouns denoting inanimates are distributed over three genders, namely II, III and IV. Throughout a story, in which a rooster has an affair with a frog, both are treated as gender III, just as in

Targets higher on the agreement hierarchy also take semantic agreement (Corbett 2006: 223–224). Note, however, that this is a more radical type of recategorization, since gender and number are involved simultaneously (as shown by the fact that plurals take singular agreement, outside the noun phrase).

[12] Similarly when number is involved: it seems natural to speakers of English that *teas* can be used for both portions (cups) of tea and types of tea, but in other languages these two different readings are not equally available. See Corbett (2000: 81–82, 84–87) for discussion, and Wiese & Maling (2005) for interesting detail on the differences between closely related languages in ‘restaurant talk’.

normal situations. No recategorization is possible (Comrie 2005). Biniŋj Gunwok (Mayali), a language of the Gunwinyguan family of northern Australia, is a little more liberal. Thus *alwanjdjuk* ‘emu’ is feminine, irrespective of the sex of a given referent (it is an epicene noun, Section 4.1 above). Yet in exceptional circumstances, the use of the masculine is possible (Evans, Brown & Corbett 2002: 130–131), for example in an account of what happens when emus divorce, and the wife emu marries another male emu, which is exceptionally treated as masculine.

It is not only sex-differentiability which can be the basis for recategorization. Savosavo, which is the easternmost Papuan language, and is spoken in the Solomon Islands, has two genders, masculine and feminine. Nouns denoting inanimates are by default masculine. However, recategorization is possible, according to a diminutive/affective classification. Nouns which are normally masculine (since inanimate) are occasionally made feminine; this may function to indicate that the referent is small compared with normal expectation or that it is in some way special (Wegener 2012: 60). An example of diminutivization is given in (17) (Wegener 2012: 63):

- (17) Pade gnari molo pono=lo te
 one small knife only=3SG.M.NOM EMPHATIC
 k-au-i.
 3SG.F.OBJ-take-FINITE
 ‘He only took a small knife.’

In this example, *molo* ‘knife’ which is normally masculine is treated as feminine, as shown by the agreement on the verb. Such recategorization in Savosavo clearly depends on properties of the referent. The property of ‘being smaller than the norm’ which prompts feminine agreements is not part of the lexical semantics of a noun but clearly a characteristic of a specific referent and the speaker’s view. Savosavo is also a good case of a language where we would not want to say that there are two homophonous lexemes, one being masculine and one being feminine and a diminutive, because this would almost double the noun lexicon of the language, given that all inanimates and all animates with not readily discernible sex can undergo multiple categorization. This would be an especially implausible analysis given that recategorization as feminine is an infrequent phenomenon in discourse and that there is considerable inter-speaker variation (Wegener 2012: 63). For Savosavo we have to say that the recategorization as feminine is a pragmatic effect which relies on features of the referent and the speakers’s view.

In a language like Savosavo we see that the less the extent to which gender agreement is fully determined by unique gender values, the greater its role in constructing meaning. We note here other examples for comparison, namely Lavukaleve (Terrill 2003: 140–141) and Walman (Brown & Dryer 2008a: 530; 2008b). In Walman, recategorization according to a diminutive/affective classification is widespread, and for this there is a set of diminutive gender agreement

forms. Note also the instances in the Iwaidjan language Mawng (Singer 2010) and in Yawuru (Nyulnyulan family, Hosokawa 1996).

Naturally all these interesting effects are non-canonical: a noun 'should' have a single gender value, unaffected by its context. Hence, all types of gender recategorization violate the Canonical Gender Principle.

9. CONCLUSION

The gender systems we find in the languages of the world are not monolithic entities, but can be analysed as clusterings of properties. The Canonical Typological approach allows us to characterize these individual properties with respect to a canonical ideal, rather than requiring us to assign a type to the system as a whole. This is invaluable if we wish to make comparisons of different noun categorization systems across languages. When defining canonical gender we can rely on established principles and criteria for canonical agreement and canonical morphosyntactic features. We also need just one specific principle and three associated criteria, which spell out its implications in order to characterize canonical gender.

Since gender is tied to the lexical entry, it is the canonical agreement feature. In many respects the varying characteristics of gender mirror those of the other morphosyntactic features. However, gender is defined as a non-canonical morphosyntactic feature because it violates the exhaustiveness criterion for canonical morphosyntactic features: in the canonical situation lexemes have access to all values of a morphosyntactic feature, which does not hold for gender in respect of nouns. This distinctive behaviour of gender among the morphosyntactic features finds expression in the Canonical Gender Principle: IN A CANONICAL GENDER SYSTEM, EACH NOUN HAS A SINGLE GENDER VALUE. The closer a system approximates to this situation, the more canonical it is. There are three associated criteria, which ensure that in the canonical situation a noun has exactly one gender value. The first criterion is that canonical gender values match agreement classes. Any deviation immediately creates non-canonicity because one or more nouns will be assigned two gender values. The second criterion states that in a canonical gender system the gender of a noun is constant across all agreement domains. This makes hybrid nouns non-canonical, since they are associated with more than one gender value, restricted by the agreement domain. The third criterion guarantees that canonical gender can be read unambiguously off the lexical entry of a noun. This is what we expect from the perspective of Canonical Typology, where classifications such as part of speech, meaning and form align in the canonical situation.

For target issues we can fall back on established principles and criteria for canonical morphosyntactic features, which apply to gender by virtue of it being a morphosyntactic feature, but which likewise apply to person, case and number. The same holds for the interaction of gender with other features. While it is canonical for the distribution of gender not to be dependent on other features,

this is a special case of the more general criterion that canonical features and their values are uniquely distinguished across the other logically compatible features and their values. Gender typically interacts with number, and in the canonical case gender and number cross-cut each other, as for example in Italian. German, Archi and Mian all illustrate deviations from this canonical ideal, in that in all of these languages the distribution of gender depends on number.

We have taken a significant step towards a typology of noun classification. In large measure we have done this by using machinery already available from the characterization of canonical morphosyntactic features. The additional Canonical Gender Principle is an idealization of an old insight, and in working through it in detail we have seen that it provides what we need for typologizing gender systems. Moreover, it anchors a canonical point from which, we believe, it will be possible to understand other types of categorization.

REFERENCES

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Aksenov, A. T. 1984. K probleme èkstralingvistièeskoj motivacii grammatičeskoj kategorii roda [On the extralinguistic motivation of the grammatical category of gender]. *Voprosy jazykoznanija* 1, 14–25.
- Apresjan, Juri D. 1974. Regular polysemy. *Linguistics* 12.142, 5–32.
- Asher, R. E. 1985. *Tamil*. London: Croom Helm. [Reprinted 1989, London: Routledge]
- Audring, Jenny. 2011. Gender. *Oxford bibliographies online: Linguistics*. <http://www.oxfordbibliography.com/view/document/obo-9780199772810/obo-9780199772810-0066.xml> (accessed 4 January 2012).
- Audring, Jenny. 2014. Gender as a complex feature. In Tania Paciaroni, Anna M. Thornton & Michele Loporcaro (eds.), *Exploring grammatical gender: Special issue of Language Sciences* 43, 5–17.
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax–morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.
- Bakker, Peter. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Oxford: Oxford University Press.
- Benni, Stefano. 1976[2011]. *Bar sport*. Milan: Universale Economica Feltrinelli.
- Birkenes, Magnus Breder, Kleopatra Chroni & Jürg Fleischer. 2014. Genus- und Sexuskongruenz im Neuhochdeutschen: Ergebnisse einer Korpusuntersuchung zur narrativen Prosa des 17. bis 19. Jahrhunderts. *Deutsche Sprache* 42, 1–24.
- Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1995*, 1–15. Dordrecht: Kluwer.
- Braun, Friederike & Geoffrey Haig. 2010. When are German ‘girls’ feminine? How the semantics of age influences the grammar of gender agreement. In Markus Bieswanger, Heiko Motschenbacher & Susanne Mühleisen (eds.), *Language in its socio-cultural context: New explorations in gendered, global and media uses*, 69–83. Frankfurt/Main: Peter Lang.
- Brown, Dunstan & Marina Chumakina. 2013. What there might be and what there is: An introduction to Canonical Typology. In Brown et al. (eds.), 1–19.
- Brown, Dunstan, Marina Chumakina & Greville G. Corbett (eds.). 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Brown, Lea & Matthew S. Dryer. 2008a. The verbs for ‘and’ in Walman, a Torricelli language of Papua New Guinea. *Language* 84, 528–565.
- Brown, Lea & Matthew S. Dryer. 2008b. Diminutive as an inflectional category in Walman. Ms., University at Buffalo. linguistics.buffalo.edu/people/faculty/dryer/dryer/BrownDryerWalmanDimin.pdf (accessed 2 June 2014).
- Comrie, Bernard. 2005. Grammatical gender and personification. In Dorit Diskin & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development: Essays in honour of Ruth A. Berman*, 105–114. Dordrecht: Kluwer.

- Contini-Morava, Ellen & Marcin Kilarski. 2013. Functions of nominal classification. *Language Sciences* 40, 263–299.
- Corbett, Greville G. 1979. The Agreement Hierarchy. *Journal of Linguistics* 15, 203–224. [Reprinted in Francis X. Katamba (ed.). 2003. *Morphology: Critical concepts in linguistics*, vol. 4: *Morphology and syntax*, 48–70. London: Routledge.]
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007. Gender and noun classes. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 1: *Grammatical categories and the lexicon*, 2nd edn., 241–279. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2011. The penumbra of morphosyntactic feature systems. In Jonathan Bobaljik, Uli Sauerland & Andrew Nevins (eds.), *Markedness and underspecification in the morphology and semantics of agreement*: Special issue of *Morphology* 21, 445–480.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013. Canonical morphosyntactic features. In Brown et al. (eds.), 48–65.
- Corbett, Greville G. 2015a. Morphosyntactic complexity: A typology of lexical splits. *Language* 91, 145–193.
- Corbett, Greville G. 2015b. Hybrid nouns and their complexity. In Jürg Fleischer, Elisabeth Rieken & Paul Widmer (eds.), *Agreement from a diachronic perspective* (Trends in Linguistics: Studies and Monographs 287), 191–214. Berlin: De Gruyter Mouton.
- Croft, William. 2013. Agreement as anaphora, anaphora as coreference. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 95–117. Berlin: De Gruyter Mouton.
- Dahl, Östen. 2000. Animacy and the notion of default gender. In Barbara Unterbeck, Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*, vol. 1: *Approaches to gender* (Trends in Linguistics: Studies and Monographs 124), 99–115. Berlin: Mouton de Gruyter.
- Dixon, R. M. W. 1982. Nominal classification. In R. M. W. Dixon (ed.), *Where have all the adjectives gone? And other essays in semantics and syntax*, 157–233. Berlin: De Gruyter.
- Donohue, Mark. 2001. Animacy, class and gender in Burmeso. In Andrew Pawley, Malcolm Ross & Darrell Tryon (eds.), *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific Linguistics 514), 97–115. Canberra: Pacific Linguistics.
- Drabbe, Petrus. 1955. *Spraakkunst van het Marind, zuidkust Nederlands Nieuw Guinea* [A grammar of Marind, south coast of Dutch New Guinea] (Studia Instituti Anthropos 11). Wien & Mödling: Missiehuis St. Gabriel.
- Emeneau, Murray B. 1955. *Kolami: A Dravidian language* (University of California Publications in Linguistics 12). Berkeley, CA: University of California Press.
- Enger, Hans-Olav. 2013. Scandinavian pancake sentences revisited. *Nordic Journal of Linguistics* 36, 275–301.
- Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 2002. The semantics of gender in Mayali: Partially parallel systems and formal implementation. *Language* 78.1, 111–155.
- Faarlund, Jan Terje. 1977. Embedded clause reduction and Scandinavian gender agreement. *Journal of Linguistics* 13, 239–257.
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: De Gruyter Mouton.
- Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In Senft (ed.), 50–92.
- Haugen, Tor Arne & Hans-Olav Enger. 2014. Scandinavian pancake constructions as a family of constructions. *Cognitive Linguistic Studies* 1, 171–196.
- Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler & Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen*, vol. 1: *Bereich und Ordnung der Phänomene*, 189–216. Tübingen: Narr.
- Hosokawa, Komei. 1996. 'My face *am* burning!' quasi-passive, body parts, and related issues in Yawuru grammar and cultural concepts. In Hilary Chappell & William McGregor (eds.), *The grammar of inalienability: A typological perspective on body parts and the part-whole relation*, 155–192. Berlin: Mouton de Gruyter.
- Kathol, Andreas. 1999. Agreement and the syntax–morphology interface in HPSG. In Robert D. Levine & Georgia M. Green (eds.), *Studies in contemporary phrase structure grammar*, 223–274. Cambridge: Cambridge University Press.

- Kibrik, A. E., S. V. Kodzasov, I. P. Olovjannikova & D. S. Samedov. 1977. *Opyt strukturnogo opisaniya arčinskogo jazyka I: Leksika, fonetika* [A structural description of Archi I: Lexis and phonetics] (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki 11). Moscow: Izdatel'stvo Moskovskogo universiteta.
- Kilarski, Marcin. 2013. *Nominal classification: A history of its study from the classical period to the present*. Amsterdam: John Benjamins.
- Köpcke, Klaus-Michael & David A. Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen. Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte* 93, 26–50.
- Köpcke, Klaus-Michael, Klaus-Uwe Panther & David A. Zubin. 2010. Motivating grammatical and conceptual gender agreement in German. In Hans-Jörg Schmid & Susanne Handl (eds.), *Cognitive foundations of linguistic usage patterns*, 171–194. Berlin: De Gruyter Mouton.
- Loporcaro, Michele. Forthcoming. Gender. In Adam Ledgeway & Martin Maiden (eds.), *Oxford handbook of the Romance languages*. Oxford: Oxford University Press.
- Loporcaro, Michele & Tania Paciaroni. 2011. Four-gender systems in Indo-European. *Folia Linguistica* 45.2, 389–434.
- Maiden, Martin & Cecilia Robustelli. 2007. *A reference grammar of modern Italian*, 2nd edn. London: Hodder.
- Massam, Diane (ed.). 2012. *Count and mass across languages*. Oxford: Oxford University Press.
- Mel'čuk, Igor A. 2013. The notion of inflection and the expression of nominal gender in Spanish. *Studies in Language* 37.4, 736–763.
- Meyer, Peter. 1994. Grammatical categories and the methodology of linguistics. Review article on W. Andries van Helden, *Case and gender: Concept formation between morphology and syntax*, 1993. *Russian Linguistics* 18, 341–377.
- Nübling, Damaris, Simone Busley & Juliane Drenda. 2013. *Dat Anna und s Eva* – Neutrale Frauenrufnamen in deutschen Dialekten und im Luxemburgischen zwischen pragmatischer und semantischer Genuszuweisung. *Zeitschrift für Dialektologie und Linguistik* 80.2, 152–196.
- Nunberg, Geoffrey D. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16, 1–43.
- Nunberg, Geoffrey D. 1996. Transfers of meaning. In James Pustejovsky & Branimir Boguraev (eds.), *Lexical semantics: The problem of polysemy*, 109–132. Oxford: Clarendon.
- Nunberg, Geoffrey D. 2008. The pragmatics of deferred interpretation. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 344–363. Oxford: Blackwell.
- Pišković, Tatjana. 2011. *Gramatika roda* [The grammar of gender]. Zagreb: Disput.
- Pullum, Geoffrey K. & Arnold M. Zwicky. 1988. The syntax–phonology interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey*, vol. 1, 255–280. Cambridge: Cambridge University Press.
- Reid, Nicholas. 1997. Class and classifiers in Ngan'gityemerri. In Mark Harvey & Nicholas Reid (eds.), *Nominal classification in Aboriginal Australia*, 165–228. Amsterdam: John Benjamins.
- Seifart, Frank. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Ph.D. dissertation, Radboud University.
- Senft, Gunter. 2000a. What do we really know about nominal classification systems? In Senft (ed.), 11–49.
- Senft, Gunter (ed.). 2000b. *Systems of nominal classification*. Cambridge: Cambridge University Press.
- Senft, Gunter. 2007. Nominal classification. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 676–725. Oxford: Oxford University Press.
- Shevelov, G. Y. 1963. *The syntax of modern literary Ukrainian: The simple sentence*. The Hague: Mouton.
- Singer, Ruth. 2010. Creativity in the use of gender agreement in Mawng: How the discourse functions of a gender system can approach those of a classifier system. *Studies in Language* 34, 382–416.
- Spencer, Andrew. 2005. Towards a typology of 'mixed categories'. In C. Orhan Örgün & 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, CA: CSLI Publications.
- Terrill, Angela. 2003. *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- Thornton, Anna M. 2009. Constraining gender assignment rules. *Language Sciences* 31, 14–32.
- Ward, Gregory. 2004. Equatives and deferred reference. *Language* 80, 262–289.
- Wechsler, Stephen. 2011. Mixed agreement, the person feature, and the index/concord distinction. *Natural Language & Linguistic Theory* 29, 999–1031.
- Wechsler, Stephen. 2013. The structure of Swedish *Pancakes*. In Philip Hofmeister & Elisabeth Norcliffe (eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*. Stanford, CA: CSLI Publications.

- Wegener, Claudia. 2012. *A grammar of Savosavo*. Berlin: De Gruyter Mouton.
- Wiese, Heike & Joan Maling. 2005. *Beers, Kaffi, and Schnaps*: Different grammatical options for restaurant talk coercions in three Germanic languages. *Journal of Germanic Linguistics* 17, 1–38.
- Zaliznjak, Andrej A. 1964. K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom ruskom jazyke [The grammatical categories of gender and animacy in modern Russian]. *Voprosy jazykoznanija* 4, 25–40.
- Zaliznjak, Andrej A. 1973. O ponimanii termina ‘padež’ v lingvističeskix opisanijax [Interpreting the term ‘case’ in linguistic descriptions]. In Andrej A. Zaliznjak (ed.), *Problemy grammatičeskogo modelirovanija* [Problems of grammatical modelling], 53–87. Moscow: Nauka. [Reprinted in Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu ruskomu jazyku i obščemu jazykoznaniju* [Russian nominal inflection: With a supplement of selected works on contemporary Russian and on general linguistics], 613–647. Moscow: Jazyki slavjanskoj kul’tury.]

Authors’ addresses: (Corbett)

*Surrey Morphology Group, School of English and Languages,
University of Surrey, Guildford GU2 7XH, UK
g.corbett@surrey.ac.uk*

(Fedden)

*Department of Linguistics, School of Letters, Art and Media,
John Woolley Bldg A20, The University of Sydney, NSW 2006,
Australia
sebastian.fedden@sydney.edu.au*