



# Splits, internal and external, as a window into the nature of features

Greville G. Corbett<sup>1</sup> 

Received: 17 February 2021 / Accepted: 12 September 2021 / Published online: 28 October 2021  
© The Author(s) 2021

## Abstract

Lexemes may be split *internally*, by phenomena such as suppletion, periphrasis, heteroclisis and deponency. Generalizing over these phenomena, which split a lexeme's paradigm, we can establish a typology of the possible internal splits. There are also lexemes whose *external* requirements are split: they induce different agreement, for instance. Again, a typology of these splits has been proposed. The next logical step is to attempt a typology of the possible *relations between internal and external splits*. This is not straightforward, since we need to avoid spurious linkages. Four lines of argument are offered: (i) general plausibility: the internal-external linkage is compelling, and so other accounts require a degree of coincidence which is unlikely; (ii) overabundance: alternative inflectional forms link to different external requirements; (iii) variation in time and space: splits in inflection and in external requirements vary, while maintaining their linkage; (iv) pluralia tantum nouns: the different types of these nouns provide intriguing confirming evidence. Case studies include Asia Minor Greek, Polish, Russian, Scottish Gaelic, Σελλε, Serbo-Croat (BCMS), Slovenian, Latin and Old Frisian. The clear instances which emerge, where an external split is demonstrably linked to an internal one, prove both surprising and significant. We discover that in split paradigms, besides overt overabundance, there may also be covertly overabundant cells. Furthermore, when external splits involve individual cells, these will not induce simple (consistent) agreement. This makes good sense, demonstrating that featural information is associated with lexemes in a natural default manner: at the lexeme level by default, unless overridden at the sub-paradigm level, unless in turn overridden at the level of individual cells.

**Keywords** Morphosyntactic features · Gender · Number · Inflection class · Heteroclites · Canonical typology · Extended lexical consistency principle · Covert overabundance · Agreement hierarchy · Morphology-free syntax

---

✉ G.G. Corbett  
[g.corbett@surrey.ac.uk](mailto:g.corbett@surrey.ac.uk)

<sup>1</sup> Surrey Morphology Group, Literature and Languages, Faculty of Arts and Social Sciences, University of Surrey, Guildford, Surrey, GU7 2XH, UK

Some lexical items show remarkable properties: they may lack internal consistency, they may be externally inconsistent and – in a fascinating minority – these two characteristics may be linked. This last type is rare, but it deserves special attention for what it tells us about the nature of lexical information and the way this is stored. I first review briefly the typology of lexeme-internal splits; these reflect phenomena such as suppletion, periphrasis, heteroclis and deponency, and they reveal higher-level patterns. Then I summarize the types of external splits (notably different case and agreement requirements), and the conditions on such splits. This outline of internal and external splits forms the essential basis for analysing the unusual instances where internal and external splits are linked. Demonstrating that the splits are linked is not straightforward, and so I give four types of argument to establish that particular splits are indeed linked. The examples, from a range of languages, show subtle variations, and these differences allow us to see clearly the hierarchy of ways in which featural information is associated with lexemes.

## 1 Essentials: internal and external splits

The best dictionary entries embody the secure accumulated results of linguistic research. If the entry for *enjoy* gives no information about the past tense, we infer that this verb is regular, there is no split in its morphological paradigm, and hence its past tense form is *enjoyed*. Conversely, the past of *go* is specified as *went*. This verb has a split in its paradigm, an internal split, induced by suppletion. Besides full suppletion, we might find the paradigm split by lesser alternations, augments, heteroclis, defectiveness or semi-deponency. Turning to external requirements, a Russian dictionary can state simply that the preposition *k* ‘towards’ takes the dative case value; this is invariably true, which means that this preposition is externally consistent. Contrast this with Russian *po* ‘for, about, by’, for which much more must be specified: since it governs different case values (accusative, dative and locative), according to complex conditions. That is, it shows a split in its external requirements.

The underlying notion here may be expressed as the Lexeme Consistency Principle:

A lexeme’s internal structuring and its external requirement are both consistent.

Of course, there are exceptions to this principle, as in the examples already given. Its useful function is as a baseline from which we can calibrate the interesting diversity we actually find. This method, setting unambiguous baselines, and calibrating carefully from them, is a hallmark of Canonical Typology.<sup>1</sup> So we start from the assumption that, given a basic lexical entry, there is nothing more to be stated: lexemes are *internally* and *externally* consistent. That is the baseline set by the principle; it is simply like agreeing to measure length from zero. We shall revisit briefly the typology of exceptions to this principle, internal splits in §1.1 and external splits in §1.2. For clarity, each type of split will be justified independently, before we go on to the main topic, the relations between such internal and external splits.

<sup>1</sup> For recent references on Canonical Typology see Round and Corbett (2020) and the bibliography at [tiny.cc/ctbib](https://tiny.cc/ctbib).

## 1.1 Internal splits

Here is a clear instance of an internal split.

- (1) Present tense sub-paradigm of Italian (ita) *andare* ‘go’ (Maiden, 2004:233):

	SINGULAR	PLURAL
1	vado	andiamo
2	vai	andate
3	va	vanno

It is evident that (1) is not internally consistent.<sup>2</sup> It shows two quite different stems, hence it is an instance of suppletion. There is a canonical typology of inflectional phenomena, which allows us to integrate suppletion and other phenomena such as defectiveness and periphrasis into a comprehensive scheme (Corbett, 2015a:149–158). But we can go further, taking a more abstract view, and ask more generally about how paradigms are split, irrespective of the phenomenon which induces the split. Four criteria have been proposed, which are introduced briefly here (see Corbett, 2015a:158–179 for the detail):

- C1: FORM VERSUS COMPOSITION/FEATURE SIGNATURE: example (1) concerns just the form. The paradigm has the same featural scheme (in terms of person and number) as other Italian verbs. Only the forms are unpredictable from the inflectional system of Italian. This contrasts with paradigms where the split is a matter of the feature signature: it involves different features (e.g. Russian verbs in the past tense mark number and gender while the present marks person and number).
- C2: JUSTIFICATION: MORPHOMIC VERSUS MOTIVATED: the pattern in (1), namely first and second persons plural versus the remainder, is one that is internal to the morphology (morphomic); it contrasts with patterns which are motivated from outside the paradigm (for instance, in terms of semantics).
- C3: SPECIFICATION OF PATTERN: LEXICALLY SPECIFIED TO FULLY REGULAR: this criterion highlights the interest of our example. In terms of the forms, Italian *andare* ‘go’ is exceptional, being the only verb with these suppletive stems (apart from the derived *riandare* ‘to go again’). More abstractly, however, there are other verbs with irregular patterning involving the same cells, and so Italian *andare* ‘go’ must be lexically specified but it is not unique.
- C4: RELEVANCE: ‘INTERNAL’ AND ‘EXTERNAL’ SPLITS: in the canonical case such splits are internal to the lexeme, and have no effect in syntax. That is what we find with our Italian example. It behaves like other verbs, in this respect. We do find internal splits which have external relevance, and these will be our main focus.

<sup>2</sup>It is conventional and convenient to lay out the forms of lexemes in this way; the key point is that this representation makes clear the cross-cutting feature values, the internal organization of inflectional morphology. How such paradigms are to be explained is open to differing styles of analysis (see Corbett, 2015a:147–149 for discussion).

This is not an arbitrary list of criteria; rather there is an underlying logic, a ‘sub-struction’ (Lazarsfeld, 1937:133–136; Round & Corbett, 2020:493–496), which establishes how the criteria cover the theoretical space (Corbett, 2015a:177–178). We have seen, then, that there are lexemes with internal splits in their paradigm, independent of any external split. But a key point from that typology for present purposes is Criterion 4. This criterion allows for the situation where a split may not just be a morphology-internal split but may correspond to an external split too.

## 1.2 External splits

We now turn to external splits, setting them out in their own right first, before looking at possible connections to internal splits. Turkish (tur) provides a clear instance:

Turkish (Kornfilt, 1997:423–424)<sup>3</sup>

- (2) Hasan bu sonat-ı **ben-im** gibi çal-dı  
 Hasan this sonata-ACC 1SG-GEN like play-PST  
 ‘Hasan played this sonata like me.’
- (3) Hasan bu sonat-ı **Rubinstein** gibi çal-dı  
 Hasan this sonata-ACC Rubinstein[NOM] like play-PST  
 ‘Hasan played this sonata like Rubinstein.’

The postposition *gibi* ‘like’ has an external requirement, namely the case value of its governee. And there is a split in this requirement: in (2) we find the genitive, but in (3) the nominative. The case value to be used is determined by specific conditions (I give fuller detail on this split in Corbett, [under review](#)). One of these conditions is that the part of speech of the governee is crucial: with nouns the nominative is used, with pronouns there is a more complex choice between nominative and genitive. There is a split too in the governors. While various postpositions in Turkish govern a single case value, *gibi* ‘like’ (and three other postpositions, *için* ‘for’, *-(y)la/ile* ‘with, by’ and *kadar* ‘as...as’) stand out as exceptional; these four are split from the remaining postpositions, since they have a split in their external requirement. This is true whether one takes a syntactic perspective, expecting adpositions to have consistent government requirements, or a semantic perspective, expecting case values (nominative and genitive here) to have consistent semantics.

This use of the term ‘split’, shares a unifying notion with other uses. We partition the lexicon into parts of speech (lexical categories); then there are regularities which can cross-cut the parts of speech, namely features. When we require an additional partition, in either dimension (parts of speech or features), we term it a split. Further, for given external relations we divide lexemes into primary (governor for case, controller for agreement) versus secondary (governee for case, target for agreement).

<sup>3</sup>I follow the Leipzig Glossing Rules <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>; [ ] indicates information that can be inferred from the use of the bare stem, <> marks infixation. In addition, bold is used as a flag to draw attention to relevant characteristics of examples. Abbreviations are the Leipzig ones with some additions: 1, 2, 3: first, second, third person, ACC: accusative, ADJ: adjective, AUX: auxiliary, DEF: definite, DEM: demonstrative, DEP: dependent, DET: determiner, F: feminine, GEN: genitive, IC: inflection class, LOC: locative, M: masculine, N: neuter, NOM: nominative, NON\_PST: non-past, PL: plural, PRS: present, PST: past, PTC: participle, SG: singular.

The additional partitioning (split) may involve the primary lexeme and/or the secondary lexeme. To be specific, in the Turkish example, there is first a split involving the primary lexeme (the postposition which is the case governor). Turkish has a set of postpositions, justified by their common syntactic properties. We need to partition this part of speech further, since the special case government we are interested in is restricted to four postpositions (hence this is an inter-lexemic split). The secondary lexeme (the governee) is also affected: there is a split here since nouns behave differently from pronouns (an inter-lexemic split), and furthermore within the pronouns the number feature is invoked (a featural split). More generally, lexemes may be referenced in their entirety (thus a verb may govern a particular case whatever its own featural specification); or according to their featural specification (nouns require attributive modifiers to agree according to their number). The secondary lexeme equally has these two possibilities (for instances, nouns and pronouns are governed differently, or governees in the singular are treated differently from plurals). These four possibilities (induced by the oppositions entire lexeme vs feature specification, and primary vs secondary lexeme) allow us to build an elegant typology using this minimal machinery (Corbett, [under review](#)).

Clearly then, there are instances of external splits, independent of internal splits, and earlier we examined internal splits, also independently (§1.1). It is time to ask whether and how they may be related.

## 2 Internal and external splits: possible relations between them

We started from the idea that, given a lexical entry, there is nothing more to be stated: lexemes are *internally* and *externally* consistent. We then reviewed the interesting phenomena that deviate from these canonical baselines. Given that there are internal and external splits, and a typology for each, the next logical step is to ask about the relation between them. It is here that we call on the Extended Lexeme Consistency Principle, which has an additional final clause:

A lexeme's internal structuring and its external requirement are both consistent. Furthermore, they are consistent with each other.

This is the baseline against which we shall calibrate the examples we find. Since splits can be internal (§1.1) or external (§1.2), there are four logical possibilities for the relation, as in Table 1.

Let us take the possibilities in turn. The Russian preposition *k* 'towards' has no internal split; it cannot have, since Russian prepositions do not inflect. Its external requirements are consistent: it always governs the dative case. In these respects it is canonical and of no further interest here. That is the simplest type. More interestingly, Russian *komnata* 'room' inflects, and it does so without any split in its paradigm. Furthermore, it is externally consistent (the relevant external property is agreement, and this requirement is consistent). Thus it too has neither an internal nor an external split, though in principle both would be possible. In the second type, the verb *go* is split internally, having the suppletive stems *go* and *went*; but externally, it is consistent. Third, we consider items which have external splits, but no internal

**Table 1** Possible relations of internal and external splits

		external split	
		no	yes
internal split	no	1. Russian <i>k</i> ‘towards’ Russian <i>komnata</i> ‘room’	3. Turkish <i>gibi</i> ‘like’, Russian <i>po</i> ‘for’ Serbo-Croat <i>gazda</i> ‘landlord’
	yes	2. English <i>go</i>	4. focus of paper

splits. There are various types of example here (to compare with type 1). We discussed Turkish *gibi* ‘like’ in §1.2. Like Russian *k* ‘towards’, the Russian preposition *po* ‘for’ cannot show an internal split, since it does not inflect; however, in contrast to *k* ‘towards’, *po* ‘for’ is a wonderful example for complex external splits, governing different cases according to the part of speech and number of the governee, as well as its particular function (see Muravenko, 2014; Corbett, [under review](#), and references in both). Still, none of these could be internally split. A more interesting example therefore is Serbo-Croat *gazda* ‘landlord’ and many similar nouns. Internally they are consistent; they have no internal split. Yet externally they are masculine when singular, but hybrid when plural; that is, their agreements vary (masculine vs feminine) according to the agreement target (Marković, 1954; Corbett, 2015b:205–207 and references there, Franks, 2020:448–464). Thus they have an external split, but no internal one.

It is the fourth type which we shall concentrate on, lexemes which are split both internally and externally. The splits may or may not be linked. The less interesting type is that where there is simply no relation between the splits, and hence we cannot ask whether they are consistent with each other. As an example, Russian verbs have an internal split involving present vs past tense; the present realizes person and number, and the past has number and gender. The split involves the feature signature (see Criterion 1 in §1.1, and Corbett, 2015a:146, 157). Some have also an external split; thus *ždat’* ‘wait for’ can take an accusative or genitive object. However, there is no relation between the internal, featural split and the external, government split. This may seem obvious, yet there are splits in government which are sensitive to the TAM properties of the verb. Hence we have to demonstrate that internal and external splits are linked, instance by instance; we cannot assume a link, nor indeed the lack of one. Where we can demonstrate a link, then we have the interesting issue of consistency between the splits.

It is worth stressing at this point that there are many examples of splits with an apparent internal-external link, but where (disappointingly) the relation proves to be at best indirect, on closer examination. An example of just such a mistaken link that is often cited by statisticians concerns Dutch storks: statistics for a series of springs suggest a correlation between the number of storks nesting and the number of human births (see, for instance, Sapsford, 2006). But there is no direct causal relationship. Rather there is a third variable, the weather nine months earlier. We can find similar

examples involving splits. Consider the following data on nouns from inflection class I in Serbo-Croat:<sup>4</sup>

(4) The Serbo-Croat augment (in inflection class I)<sup>5</sup>

		‘city’	‘window’
SINGULAR	NOMINATIVE	grad	prozor
	VOCATIVE	grad-e	prozor-e
	ACCUSATIVE	grad	prozor
	GENITIVE	grad-a	prozor-a
	DATIVE	grad-u	prozor-u
	INSTRUMENTAL	grad-om	prozor-om
	LOCATIVE	grad-u	prozor-u
PLURAL	NOMINATIVE	grad- <b>ov</b> -i	prozor-i
	VOCATIVE	grad- <b>ov</b> -i	prozor-i
	ACCUSATIVE	grad- <b>ov</b> -e	prozor-e
	GENITIVE	grad- <b>ov</b> -ā	prozor-ā
	DATIVE	grad- <b>ov</b> -ima	prozor-ima
	INSTRUMENTAL	grad- <b>ov</b> -ima	prozor-ima
	LOCATIVE	grad- <b>ov</b> -ima	prozor-ima

The paradigm of the noun *grad* ‘city’ is split by the absence or presence of the augment *-ov-* as shown (Browne, 1993:319–320). The conditions determining which nouns take this augment are complex, interesting and changing (see Baerman et al., 2017:93–98). What is relevant here is that for all cells of the paradigm which take this augment, agreement is plural. We might be tempted to suggest that the agreements are causally linked to the augment. But this would be falling into the Dutch stork trap. There are other nouns, like *prozor* ‘window’, which do not take the augment. But they take exactly the same set of agreements as those with the augment. In other words, nouns like *grad* ‘city’ in (4) have an internal split in their paradigm, conditioned by number. Agreement is also conditioned by number, and equally so for

<sup>4</sup>In accord with the 2017 ‘Deklaracija o zajedničkom jeziku’ (<http://jezicinalacionalizmi.com/deklaracija/>), I treat Serbo-Croat (hbs) as a pluricentric language, like English or German, with four standards: Bosnian (bos), Croatian (hrv), Montenegrin (cnr), and Serbian (srp); in ISO terms it is a macrolanguage. See Corbett and Browne (2018) for a linguistic outline, Bailyn (2010) for an experimental translation study of the degree of difference of two of the standards, and Bugarski (2012, 2019 and references there) for the complex issues of sociolinguistic background and language status.

<sup>5</sup>The particular labelling of inflection classes is, of course, simply a convenience. Here are the labels I use (top line), with examples, and the corresponding traditional terms of Slavists and Indo-Europeanists (bottom line).

label	I	II	III	IV
example (Serbo-Croat)	<i>prozor</i> ‘window’	<i>duša</i> ‘soul’	<i>stvar</i> ‘thing’	<i>selo</i> ‘village’
inflection in nominative singular	-∅	-a	-∅	-o
gender (unless pre-empted by semantics)	MASCULINE	FEMININE	FEMININE	NEUTER
origin of inflection class	<i>o</i> -stem	<i>a</i> -stem	<i>i</i> -stem	<i>o</i> -stem

Table i: Labels for inflection classes

nouns like *grad* ‘city’ and for *prozor* ‘window’. Number is the third variable here: there is no direct link between the augment (a purely morphological phenomenon) and the agreement; we would not argue that *grad* ‘city’ has an external split directly linked to its internal split. Rather both are linked to number. The effect of the third variable, number here, does mean that the data fit with the Extended Lexeme Consistency Principle. This is to be expected, since morphological and morphosyntactic factors align in many instances. But this is not our prime concern: rather we are interested in examples where there is demonstrably a direct linkage between internal and external splits, and where it therefore makes sense to analyse whether the splits are consistent.<sup>6</sup>

To clarify what is needed, here is an operationalization of how to identify the phenomenon of interest. The operationalization consists of the following steps for analysing a given lexeme:

1. specify the lexeme’s internal split  
that is, give a morphosyntactic feature specification (e.g. singular vs plural), or list the cells (e.g. genitive singular vs remaining cells), for internal behaviour A vs internal behaviour B (e.g. suppletive stem A vs suppletive stem B);
2. specify the lexeme’s external split  
that is, give the different external behaviours: external behaviour A vs external behaviour B (e.g. takes feminine agreement vs takes masculine agreement);
3. establish that external behaviour A is found together with internal behaviour A, and external behaviour B is found together with internal behaviour B;
4. check that this external split is not found with lexemes which lack the internal split (the example of the Serbo-Croat augment (4) fails at this point).

A worked example is presented in §3.1.

So far we have discussed how to demonstrate a linkage between an internal and external split, namely that they co-vary. When such a linkage can be demonstrated, we can ask further whether there is evidence to demonstrate directionality, namely that one split determines the other. As we shall see, there are instances where there is such evidence.

Given that these linkages need careful analysis, we should ask what type of evidence would be relevant, for examples which are less straightforward than the augment in (4). In order to establish which data count, there are four lines of argument.

<sup>6</sup>It is important to keep in mind that we are analysing lexemes here, as the name of the principle implies. This is to be distinguished from ‘split’ used in the sense of partitioning of the lexemes of a language, in addition to partitioning into parts of speech (lexical categories), as discussed in §1.2. For instance, we may split (additionally partition) verbs into modal verbs and other. But that is not an internal split, and so is not our concern here. Thus English modals differ in their morphology from other verbs, and have special syntactic behaviour (the NICE properties, Sag et al., 2020). That is, modals share special morphological and syntactic properties as compared to other verbs (just as verbs share special morphological and syntactic properties as compared to nouns). But this is a split of lexeme from lexeme (an inter-lexemic split), not an internal split. To qualify, according to our operationalization, we would need some part of the paradigm of modal verbs to determine different external behaviour from that of the rest of their paradigm. (Naturally the fact that these modals are defective, lacking non-finite forms, has an impact on syntax; but this is a general regularity. Missing forms cannot by definition appear, in general, not specifically in the case of the modals.)



The first is general plausibility: alternative explanations would require a degree of coincidence which is highly unlikely (§3). The other three arguments are more specific versions of this one. In §4 we look at overabundance, where alternative inflectional forms link to different external requirements; in §5 we move on to variation in time and space, and consider instances where differences in inflection are linked to different external requirements, as both vary over time or space. In §6 we consider the specific and revealing types of split in pluralia tantum nouns. §7 is devoted to the significance of all these splits, including a constraint on featural information, and an extension to the scope of the Agreement Hierarchy. As we analyse the examples, we shall find some evidence for internal splits determining external splits. Given this, we should bear in mind the Principle of Morphology-Free Syntax, according to which rules of syntax do not have access to purely morphological features (such as inflection class) nor to the internal structure of a word (such as whether it is split); it might appear that our instances represent challenges to the principle. As we shall see when we review the data in §7.4, the principle remains secure. General conclusions are presented in §8.

### 3 Argument 1: plausibility

In several instances, the relation of internal inflectional form with external requirement is simply too unlikely to be coincidental. The relation may be seen as unusual within the confines of the particular language and/or cross-linguistically. Some of our examples are heteroclites, and these can usefully be introduced together. ‘Heteroclite’ was used by classicists to mean different or irregular; thus Matthews (2007:323) ‘a noun whose inflection follows something other than the regular pattern is traditionally “heteroclite”’. In modern linguistics its range has been narrowed to indicate specifically those items whose paradigms are split between two or more inflection classes (see Kaye, 2015:1–7, 15–29).<sup>7</sup> Typically heteroclites have only internal splits: we shall be interested in the small subset of them which also have external splits. However, whatever the type of internal split, the key point in this section is that the relation between the internal and the external split is striking.

#### 3.1 Serbo-Croat *oko* ‘eye’

We begin with a clear-cut example of a noun with an internal split and a split in its external requirements:

---

<sup>7</sup>Interest in heteroclisis was stimulated by Stump (2006), and Maiden (2009) initiated discussion of its origins. Stump extends the use of ‘heteroclite’ beyond the irregular to include instances which are ‘highly systematic’ (2006:280); this makes sense, since ideally we should define phenomena independently of the number of instances. However, the examples we shall discuss are unique or belong to small classes (as in Noyer, 2005 and Kaye, 2015, for instance). This means that we can avoid the issue of whether potential heteroclites should not simply be assigned to a separate inflection class. Rather, in systems where there are inflection classes which are relatively distinct in form, and these classes have large numbers of members (Corbett, 2009a), we are fully justified in treating items as heteroclite when they combine forms from different inflection classes and are themselves few in number.

(5) Serbo-Croat *oko* ‘eye’ compared with regular nouns<sup>8</sup>

		‘skein of yarn’	‘eye’	‘bone’
SINGULAR	NOMINATIVE	klupk-o	ok-o	kost
	ACCUSATIVE	klupk-o	ok-o	kost
	GENITIVE	klupk-a	ok-a	kost-i
	DATIVE	klupk-u	ok-u	kost-i
	INSTRUMENTAL	klupk-om	ok-om	kost-i / košču
	LOCATIVE	klupk-u	ok-u	kost-i
PLURAL	NOMINATIVE	klupk-a	oč-i	kost-i
	ACCUSATIVE	klupk-a	oč-i	kost-i
	GENITIVE	klubāk-ā	oč-ī / oč-ijū*	kost-ī / kost-ijū
	DATIVE	klupc-ima	oč-ima	kost-ima
	INSTRUMENTAL	klupc-ima	oč-ima	kost-ima
	LOCATIVE	klupc-ima	oč-ima	kost-ima
	INFLECTION CLASS	IV		III
	gender from IC	NEUTER		FEMININE

\* *očijū* is the more frequent alternative for this noun (SANU, 1959)

The noun *oko* ‘eye’ has a clear internal split between singular and plural, and in two interrelated ways. First, it has an irregular stem alternation; the change of consonant  $k \sim \check{c}$  [tʃ] before *-i* is not a synchronic alternation in the inflectional morphology; a different alternation *would* be expected in the plural dative, instrumental and locative, as we see in *klupko* ‘skein of yarn’, namely  $k \sim c$  [ts]. Second, it is a heteroclite: we find inflections from two different inflection classes, class IV in the singular and class III in the plural.<sup>9</sup> While there are several thousand nouns in inflection class IV and in III (Tošović, 2016), there are just two nouns like *oko* ‘eye’; the only other noun that behaves in the same way is *uho* / *uvo* ‘ear’, with plural *ušī*.

<sup>8</sup>None of these nouns has a vocative distinct from the nominative. The regular inflections can be found in Browne and Alt (2004:33, 35) and Corbett and Browne (2018:347). There are thousands of nouns in *-o*, but ordinary nouns specifically in *-ko*, like *klupko* ‘skein of yarn’, are rare, as can be seen in Matešić (1965–1967). The only other noun like *oko* ‘eye’ is *uho* ‘ear’; a normal noun in *-ho* to compare with that is *pazuho* ‘armpit’; its inflection is similar to that of *klupko* ‘skein of yarn’, except that in the dative, instrumental and locative plural it may have the corresponding alternation  $h \sim s$  (hence *pazusima*) or no alternation (then *pazuhima*); for further information on this stem alternation see Browne (2021).

<sup>9</sup>The syncretic form for the dative, instrumental and locative plural is shared with inflection class IV, and indeed with inflection class I, but not with inflection class II, which has *-ama*. Why then is it appropriate to say simply that the plural inflections are from class III? There are two reasons. First, it is more economical to specify an inflection class for the whole plural than to posit an additional split. Second, if we add a relative clause with the relative pronoun in a direct case, we can see the gender: it is feminine, which matches the expectation from inflection class III but not from classes IV or I. The issue of forms shared across inflection classes will be highly significant when we discuss *dokument* ‘document’ in §4.2; see further discussion there.

Regular forms of the plural are found rarely, and usually in other meanings, like ‘mesh’ (SANU, 1959:425–426; Benson, 1971:364–365). However, regular forms exist only for the nominative, accusative and genitive (reference works typically just omit mention of the remaining forms, but Maretić, 1931:134 is explicit that there are no regular forms here). Where there are distinct regular forms and distinct agreement (nominative and accusative plural), we then find the neuter plural agreement, as would be predicted from the regular plural inflection.

What then of the external requirement? Gender assignment in Serbo-Croat depends first on semantics (nouns denoting females are feminine and those denoting males are masculine). That does not apply here. And second it follows inflection class (nouns of inflection class I are masculine, II and III feminine, and IV neuter). This is what we find with *oko* ‘eye’ and *uxo* ‘ear’. We expect class IV nouns to take neuter agreement, and class III to take feminine agreement, and so these two nouns switch gender between singular and plural (as Wechsler & Zlatić, 2003:40 put it, they co-vary ‘in lockstep across the singular/plural divide’):

Serbo-Croat

- (6)    *njezin-o*            *ok-o*  
          her-SG.NOM.N eye-SG.NOM  
          ‘her eye’
- (7)    *njezin-e*            *oč-i*  
          her-PL.NOM.F eye-PL.NOM  
          ‘her eyes’

We should now consider the strength of the evidence provided by the co-occurrence of these two splits. The first point is that we can find the two types of split independently. We can have an internal split according to number, with no external effect. Within Serbo-Croat, in (5), we saw the split in the inflection of *grad* ‘town’, which has an internal split conditioned by number but this internal split has no additional external effect; for agreement purposes it behaves like any normal noun. Similar examples outside Serbo-Croat are not hard to find; for example the split in Slovenian *človek* ‘man, person’ plural *ljudje* (Corbett, 2007a:30) has no external effect. We can also find the external split without the internal. Within Serbo-Croat, as noted earlier (Table 1), there are nouns like *gazda* ‘landlord’ and *vladika* ‘bishop’, which are masculine in the singular but masculine and feminine in the plural (with Agreement Hierarchy effects, that is, agreements vary according to the agreement target, see §7.2). These have an external split but no internal split.

We see that the particular internal split of *oko* ‘eye’ has analogues both within Serbo-Croat and cross-linguistically. Similarly, parallels to the external split exist within and outside Serbo-Croat. The *linkage* of this internal and this external split is, however, rare within Serbo-Croat. There are just two nouns which behave this way, *oko* ‘eye’ and *uho* ‘ear’. It is not even all the paired body parts, since *ruka* ‘hand, arm’, and *noga* ‘foot, leg’ behave normally (internally in terms of inflection and externally in terms of agreement).

To check that this example meets our definition, let us run through the operationalization from §2.

1. specify the lexeme’s internal split:  
       singular A vs plural B (evidence: stem alternation and inflection class);
2. specify the lexeme’s external split:  
       neuter agreement A vs feminine agreement B;
3. establish that external behaviour A is found together with internal behaviour A,  
    and external behaviour B is found together with internal behaviour B:  
       yes: neuter agreement A is found with singular inflection A,  
       and feminine agreement B is found with plural inflection B.

4. check that this external split is not found with lexemes which lack the internal split:

yes: one other item shares the external split (*uho* ‘ear’) and it also has the internal split.

Thus *oko* ‘eye’ is indeed an example of what we are looking for; the operationalization makes that clear. A further argument that this link between the internal and external split is not coincidental is that it has persisted over time. The link goes back to the earliest attestations of Slavonic: in Old Church Slavonic (*chu*), *oko* ‘eye’ and *uxo* ‘ear’ were neuter in the singular; *oči* and *uši* were then the corresponding dual forms, and they took mainly feminine agreement (Vaillant, 1964:111–112, 168–169; Olander, 2015:189–191). According to Vaillant, with paired body parts the dual forms tended to supplant the plurals, which were rare. It is the dual forms *oči* and *uši* which have survived in Serbo-Croat as the plural forms. Thus this internal-external split is stable: it has persisted for some thousand years.<sup>10</sup>

We can go further; it is not just that the two splits are linked. It is reasonable to conclude that the external split is determined by the internal split. The inflection class of the singular (IV) determines the gender of the singular (neuter), and the inflection class of the plural (III) determines the gender of the plural (feminine). This follows the gender assignment system of Serbo-Croat, mentioned earlier, which has been justified independently of the special case of *oko* ‘eye’ (Corbett, 2009b:152); in the full system, four inflection classes map onto three gender values. With *oko* ‘eye’ the directionality is particularly clear. We see this if we try the opposite prediction, gender to inflection. Suppose we were to specify the plural as irregularly feminine, and then attempt to predict the inflection class from that. We would expect to find the majority class II, and this would incorrectly predict the plural in *-e*, namely *\*oke*, like *ruke* ‘hands, arms’. However, by specifying the plural as irregularly belonging to inflection class III, and predicting the gender value from this, we obtain the right prediction, namely feminine gender.

Serbo-Croat *oko* ‘eye’ demonstrates a linkage between an internal and an external split; furthermore it shows the directionality: the internal split determines the external split. I have spelled out this example in some detail; this will allow further cases to be presented more briefly.

### 3.2 Latin *balneum* ‘bath’

This is a venerable example of a heteroclite, cited by Baerman (2007:16), and discussed over two millennia earlier, by Varro (de Melo, 2019). *Balneum* ‘bath’ is com-

<sup>10</sup>In Slovenian, Serbo-Croat’s closest relative, *oko* ‘eye’ is neuter, *oči* ‘eyes’, the old dual, is retained as a plural, and is feminine. There is an innovative dual *očesi*, and here the agreements are syncretic between feminine and neuter (Harrison, 2009:32). When the dual is used in Slovenian is a complex issue (Priestly, 1983:440–441), and there is variation here (Herrity, 2000:37–39, Dvořák & Sauerland, 2006). For two eyes, the plural is typically used, with the dual largely reserved for instances strengthened by the inclusion of ‘two’ or ‘both’: Herrity (2000:31) gives these examples: *škiliti na lévo* (N) *okó* ‘to squint with one’s left eye’; *síve* (PL.F) *oči* (PL) ‘blue eyes’; *Ōn je slép na obé* (F/N) *očési* (DUAL) ‘He is blind in both eyes’. See also <http://www2.arnes.si/~lmarus/suss/arhiv/suss-arhiv-000095.html> for the plural *očesa* in transferred meanings. In Russian, *oko* ‘eye’ had a very different, but equally interesting fate, being largely replaced, first through invasive suppletion and then by regularization of the paradigm, for which see Chumakina et al. (2004:283–286), and references there.

parable to Serbo-Croat *oko* ‘eye’, except that its stem does not change. (The regular nouns in (8) are from Risch, 1977:231.)

(8) Latin (lat) *balneum* ‘bath’ compared with regular nouns

		‘yoke’	‘bath’	‘she-goat’
SINGULAR	NOMINATIVE	iug-um	balne-um	capr-a
	ACCUSATIVE	iug-um	balne-um	capr-am
	GENITIVE	iug-ī	balne-ī	capr-ae
	DATIVE	iug-ō	balne-ō	capr-ae
	ABLATIVE	iug-ō	balne-ō	capr-ā
PLURAL	NOMINATIVE	iug-a	balne-ae	capr-ae
	ACCUSATIVE	iug-a	balne-ās	capr-ās
	GENITIVE	iug-ōrum	balne-ārum	capr-ārum
	DATIVE	iug-īs	balne-īs	capr-īs
	ABLATIVE	iug-īs	balne-īs	capr-īs
	INFLECTION CLASS	II		I
	gender from IC	NEUTER		FEMININE

Varro’s comments on this heteroclite are given, translated and discussed by de Melo (2019:151, 514, tr. 515, 566 tr. 567; 1068, 1154–1155, 1257).<sup>11</sup> Varro does not comment on our main concern – the external requirements of *balneum* ‘bath’. However, the examples he gives illustrate nicely what is needed:

Latin (Varro: de Melo, 2019: 566)<sup>12</sup>

(9) ūn-um                  balne-um  
one-SG.NOM.N bath-SG.NOM  
‘one bath’

(10) pūblic-ae              balne-ae  
public-PL.NOM.F bath-PL.NOM  
‘public baths’

Thus the heteroclite *balneum* ‘bath’ takes its singular and plural forms from different inflection classes,<sup>13</sup> and the gender values required are those we would expect according to these inflection classes. The internal split and the external requirement (agreement) co-vary.<sup>14</sup> *Balneum* ‘bath’ is not quite unique: *epulum* ‘banquet, feast’, plural *epulae*, behaves similarly.

<sup>11</sup>Varro considers a difference in meaning between singular and plural, but de Melo (2019:151) takes issue with this. In any case, it is important to note that for many nouns number is not fully compositional (think no further than English *ground* vs *grounds*). Since we find lack of full compositionality with morphologically regular nouns, this is not an issue specific to heteroclites (I return to this point in §7.3).

<sup>12</sup>While the agreements in these examples happen to be alliterative, this is not always the case in Latin.

<sup>13</sup>There are also instances of the regular *balnea* ‘baths’; see Lewis and Short (1879) for the detail: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059:entry=balneum> accessed 25.09.2020.

<sup>14</sup>Note that there are shared forms in (8), in the dative and ablative plural; for *balneum* ‘bath’ I have attributed these to inflection class I and not to inflection class II. This is the simple solution: if *balneum* ‘bath’ belongs to inflection class I in the plural, its gender follows automatically from the normal assignment rules. And this fits the general picture where inflectional forms that are shared across inflection classes typically do not affect gender assignment. This point is taken up in §4.2.

### 3.3 Sɛlɛɛ inqorate genders

The large Niger-Congo family provides numerous examples of external splits linked to internal splits. The linguistic tradition for these languages is slanted towards diachronic investigation, in a way that tends to mask the regularities of the synchronic systems; for recent discussion see Babou and Loporcaro (2016:3–6) and Bach (2018:225–233). However, the notion of ‘inqorate genders’ (agreement classes which comprise a small number of nouns, and whose agreements can be readily specified as an unusual combination of forms available for agreement with nouns with the normal gender values) is discussed, particularly since researchers working on this family are often careful to give good data on the number of nouns in each gender.

We look particularly at Sɛlɛɛ (snw), based on Agbetsoamedo (2014); see also Di Garbo and Agbetsoamedo (2018:185–186). Sɛlɛɛ is spoken in Santrokofi in the Volta Region of Ghana, and is one of the Ghana-Togo-Mountain languages. The classification of these languages is difficult, but they belong within Atlantic-Congo and ultimately within Niger-Congo. Sɛlɛɛ has five main gender values, as shown by various agreement targets, particularly within the nominal phrase. Table 2 gives the markers on the proximal demonstrative (*-mle*) for illustration:

**Table 2** The main gender values of Sɛlɛɛ

gender	agreement of demonstrative		inflection class	example
	SINGULAR	PLURAL		
I	wɔ-	ba-	1/2	o-tii ‘person’;
IV	wɔ-	se-	1/4	o-nu ‘nose’
II	le-	ya-	5/8	le-foso-bi ‘lamb’
III	ko-	ya-	7/8	kɔ-la ‘python’
V	ka-	be-	3/6	ka-nya ‘mouth’

Note the syncretisms in Table 2: genders I and IV share their singular agreement (hence the ordering in the table), while II and III share their plural agreement. Under ‘inflection class’ the numbers indicate the prefixal markers on nouns for singular and plural (and the inflection class is a strong predictor of gender). As example (11) shows, the prefixes on the nouns can be phonologically similar to the agreement marker, but this is not invariably the case (as seen in the plural of (14) and (15)).

Sɛlɛɛ: Agbetsoamedo (2014:110, from Acts 1.14)

- (11) ba-tii      ba-mle  
       PL-person I.PL-DEM  
       ‘these people’

Each of these five gender values comprises a substantial portion of the noun lexicon (at least 10% in Agbetsoamedo’s corpus of 552 nouns). Our interest is in those which do not fit into the main gender values. Consider, for instance, the noun *kɔ-nɛɛ* ‘hand, arm’. To make clear the similarity to previous examples from other languages, I present it with other regular nouns, with its small paradigm set out in a column (hence rotated from Table 2).

(12) Selɛɛ *kɔ-nɛɛ* ‘hand’ compared with regular nouns

	‘machete’	‘hand’	‘stomach’
SINGULAR	<i>kɔ-pa</i>	<i>kɔ-nɛɛ</i>	<i>ka-futu</i>
PLURAL	<i>a-pa</i>	<i>n-nɛɛ</i>	<i>n-futu</i>
INFLECTION CLASS	7/8		3/6
gender from IC	III		V

When singular, *kɔ-nɛɛ* ‘hand’ has the same inflection as nouns like *kɔ-pa* ‘machete’. When plural it takes the same marker as *n-futu* ‘stomachs’. Inflectionally, then, it is like the examples we have seen from other languages, except that it has a smaller paradigm. What then of the agreements? Here are the key examples:

Selɛɛ: Yvonne Agbetsoamedo (personal communications 27 Oct. 2020 and 1 Dec. 2020)

- (13) *kɔ-pa*      *ko-mle*      *a-pa*      *ya-mle*  
 SG-machete SG.III-DEM      PL-machete PL.III-DEM  
 ‘this machete’      ‘these machetes’
- (14) *kɔ-nɛɛ* *ko-mle*      *n-nɛɛ* *be-mle*  
 SG-hand SG.III-DEM      PL-hand PL.V-DEM  
 ‘this hand’      ‘these hands’
- (15) *ka-futu*      *ka-mle*      *n-futu*      *be-mle*  
 SG-stomach SG.V-DEM      PL-stomachs PL.V-DEM  
 ‘this stomach’      ‘these stomachs’

These examples show that *kɔ-nɛɛ* ‘hand’, when singular, not only has the same marker as nouns like *kɔ-pa* ‘machete’, but also takes the same agreement. When plural it takes the same inflection as *n-futu* ‘stomachs’, and the same agreement. In other words, it is like a gender III noun when singular and a gender V noun when plural. Why do we not simply add another gender value? When we look for similar nouns we find just three more: *kɔ-kpa* ‘leg’ (plural *n-kpa*) and their derived diminutive/derogatory nouns *ka-nɛɛ-nyi* or simply *ka-nɛɛ* ‘tiny hand/arm’ and *ka-kpa-nyi* or *ka-kpɛɛ* ‘tiny leg’. Thus there are only four nouns of this type, out of Agbetsoamedo’s corpus of 552 nouns. Seen in this way, the data from Selɛɛ are fully comparable to those from Serbo-Croat and Latin (and there are further examples in Selɛɛ, for which see Agbetsoamedo, 2014). The nouns in question are heteroclites, in that they take material from two different inflection classes (the noun paradigms are small in Selɛɛ), and the resulting gender values are inquorate, with just four members.

The Selɛɛ situation is not unusual; there are numerous comparable examples within Niger-Congo, and I list just three here. In Noni (nhu), an Eastern Beoid language of Cameroon, there are just six nouns with an irregular gender pairing (an external split), as detailed in Hyman (1981:8). We do not recognize an additional gender value (they are inquorate). Of these six, four have irregular plurals (an internal split). For Cicipu (awc), a Benue-Congo language of northwest Nigeria, McGill (2009:241–252) provides a clear discussion, giving the number of nouns in each gender; again the internal and external splits line up. Finally, Sagna (2019:594–597) describes Eegimaa (Banjal, bqj), where there are two inquorate genders with one or

two members, where an unusual internal split (inflection) co-varies with an external split (agreement).

### 3.4 Scottish Gaelic *muir* ‘sea’

A remarkable instance is found in Scottish Gaelic (gla). In some dialects, such as that of Lewis, we find a change in gender value, conditioned by case. The item in question is *muir* ‘sea’ (Lamb, 2008:206):

(16) Scottish Gaelic *muir* ‘sea’, with definite article

	SINGULAR	GENDER (SINGULAR)	PLURAL
NOMINATIVE	am muir	MASCULINE	na mar-annan
DATIVE	a’ mhuir	MASCULINE	
GENITIVE	na mar-a	FEMININE	

The paradigm is internally split, with the genitive singular having an unpredictable form (irregularity in Scottish Gaelic nouns usually targets the genitive); the expected genitive singular would be: *a’ mhuir* (definite), and the indefinite *muir* (William Lamb, personal communication 9.10.2020). And this irregular cell brings an external split, in requiring a different gender value, as will be demonstrated. Use of the dative is restricted to government by some prepositions; the gender value here can be established only for older speakers. This is because the evidence for the gender value in the dative is adjectival agreement, but as William Lamb points out (personal communication) for younger speakers this particular gender distinction tends to be lost. Hence we concentrate on nominative and genitive. The stem *mar-* is also found in the plural. However, Scottish Gaelic does not distinguish gender in the plural; this is why gender values are given after the singular in (16).<sup>15</sup>

Consider now the agreement data:

Scottish Gaelic (Joan MacDonald consultant, from Lewis; William Lamb, personal communications, November 2020)

- (17)   tha       am               muir       thall               an-sin  
           be.PRS DEF.NOM.M sea.NOM over.STATIVE there  
           ‘the sea is over there’
- tha       e           /\*i       thall               an-sin  
           be.PRS 3SG.M /\*3SG.F over.STATIVE there.  
           ‘it’s over there’

<sup>15</sup>The origin of the split in *muir* ‘sea’ is that the noun was originally neuter, before this gender value was lost in Celtic (Mackinnon, 1910:302). In Oftedal’s classic account of the dialect of Leurbost, Isle of Lewis, he reports that there *muir* /muð/ ‘sea’ and also *talamh* /taLu/ ‘earth, land, soil’ are masculine in the nominative singular, and feminine in the genitive singular (1956:180). In that dialect, gender is not distinguished in the dative singular. Dialects vary here (see Dwelly’s dictionary 1902–1911), and these instances are being regularized (Mark, 2003:443), but there are several independent reports of the remarkable behaviour of these nouns. I am grateful to Bernard Comrie for pointing me to Oftedal’s grammar. See also Iosad and Lamb (2020) for a dialectometric analysis of Scottish Gaelic morphology, devoting particular attention to nominal phrases.



- (18)    tha        fuaim na                mara        àrd    an-seo  
          be.PRS sound DEF.GEN.F sea.GEN high here  
          ‘the sound of the sea is high (loud) here’
- feum-aidh        gu bheil        e        /\*i        faisg oirnn  
          must-NON\_PST to be.PRS.DEP 3SG.M /\*3SG.F close on.us  
          ‘it must be close to us’

For these examples, the first part is as quoted previously (Corbett, 2015a:170) and here we do indeed see a change of gender from *am muir* ‘the sea’ (masculine) to *na mara* ‘of the sea’ (feminine), as shown by the gender agreement of the article. This follows what is traditionally stated. However, I take being of a particular gender to mean controlling agreement in that gender value consistently (that is, under all syntactic conditions, for all agreement targets); this leads us to seek further evidence. Relative clauses provide no data. But when we consider the agreement of personal pronouns, as in the second part of these examples, we see that in both instances we have the masculine pronoun, with the feminine being unacceptable. Thus *muir* ‘sea’ does not simply change gender; rather, in the genitive (example (18)), it is a hybrid: it is feminine for attributive agreement but masculine for agreement of the personal pronoun.

The data are significant for how featural information is specified. Nevertheless, it is hard to tease out the different possibilities here. I therefore defer further discussion to §7.1.2 and §7.1.3, by which point further key data will be available.

### 3.5 Serbo-Croat *dete* ‘child’, *deca* ‘children’

The noun *dete* ‘child’, with the plural *deca*, is a dramatic instance of linked internal and external splits. This is its paradigm (Ekavian forms):

- (19)    Serbo-Croat *dete* ‘child’ (compared with regular nouns)

		‘button’	‘child’		
SINGULAR	NOMINATIVE	dugm-e	det-e		
	VOCATIVE	dugm-e	det-e		
	ACCUSATIVE	dugm-e	det-e		
	GENITIVE	dugm-et-a	det-et-a		
	DATIVE	dugm-et-u	det-et-u		
	INSTRUMENTAL	dugm-et-om	det-et-om		
	LOCATIVE	dugm-et-u	det-et-u		
PLURAL	NOMINATIVE	dugm-et-a	dec-a	žen-a	SINGULAR
	VOCATIVE	dugm-et-a	dec-o	žen-o	
	ACCUSATIVE	dugm-et-a	dec-u	žen-u	
	GENITIVE	dugm-et-ā	dec-e	žen-e	
	DATIVE	dugm-et-ima	dec-i	žen-i	
	INSTRUMENTAL	dugm-et-ima	dec-om	žen-om	
	LOCATIVE	dugm-et-ima	dec-i	žen-i	

The singular of *dete* ‘child’ follows the small class of nouns like *dugme* ‘button’.<sup>16</sup> However, its plural is quite different. It has a different stem (and  $t \sim c$  [t̥s] is not a regular alternation in inflection). But more significantly, the inflections are not normal plural inflections; rather they match a *singular* inflection type (class II), as shown by the noun *žena* ‘woman’ in (19). As a result, the ‘wrong’ distinctions are available: *deca* ‘children’ has a distinct form of the vocative (something that no regular plural noun has), and its instrumental is also distinct, while for regular plurals it is syncretic with the dative and locative (Corbett, 2007b:39).

This striking internal split is clearly linked to an unusual external split. In the singular the agreements are neuter, since the other nouns which inflect according to this class are neuter in the singular. But in the plural the agreements of *deca* ‘children’ are unusual and complex (Corbett, 1983:76–93; Wechsler & Zlatić, 2003:50–60, 206–219, 2012; Hristov, 2013:336–341, and references there). There are situations where unambiguously feminine singular agreement is found (there are several instances of these examples on the web):

- (20)    vid-im    sv-u                dec-u  
           see-1SG all-SG.ACC.F children-ACC  
           ‘I (can) see all the children’

Then there are others where the agreement is clearly plural:

- (21)    dec-a                su                doš-l-a  
           children-NOM AUX.3PL come-PST-PL.N  
           ‘the children have arrived’

In (21), the auxiliary is unambiguously plural, and there are good arguments for saying that the gender/number form of the participle *došla* is neuter plural (first, the evidence from conjoining, and second the constraint of the Predicate Hierarchy, Corbett, 1983:77–78 and 87). If personal pronouns are taken as agreement targets, then here we find neuter plural *ona* or masculine plural *oni*, dependent on the type of reading, with feminine plural possible if the children are all girls (Wechsler & Zlatić, 2003:51, 200, 205–211). In sum, *deca* ‘children’ can control different types of agreement; they differ according to the agreement target, and their distribution is subject to the Agreement Hierarchy (Corbett, 1983:76–88). The special interest of the effect of the case of the agreement target is discussed in Corbett (under review). The key point is that the remarkable agreement possibilities of *deca* ‘children’ are directly attributable to its remarkable internal split.<sup>17</sup>

<sup>16</sup>I give the regular plural forms of *dugme* ‘button’ to provide the clearest contrast; however, there is also the ‘collective’ plural *dugmad* ‘buttons’, which is more frequent. In fact, this class of nouns is characterized by non-standard relations between singular and plural (Stevanović, 1970:181–184; Pišković, 2011:154).

<sup>17</sup>For additional variation see Pišković (2011:241–243); for new data (including a survey of 42 consultants) on the partly similar noun *brat* ‘brother’, plural *braća*, particularly when in the secondary predicate position, see Despić (2017:262–264).

### 3.6 What we learn from the five examples for the plausibility argument

The case studies in §3 come from a range of languages. In each there is co-variance between the internal split and the external split. Furthermore, there is evidence as to the nature of this linkage. This is clearest with Serbo-Croat *oko* ‘eye’. The internal split involves an irregular (semi-suppletive) stem alternation. The resulting set of the noun’s inflected forms do not match one of the regular inflectional patterns. This indicates that it is the internal split which determines the external one, namely the split in gender. Specifying an irregular set of gender values would not lead to the prediction of the formal irregularity, but predictions from inflection class to gender, for the parts of the split, are straightforward. Serbo-Croat *dete* ‘child’, *deca* ‘children’ is largely comparable. Again there is an irregular (semi-suppletive) stem alternation, and the set of inflected forms is highly unusual, showing that it is the internal split which determines the external one, namely the split in gender and number. The prediction from inflection class to gender in the singular is normal; in the plural, the strange set of agreements is predictable to a limited degree. Scottish Gaelic *muir* ‘sea’ belongs next, since it too has an irregular (semi-suppletive) stem alternation, showing a morphomic pattern. The evidence is harder to interpret, but it appears that the internal split determines the external one. Latin *balneum* ‘bath’ and Sælc inquirate genders also show co-variance between the internal split and the external split. They do not provide an additional argument for directionality. That is, in each language there is a good argument that in general gender depends on inflection class (except for nouns whose gender is assigned according to semantic principles); however, the nouns in question, where only the pairing of inflection classes is irregular, offer no additional argument.

## 4 Argument 2: overabundance

Items which show overabundance (Thornton 2019a, 2019b) provide an additional type of argument. In the canonical situation each lexeme has a single realization for each featural specification; overabundance is the situation in which a given lexeme has more than one possibility in a particular cell or cells. For example, English *burn* has the past participle *burned* and *burnt*. If a lexeme has forms from different inflection classes, which are cell-mates, and these different possibilities also have different external requirements, this is a strong indicator that the two splits are linked. We focus on two main examples, which are interestingly different in nature.

### 4.1 Polish *ręka* ‘hand, arm’

Polish provides a cogent instance of an external split linked to internal overabundance; see the paradigm of *ręka* ‘hand, arm’ in (22):<sup>18</sup>

<sup>18</sup>I am grateful to Axel Holvoet for pointing out *ręka* ‘hand, arm’ and to Dávid Győrfi for all his help with work with consultants.

(22) Polish (pol) *ręka* ‘hand, arm’

	SINGULAR	PLURAL
NOMINATIVE	<i>ręka</i>	<i>ręce</i>
VOCATIVE	<i>ręko</i>	<i>ręce</i>
ACCUSATIVE	<i>rękę</i>	<i>ręce</i>
GENITIVE	<i>ręki</i>	<i>rąk</i>
DATIVE	<i>ręce</i>	<i>rękom</i>
INSTRUMENTAL	<i>ręką</i>	<i>rękami /rękoma</i>
LOCATIVE	<i>ręce / ręku</i>	<i>rękach</i>

We see two instances of overabundance. The one in the instrumental plural need not detain us; while *rękami* is the regular form and *rękoma* is highly irregular, the patterns of systematic syncretism mean that there is no relevant choice for the possible agreement targets. That is, there is no feasible external split. The instance of overabundance in the locative singular, on the other hand, is highly significant. The rest of the paradigm implies a noun of inflection class II, like *książka* ‘book’, and hence feminine. The consonant alternation in *ręce* is expected in the locative singular. On the other hand, the cell-mate *ręku* is totally unexpected. Its origin is in the dual, but the form has been taken over as a locative singular; this form would fit into inflection class I (predicting masculine) like *rok* ‘year’, for instance, or inflection class IV (predicting neuter), like *biurko* ‘desk’, given that we have a velar-final stem (Rothstein, 1993:698–699). In the locative singular, masculine and neuter share the same agreement form (distinct from the feminine).

We examine each locative singular with an attributive modifier. The numbers to the right are the number of examples in the National Corpus of Polish (<http://nkjp.pl/poliqarp/>), searched 28 May 2020, for exactly these phrases:<sup>19</sup>

Polish

(23) w lew-ej                      ręc-e                      109 examples  
in left-SG.LOC.F hand-SG.LOC  
‘in the left hand’

(24) w lew-ym                      ręk-u                      7 examples  
in left-SG.LOC.M/N hand-SG.LOC  
‘in the left hand’

The form that would be expected for a regular noun, that in (23), has feminine agreement, as would be predicted. Its irregular cell-mate in (24) has agreement forms which are clearly not feminine. Switching the agreements leads to unacceptable variants: consultants do not accept them, and there are no such examples in the corpus (checked 29 October 2020). Thus we see that there is one cell in the paradigm (the locative singular) where there are cell-mates which could induce different agreement; and we do indeed find different agreements. These are determined ultimately by the

<sup>19</sup>This choice of phrase ensures that only one hand could be involved, hence the form’s origin in the dual is not synchronically relevant.

inflection class of the cell-mates, through the gender assignment rules. Elsewhere in the paradigm agreements are as expected. Thus we have an internal split which determines an external split in agreement requirements.

The situation is actually more complex, in a way which appears not to have been noted previously. The locative singular forms are not simply feminine or masculine/neuter, as we see when we check other agreement targets. The locative cannot stand in subject position, so we cannot look for predicate agreement. We examine therefore the agreement of the relative and personal pronoun:<sup>20</sup>

- (25)    W lew-ym                      ręk-u                      trzym-am                      ten  
           in left-SG.LOC.M/N hand-SG.LOC hold.PRS-1SG this[SG.ACC.M]  
           długopis.  
           pen(M)[SG.ACC]  
           ‘I’m holding this pen in (my) left hand.’
- Złama-ł-a-m                      ją /                      \*go /                      \*je                      dwa miesiąc-e  
           (I) break-PST-SG.F-1SG 3SG.F / \*3SG.M / \*3SG.N two month.PL.ACC  
           temu.  
           ago.  
           ‘I (woman) broke it two months ago.’
- Ale teraz                      jest                      już                      zdrow-a /                      \*zdrow-y /  
           But now (it) be.PRS.3SG already healthy-SG.F / \*healthy-SG.M /  
           \*zdrow-e.  
           \*healthy-SG.N  
           ‘But already it’s fine.’

The personal pronoun is always feminine as in (25), and consultants had the same judgement for the relative pronoun. (A web example of the latter is: *w ręku, która służy do pisania* ‘in the hand, which (SG.F) serves for writing’.)

We shall consider further the significance of these data in §7.1.2 and §7.1.3. For now, note that the effect within the nominal phrase is not constrained by word order (*ręku* takes the same agreement for postposed as well as preposed modifiers).

## 4.2 Serbo-Croat *dokument* ‘document’

Here we investigate a group of nouns with a highly significant split. Tošović (2016:97) lists 15 nouns with similar behaviour; they differ slightly, and we shall focus on *dokument* ‘document’, since it occurs relatively frequently, and judgements are clear.<sup>21</sup> The data come mainly from consultant work with Ljubomir Popović (Belgrade), also from Andrija Petrović, Marko Simonović, advice from Wayles Browne,

<sup>20</sup>When searching for such examples in the corpus it became clear that with a phrase such as *w lew-ym ręk-u* ‘in the left hand’ typical continuations concern either the person involved, or the object in the hand. Yet consultants had clear judgements when given an appropriate context, namely that the hand had been broken and so being able to hold something in it was significant.

<sup>21</sup>Some have *dokumenat* (with ‘fleeting’ *-a-*) as the nominative and accusative singular; that possibility does not affect the argument here.

and from corpus work. As will become clear, there are differences according to area, and I shall concentrate on the situation in the east. Here are the forms:

(26) Overabundance: Serbo-Croat *dokument* ‘document’

	SINGULAR	PLURAL	
NOMINATIVE	dokument	dokument-i	dokument-a
ACCUSATIVE	dokument	dokument-e	dokument-a
GENITIVE	dokument-a	dokumenāt-ā	
DATIVE	dokument-u	dokument-ima	
INSTRUMENTAL	dokument-om	dokument-ima	
LOCATIVE	dokument-u	dokument-ima	
	IC I → M	IC I → M / IC IV → N	

We see that in the nominative and accusative plural, this noun is overabundant. We find the forms expected for a class I noun like *prozor* ‘window’ (as in (4) above), but also the forms for a class IV noun like *klupko* ‘skein of yarn’ in (5) above. This internal split is linked to an external split: the internal split and the agreements co-vary. Class I nouns are masculine and *dokumenti* takes masculine plural agreement, class IV nouns are neuter, and *dokumenta* takes neuter agreements, as these examples show:

(27) ov-i                      dokument-i                      su                      vrlo važn-i  
DEM-PL.NOM.M document-PL.NOM be.3PL very important-PL.NOM.M

(28) ov-a                      dokument-a                      su                      vrlo važn-a  
DEM-PL.NOM.N document-PL.NOM be.3PL very important-PL.NOM.N  
‘these documents are very important’

We see that the different inflections in (27) and (28) give rise to different agreements: the internal split (overabundance in inflection class) co-varies with an external split (different agreements required). In an early discussion of *dokument* ‘document’ and the few similar nouns, Ignjatović (1963:216–219) provides textual examples from 1879–1960 and agreement targets always take the gender value implied by the inflection class, as indicated in (26) above.

We have examined the overabundant cells in (26), and seen that the internal alternatives co-vary with the external alternatives. We must now consider the remaining part of the plural paradigm, the oblique cases, whose forms are shared by inflection classes I and IV. There are two reasonable hypotheses here.

Hypothesis I:

the unusual behaviour of this noun (and similar ones) is restricted to the nominative and accusative plural. It is basically a class I noun, and there is nothing unusual about its oblique plural cells.

Hypothesis II:

gender is by default a property of the lexeme. If this default is overridden, we then expect gender to be a property of the number sub-paradigms. The impli-

cation is that the plural sub-paradigm would be masculine or neuter, and there should be a difference in agreement with the oblique case forms.

For choosing between these two hypotheses, attributive modifiers will not help, since they show systematic syncretism across the genders in the oblique plural. However, we can use the relative pronoun, in a direct case, to tease apart the two hypotheses:

- (29) Govorili su o **tim** dokumentima ‘they spoke about those documents ...’  
 (29a) ... koj-i su pored tebe  
       ... which-PL.NOM.M be.3PL beside you  
 (29b) ... koj-a su pored tebe  
       ... which-PL.NOM.N be.3PL beside you  
       ‘... which are beside you’

Both forms are accepted. This means that the noun, when in the oblique case values in the plural, can be masculine or neuter, as in the direct cases. This remarkable situation is confirmed by numerous examples found in the Serbian Web Corpus (srWaC, accessed 23.11.2020). We have masculine and neuter possible through the plural: in the direct cases with overt overabundance, and in the oblique cases with ‘covert’ overabundance. It is not that the oblique cases are simply unspecified for gender, since the feminine relative pronoun *\*koje* is excluded. All this shows that Hypothesis II is correct.

Consultants suggest that the forms in *-a* and the associated neuter agreement are more prevalent in the east than the west. This makes sense, since Ignjatović (1963:218) points out that the *-a* plural is ultimately of Latin origin, entering the language through administrative, legal or ecclesiastical terminology.<sup>22</sup> Serbian was traditionally open to borrowings where Croatian would create neologisms (Klajn, 2001:90–91). Nevertheless, we find several instances also in the west: in the Croatian National Corpus (Hrvatski nacionalni korpus v 3.0 beta, accessed 23.11.2020), and in the Croatian web corpus (hrWaC, accessed 23.11.2020).

From the perspective of borrowing too, the phenomenon is noteworthy. Nouns originally from Latin, with singular in *-um* and plural in *-a*, of neuter gender, were borrowed as internationalisms (via French or German, for instance). The singular *-um* is not a Serbo-Croat inflection, so it was either treated as a part of the stem, or dropped (the latter especially in the west).<sup>23</sup> Either way this gave a singular stem fitting readily into inflection class I. The plural *-a* is a regular inflection (from inflection class IV), but not one that previously combined with the singular of inflection class I, hence there arose the combination of heteroclisis and overabundance we have analysed. This is something outside the current typologies of borrowing (Arkadiev & Kozhanov, 2021). Consider, in particular, Gardani’s detailed typology (2020:270–272), based on the established notions of MATerial and PATtern borrowing. In our example we find MAT borrowing of the specific items, but no PAT bor-

<sup>22</sup>See also (see also Klajn, 1998:69–70); I thank Jasna Vlajić-Popović for bringing this reference to my attention.

<sup>23</sup>Vujaklija (1980) gives numerous examples. The two possibilities are still evident today: in this report, forms both with and without *-um* are used (but sadly no plural): <https://www.nezavisne.com/novosti/drustvo/Sarajevo-ce-imati-prvi-krematorijum-u-BiH/540327>.

rowing. The pattern which results is novel (even though the *-a* inflection was well established in Serbo-Croat): a consonant-final stem with no inflection in the nominative singular is paired with *-a* in the nominative plural; given these forms, regular gender assignment gives a masculine-neuter pairing. What is special here is that the borrowing triggered a pattern of covert overabundance: the native oblique plural inflections in the nouns in question are treated as belonging to two different inflection classes, within the same nouns, and controlling the two appropriate gender values.

### 4.3 What we learn from the examples showing overabundance

The Polish and Serbo-Croat examples are intersections of heterocclisis and overabundance. They differ in ways that will prove crucial for what we can learn from internal-external splits.

In the paradigm of the Polish noun *ręka* ‘hand, arm’, there is a clear split between the cells belonging to inflection class II and the locative singular variant form. This is ‘distinct’ heterocclisis. The key cell, the locative singular, has two available forms, two ‘cell-mates’, and they belong to two different inflection classes. Those inflection classes assign different values: inflection class II assigns feminine, while inflection class I assigns masculine and IV assigns neuter. And this is what we saw in examples (23) and (24): the form belonging to inflection class II, as the rest of the paradigm, induces feminine agreement, and that belonging to inflection class I / IV induces masculine / neuter agreement.

Serbo-Croat *dokument* ‘document’ is a different type of heterocclite in that there are cells which are ‘shared’ between the contributing inflection classes. The plural oblique case forms could belong to inflection class I, which would induce gender value masculine, or to inflection class IV, which would induce gender value neuter. And as examples (29a) and (29b) show, we find both gender values. Again, there is a straightforward assignment of gender value according to inflection class. Since the mapping of inflection classes to gender values is typically many to one, the evidence from these heterocclite nouns can be readily accommodated. If one tried to analyse gender as the predictor, then it is not clear how specifying an irregular gender value could lead to the right prediction of inflection class. Hence we see again that the internal split determines the external one.

At this point we should ask why such options are not more frequently available. Since the Serbo-Croat plural inflection *-ima* (for dative, instrumental and locative) is shared by inflection classes I, III and IV (the genitive differs), we can consider why we do not find alternative agreements for all nouns in these classes (not, of course, for attributive modifiers since they show syncretism for gender, but for the relative pronoun) in examples comparable to (29a) and (29b). The reason is that, by default, gender is assigned to nouns, as complete lexemes. Thus Serbo-Croat *grad* ‘city’ and *prozor* ‘window’ in (4) above inflect according to inflection class I, and so are exclusively masculine, even for forms that are shared with other inflection classes. This holds for the lexeme, and whether particular inflections are unique to a given inflection class, or are shared across classes, has no effect. *Dokument* ‘document’ is different in that it has cells with overtly overabundant forms which belong to different inflection classes. These are the nominative and accusative plural, where the



forms *dokumenti* and *dokumente* belong to inflection class I (assigning masculine), while *dokumenta*, which can be nominative and accusative plural, belongs to inflection class IV (assigning neuter). The remaining plural cells fit with inflection class I and with class IV. For this noun, then, we have an override to the usual pattern of assigning gender to the entire lexeme. This means that *dokument* is masculine, except that this specification is overridden for the whole plural sub-paradigm. In the plural sub-paradigm the overtly overabundant nominative and accusative induce masculine or neuter agreement, according to the inflection class selected. The remaining cells of the plural sub-paradigm are covertly overabundant; they fit with both inflection classes and take both the corresponding gender values; this is seen given an agreement target like the relative pronoun, which can make the relevant distinction. Conversely, in the singular, forms like *dokumentom* (instrumental) are ambiguous between inflection classes, but corpus work confirms that only masculine agreement is found (for instance, the masculine relative pronoun *koji*). Thus there is no override in the singular. Here the noun is of inflection class I, and hence masculine. All this shows again the general point that agreement deals in feature values, not in particular inflections or morphemes.

Polish *ręka* ‘hand, arm’ is a more restricted instance than Serbo-Croat *dokument* ‘document’, since the override is for the locative singular only. The rest of the singular sub-paradigm belongs unambiguously to inflection class II. Therefore, only the locative singular shows the split in external gender requirement. The point that we need to take further, once we have examined other relevant examples, is that these two examples differ in the nature of the feature specification. In the plural, the variant *dokumenta* ‘documents’ is straightforwardly neuter: all agreement targets stand in the neuter. In contrast, with Polish *ręka* ‘hand’ in the locative singular form *ręku*, targets differ. Attributive modifiers are masculine / neuter, while others are feminine. Rather than having a straightforward gender specification, it is a hybrid. We return to this issue in §7.1.2 and §7.1.3 below. Further examples, which underpin the evidence provided by overabundance, are given in Appendix 1; these involve Polish, and Italian and Italo-Romance.

## 5 Argument 3: variation in time and space

Variation through time and across space provides a third type of argument. Suppose that at the level of individual lexemes we can observe forms from one inflection class replacing those of another, and going hand-in-hand with this a change in the external requirement, then this is strong evidence for a linkage. Our evidence here comes from Asia Minor Greek (§5.1). Equally if there is variation across space, and the two types of split co-vary, this too is good evidence for a linkage, and here we examine data from Old Frisian (§5.2). Scattered remnants, preserving the internal-external link, provide further support, as shown by data from Slovenian (Appendix 2).

### 5.1 Asia Minor Greek heteroclites

The Greek varieties (grk) spoken in eastern Asia Minor until 1923 provide more extensive evidence of internal-external linkage than in most of the examples we have

discussed. We find suffixes invading paradigms where they did not figure originally. These suffixes combine a derivational element (originally diminutive, but this meaning had long been lost) and an inflectional element. This creates heteroclititic nouns, and their agreements link to the mix of forms. The data are from Petros Karatsareas (2011a:228–253, 2011b, and personal communications 7 and 8 Aug. 2020 and 21 Jan. 2021). The suffixes in question are the genitive singular *-iu*, the genitive plural *-ion* and the nominative / accusative plural *-ia*; these belong to an inflection class associated with neuter nouns, but they have entered nouns in inflection classes whose members are originally masculine or feminine. The conditions sound ideal for our purposes. However, when we come to the external split, we face two problems in Asia Minor Greek. First, in some varieties, notably in Cappadocian, gender distinctions have been lost. Second, in varieties that preserve gender distinctions to some degree, gender is not distinguished in the genitive plural. This means that the best place to look is the nominative / accusative plural *-ia*, where it has become part of paradigms that were otherwise masculine or feminine, and in those dialects that preserve gender distinctions, for instance in Pontic:

Pontic Greek (Drettas, 1997:129, glossing from pp. 110–111, 119, 124)

- |      |              |                   |                  |                    |
|------|--------------|-------------------|------------------|--------------------|
| (30) | o=           | ðróm-on           | ta=              | ðróm-æ (< ðróm-ia) |
|      | DET.SG.NOM.M | street-SG.NOM/ACC | DET.PL.NOM/ACC.N | street.PL.NOM/ACC  |
|      | ‘the street’ |                   | ‘the streets’    |                    |
|      |              |                   |                  |                    |
| (31) | i=           | pf-i              | ta=              | pf-ía              |
|      | DET.SG.NOM.F | soul-SG.NOM       | DET.PL.NOM/ACC.N | soul-PL.NOM/ACC    |
|      | ‘the soul’   |                   | ‘the souls’      |                    |

In (30) we have a noun that belongs to an inflection class inducing masculine agreement, but the nominative plural inflection *-ia* leads to consistent neuter plural agreement (not just of the clitic article as here, but with other agreement targets too, Petros Karatsareas, personal communication). Similarly in (31) the noun belongs to an inflection class that normally leads to feminine agreement, but with *-ia* in the nominative and accusative plural we again find consistent neuter agreement.

There are two key points to retain from the Greek data. First, the phenomenon is widespread in Pontic, but it is also found in other Greek dialects and also in the standard language (see sources above). So this is a more general phenomenon than several of those discussed earlier. And second, the agreements are consistently neuter plural (in dialects where gender is distinguished), for all agreement targets. No Agreement Hierarchy effects have been observed. Since we are dealing with just the nominative plural (and accusative syncretic with the nominative), how then does this fit with the data from Scottish Gaelic and from Polish? The point is that gender is not distinguished in the genitive plural, hence the gender of the nominative / accusative plural is the gender of the plural. For nouns which take *-ia*, then, this leads to neuter gender being specified for the ‘entire’ plural sub-paradigm (analogous to what we saw for Serbo-Croat *dokument* ‘document’). Indeed it is part of a trend for the neuter to expand (for non-human nouns) in Asia Minor Greek (Karatsareas, 2011b:114). We should not be misled by the small paradigm here: if a noun is neuter ‘only’ in the nominative and accusative plural, it is neuter in the plural. Hence these Asia Minor

Greek nouns fit well into the bigger picture: we see a singular-plural split, albeit in a small paradigm, hence the plural is consistently neuter.

## 5.2 Old Frisian *wīf* ‘woman, wife’

Old Frisian (ofs) *wīf* ‘woman, wife’ allows us to see internal splits (the paradigm is split between inflection classes) and external splits (in gender agreement) co-varying in step with each other. The data come from the careful study by Fleischer and Widmer (2016) and personal communications from Jürg Fleischer (4 and 6 October 2020). The texts were selected from classical Old East Frisian legal manuscripts, written originally between the end of the thirteenth century and the middle of the fifteenth century (Bremmer, 2009:13–14). The item of note, Old Frisian *wīf* ‘woman, wife’, is a hybrid. It takes a feminine personal pronoun. For the relative pronoun and the predicate, Fleischer & Widmer are not aware of any examples with clear gender agreement (2016:221). However, in attributive position there are examples of neuter and of feminine agreement. This is a considerable move to the feminine as compared to the situation in Old High German (goh, Fleischer, 2012:175, 177, 178, discussed in Corbett, 2015b), and indeed as compared to modern German, while staying in accord with the Agreement Hierarchy (§7.2). For the noun’s external requirements, then, we look at attributive agreement and the linkage to the inflection of this noun. The essential information is given in Table 3 (modified from Fleischer & Widmer, 2016:223):<sup>24</sup>

**Table 3** Inflection of Old Frisian *wīf* ‘woman, wife’ and attributive gender agreement

	<i>wīf</i> : ending	<i>wīf</i> : attributive agreeing form	<i>a</i> -stem: ending (NEUTER)	<i>ō</i> -stem: ending (FEMININE)	<i>i</i> -stem: ending (FEMININE)
NOMINATIVE	-Ø	NEUTER	-Ø	-e, -Ø	-e, -Ø
ACCUSATIVE	-Ø	NEUTER	-Ø	-e	-e, -Ø
GENITIVE	-e -es	FEMININE NEUTER	-es	-e	-e
DATIVE	-e	NEUTER, FEMININE	-e	-e	-e

Fleischer and Widmer (2016:223) give the inflections of *wīf* ‘woman, wife’, and for comparison the forms of the relevant inflection classes (neuter *a*-stems, feminine *ō*-stems and feminine *i*-stems), since the options are neuter and feminine.<sup>25</sup> In the nominative and accusative (syncretic for *wīf* ‘woman, wife’ but with some distinct forms elsewhere), the former system is preserved: no inflection (bare stem) and neuter agreement. In the dative case the inflections are ambiguous, shared across different

<sup>24</sup>I have changed the order of cases both for consistency with other data in the current paper and because, as Table 5 shows, the change is progressing from dative, to genitive and has not reached the nominative-accusative.

<sup>25</sup>These classes are according to Steller (1928:37, 38 and 40, respectively), which is a helpful orientation, though Fleischer and Widmer point out (2016:223) that Steller’s paradigms represent something of an idealization of the variation that is found in the manuscripts.

inflection classes. Here we find both the earlier agreement, neuter (syntactic agreement) and the innovative feminine agreement (semantic).

The key case for our purposes is the genitive. This case value shows that there is an internal split in the paradigm. We find inflections from different inflection classes, *-e* from inflection classes that would predict feminine gender, and *-es* from the original inflection class, leading to neuter gender. Fleischer and Widmer (2016:225) give the following examples of genitives (with unattested forms in standardized orthography):

#### Old Frisian

*a*-stem inflection with neuter agreement (not *\*ō*-stem inflection with neuter agreement)

- |      |               |              |                             |                           |
|------|---------------|--------------|-----------------------------|---------------------------|
| (32) | the-s         | wiu-es       | not attested: <i>*the-s</i> | wiv-e                     |
|      | DET-GEN.SG.N  | woman-GEN.SG |                             | DET-GEN.SG.N woman-GEN.SG |
|      | 'the woman's' |              |                             |                           |

*ō*-stem inflection with feminine agreement (not *\*a*-stem inflection with feminine agreement)

- |      |              |              |                              |                        |
|------|--------------|--------------|------------------------------|------------------------|
| (33) | sin-re       | wiw-e        | not attested: <i>*sin-re</i> | wīv-es                 |
|      | 3SG.GEN-F    | woman-GEN.SG |                              | 3SG.GEN-F woman-GEN.SG |
|      | 'his wife's' |              |                              |                        |

When *wīf* 'woman, wife' is inflected according to one inflection class (the *a*-stem inflection) we find only neuter agreement (32) in the nominal phrase, and when according to a different inflection class (the *ō*-stem inflection) there is only feminine agreement (33). Thus we see evidence for a linkage between inflection class and gender.

Thus far we have been generalizing over time and space. But we can do better, and look at the different manuscripts investigated. Table 4 is modified from Fleischer and Widmer (2016:227); I have arranged the manuscripts in chronological order (following Bremmer, 2009:13), running from the First Brokmer Manuscript (after 1276 but before circa 1300) to the Second Emsigo Manuscript (shortly after 1450).

In Table 4 the nominative/accusative shows a clear picture: the original inflectional form is preserved and attributive agreement is always neuter, in all the manuscripts. In the dative, the inflectional marker could be from different inflection classes, and we find several instances of the conservative neuter and innovative feminine agreement. In the genitive, the key case, we see a clear linkage: the agreement is always that predicted from the inflection class.<sup>26</sup>

<sup>26</sup>For the genitive, we need to bear in mind that Fleischer & Widmer's interest was in establishing the development of *wīf* 'woman, wife' as a hybrid. In looking for unambiguous evidence of gender, they noted this clear pattern: when the noun was marked with *-es*, attributive agreement was always neuter; and when marked with *-e*, agreement was always feminine. (Fleischer & Widmer reasonably restrict their consideration to neuter versus feminine.) Based on this regular distribution, they treated the inflectional morphology (for the genitive only, hence the brackets for attributive agreement in the table) as sufficient evidence for the gender of the noun for a given instance. To relate the evidence to our concerns, there are two key points to retain. First, all the instances where there is an attributive modifier follow the pattern

**Table 4** Old Frisian *wīf* ‘woman, wife’: examples with unambiguous indication of gender

case of noun		NOMINATIVE / ACCUSATIVE		GENITIVE		DATIVE	
marker		-Ø		-es	-e	-e	
attributive agreement		NEUTER	FEMININE	(NEUTER)	(FEMININE)	NEUTER	FEMININE
manuscript	Brokmer 1	1	-	-	2	-	7
	Riustring 1	4	-	3	-	-	7
	Brokmer 2	1	-	-	4	-	7
	Riustring 2	1	-	-	-	-	2
	Hunsingo 1	3	-	3	-	3	2
	Hunsingo 2	3	-	3	-	2	3
	Emsigo 1	3	-	2	1	-	4
	Emsigo 3	3	-	-	1	-	1
	Fivelgo	6	-	3	-	8	2
	Emsigo 2	-	-	1	1	-	1

In broad-brush terms, the Old Frisian hybrid noun *wīf* ‘woman, wife’ is moving towards becoming a noun of feminine gender. But the progression is not straightforward, and arranging the examples in date order, as in Table 4, does not suggest a simple development. However, the extant manuscripts are distributed in space as well as time. As Fleischer & Widmer show, the data make better sense if interpreted geographically, as in Table 5, modified from Fleischer and Widmer (2016:229):

**Table 5** Old Frisian *wīf* ‘woman, wife’: geographical distribution of examples from Table 4

	West		← →		East
manuscripts	Hunsingo 1,2	Fivelgo	Emsigo 1,2,3	Brokmer 1,2	Riustring 1,2
NOM./ACC.	N	N	N	N	N
GENITIVE	N	N	N and F	F	N
DATIVE	N and F	N and F	F	F	F

In Table 5 the same manuscripts are arranged now according to their presumed place of origin (based on Bremmer, 2009:16). Innovative feminine agreement starts with the dative, where the inflectional morphology is ambiguous, and then may appear also in the genitive. Moreover the development is least advanced in the manuscripts originating in western areas and most advanced in the east.<sup>27</sup> There is an apparent issue, in that the easternmost documents, Riustring 1 and 2, do not show

illustrated in (32) and (33) above. And second, instances of a noun in the genitive without an agreeing attributive modifier are rare, since in the absence of any other modifier the definite article will normally be found. Thus the data are completely consistent with the generalization that the inflection class and the attributive agreement are fully linked in the genitive (but the number of instances demonstrating that linkage is slightly smaller than the figure in Table 4).

<sup>27</sup>This development has come to its conclusion (the noun is simply feminine), in modern North Frisian (also in the East Frisian dialect of Wangerooge, now extinct), while in West Frisian (and Saterlandic) the noun remains a hybrid (Fleischer & Widmer, 2016:236–237).

instances of feminine agreement in the genitive. This apparent issue is actually grist to our mill in that, as Fleischer and Widmer (2016:229) explain, in these documents we do not find genitives in *-e*. In harmony with this inflectional difference, we do not find feminine agreement in these documents either.

The careful analysis of the Old Frisian legal texts by Fleischer & Widmer, for all the difficulties which such data present, reveals a clear picture, one which fits well with our typology. We see a noun whose inflectional morphology is split (with variability according to dialect area), and whose hybrid agreements are linked to this split. This is particularly notable in the genitive where forms from different inflection classes are found. These classes induce different gender values and, indeed, the gender of attributive modifiers co-varies with the inflection class. The pattern is seen most clearly when we look at the geographical distribution.

### 5.3 What we learn from the examples showing variation in time and space

Diachrony, as we saw with Asia Minor Greek heteroclites, and geographical distribution, as in Old Frisian *wīf* ‘woman, wife’, both provide clear evidence supporting the claim that the internal and external splits co-vary. Both too support the claim that the internal split determines the external. In Asia Minor Greek, it is the incursion of the specific inflection, from outside the paradigm, that leads to the internal split, and creates the external split (gender agreement). The fact that this occurs with nouns which were previously masculine and others which were feminine shows that this is first a matter of inflection and then of gender. With Old Frisian *wīf* ‘woman, wife’ we see a step by step shift in inflection class, with the inflectional form which is shared by the inflection classes allowing both gender values, and the incursive inflectional form bringing the new gender value (feminine). Given the distribution of the Old Frisian manuscripts, it is their geographical distribution that allows us to see this effect most clearly.

## 6 Argument 4: pluralia tantum nouns

Several of our case studies have involved few examples (even just one) in a single language; the fact that these rarities nevertheless support significant generalizations (as we see in §7) is therefore impressive. We now turn to something found in larger groups of nouns in many languages, namely pluralia tantum nouns. These nouns show a particular type of split, between the existing and missing part of the paradigm. Their properties have recently been analysed in detail (Corbett, 2019), so this section is short. What is important is that these nouns are of various types (not all are like *scissors*), and the differences prove significant for our purposes.

Consider first Russian *sani* ‘sledge’ (Zaliznjak, 1967/2002:57–61, 75–80), one of several similar nouns. Its behaviour is comparable to English *scissors*, and since it is found within a nominal system which inflects for (at least) six case values and two number values, the lack of singular forms is evident:

(34) Russian *sani* ‘sledge’

	SINGULAR	PLURAL
NOMINATIVE	---	san-i
ACCUSATIVE	---	san-i
GENITIVE	---	san-ej
DATIVE	---	sanj-am
INSTRUMENTAL	---	sanj-ami
LOCATIVE	---	sanj-ax

*Sani* ‘sledge’ has regular forms for the plural, and has an internal split, since it has no regular singular forms. And it has a linked external split, since it requires plural agreement:

Russian

- (35) èt-i                      san-i  
 DEM-PL.NOM sledge.PL.NOM  
 ‘this sledge’ / ‘these sledges’

Like English *scissors*, Russian *sani* ‘sledge’ can be used of one entity or more than one, but it always takes plural agreement. Indeed, the numeral *odin* ‘one’ is used in the plural with such nouns:

- (36) odn-i                      san-i  
 one-PL.NOM sledge-PL.NOM  
 ‘one sledge’

Such nouns lack a morphological singular, and this lack links to the syntax, where they are consistently plural too. Their mismatch is between their morphology and syntax on the one hand (where only plural is possible) and the morphosemantics of number on the other (since reference can be to one or more than one entity). Nouns like *scissors* and *sani* ‘sledge’ represent a common type, but it is not the only one. Thus Tsez *xexbi* ‘child(ren)’ has only plural forms (like Russian *sani* ‘sledge’). Externally, it behaves differently. For one child it takes singular agreement, and for more than one it takes plural agreement (Comrie, 2001:381–383, discussed in Corbett, 2007b:31–38 and 2019:55–59). It is unremarkable in terms of morphosemantics and syntax: it is only its inflectional morphology which is unusual. Tsez *xexbi* ‘child(ren)’ has an internal but no external split. Thus pluralia tantum nouns may have an external split linked to an internal split, as shown by English *scissors* and Russian *sani* ‘sledge’ (Type 4 in Table 1), or they may have an internal but no external split, as with Tsez *xexbi* ‘child(ren)’ (Type 2 in Table 1).<sup>28</sup>

<sup>28</sup>Analyses tend to be done, consciously or unconsciously, to avoid splits, and this makes good sense. But the different types of pluralia tantum nouns show that we should not draw conclusions on the basis of just one type. These nouns demonstrate that, while we often assume that the morphosemantic and morphosyntactic feature values associated with a given cell in a paradigm are the same (as they are with the Tsez example), they may well not be (as our Russian example shows). The distinction between morphosemantic and morphosyntactic features is discussed in Corbett (2012:49–50) and Spencer (2013:219–232); this distinction suggests that Stump’s interesting work on paradigm types needs further elaboration, requiring us to split his content paradigm, which is the ‘interface with syntax and semantics’ (Stump, 2016:104).

There is a further important distinction to be drawn here. Pluralia tantum nouns like *scissors* and *sani* ‘sledge’ take consistent plural agreement. But a plurale tantum noun can also be a hybrid. Karlsson (1968) reported that in Finnish (fin) proper nouns like *Yhdysvallat* ‘the United States’, *Filippiinit* ‘the Philippines’ normally took a singular predicate. In attributive position, however, both singular and plural adjectives were possible, with singular on the increase. This trend has continued so that according to Hannu Tammola (personal communications, 5 June 2017), the singular is now usual, though occasional plurals are found in attributive position. The fact that pluralia tantum nouns can be hybrids will prove useful in §7.3.<sup>29</sup>

Pluralia tantum nouns also provide a different argument bearing on directionality within the linkage of splits. We find examples as with English *scissors* and Russian *sani* ‘sledge’, where there is an internal split, and a co-varying external split. Then we find examples like Tsez *xexbi* ‘child(ren)’, where there is an internal but no co-varying external split. But we do not find the converse of Tsez, that is, a language where some nouns have to be lexically marked as taking only plural agreement, while having no internal split (that is, having a normal singular and plural), see Corbett (2000:66–67). This distribution of possibilities implies that it is the internal split determining the external one.

## 7 Significance: what we learn about features

When evaluating the significance of the examples we have analysed, we must bear in mind that the data are scarce. Indeed finding examples is like panning for gold. A majority have been found in Indo-European languages, simply thanks to the long history of previous work, leading to reference grammars and large dictionaries, which greatly improve the chances of finding relevant data. We also saw evidence from Selee (§3.3); there are further instances in Niger-Congo but of the same general type. We also noted briefly relevant data from Finnish (§6). A second property of the data is that nouns but not verbs are represented; related to this, our examples involve agreement rather than government.<sup>30</sup> We need a fuller range of languages here. The examples that have been found, despite these issues, are highly significant. First, they help establish the locus of featural information (§7.1), and indicate how the scope of the

<sup>29</sup>Modern Hebrew *be'alim* ‘owner(s)’, as analysed by Landau (2016), is also a remarkable hybrid, but it is not straightforwardly a plurale tantum noun; there is a singular form *ba'al*, but with rather different properties

<sup>30</sup>There are apparent examples of verbal government being split, with this being linked to an internal split. Verbs may require a different case frame according to their tense-aspect-mood, and these feature values may be realized variously. However, the examples to date do not qualify, for the reason discussed in §2, namely that there is a third variable. Tense-aspect-mood determines the case frame, irrespective of whether the verb has an internal, morphological split or not. There are (also only apparent) examples of internal-external links involving nouns and case. Various Dagestanian languages of the Tsezic group have two genitive case values, split according to the case value of the head noun: one genitive marks dependents in phrases headed by a noun in the absolutive, and the second when the head noun stands in any other case value (the obliques). These latter values attach to an oblique stem, which may be formed in a variety of ways. But again there is no direct linkage: the particular way in which the direct (absolutive) versus oblique opposition is realized on the head noun plays no role in the selection of the genitive: the split is determined purely featurally.



Agreement Hierarchy can be extended (§7.2). They require us to be clear about the role of frequency (§7.3), after which we can assess the importance of the data for lexicalism and morphology-free syntax (§7.4).

## 7.1 The locus of featural information

We can make sense of our diverse examples by starting from the assumption that lexical information is specified, by default, at the level of the lexeme. For instance, a lexeme has a particular gender value (and this may often be inferred from its inflection class). This is what we find in items with no external split. The interesting examples involve overrides to this general default, involving either a sub-paradigm (§7.1.1) or individual cells (§7.1.2).

### 7.1.1 Featural information at the level of the sub-paradigm

We saw that with Serbo-Croat *oko* ‘eye’ (§3.1), its singular is straightforwardly neuter and its plural feminine. This holds for all agreement targets. Similarly Latin *balneum* ‘bath’ (§3.2) is neuter in the singular and feminine in the plural. Serbo-Croat *dokument* ‘document’ (§4.2) has the complication that it is overabundant, having two possible forms for the nominative and accusative plural. As we saw, it is in fact the whole plural sub-paradigm which has the two corresponding gender values: even in the cases with a single (ambiguous) form, both gender values are available for agreement. Pluralia tantum nouns like Russian *sani* ‘sledge’ (§6) also belong here: they have consistent agreement involving the plural sub-paradigm.

Asia Minor Greek helps motivate the distinction drawn above. We saw examples where the nominative / accusative form is taken from a different paradigm, and induces neuter agreement. However, the remaining form, the genitive plural, does not distinguish gender. Hence once the nominative and accusative are specified as neuter, based on the inflection class of the form, that is essentially the specification of the plural sub-paradigm. Finally, consider Σελε (§3.3). There the paradigm consists of just two cells. Specifying the plural sub-paradigm is, in a sense, the same as specifying the plural cell. But in fact it makes a difference. If we adopt the analysis requiring the lesser override, treating Σελε as overriding the plural paradigm, then we might suggest:

Preliminary generalization 1 (to be refined in §7.1.3)

If there is an internal split at the sub-paradigm level,  
and this determines an external split,

then the agreement is straightforward: that is, a single value is involved.

As we shall see, this is in marked contrast to those splits where the cell level is involved. Nevertheless, the proposed generalization needs refining, since it does not account for Serbo-Croat *deca* ‘children’ nor Finnish *Yhdysvallat* ‘the United States’ (a problem remedied in §7.1.3).

### 7.1.2 Featural information at the level of the cell

Polish *ręka* ‘hand’ differs in that the single cell involved, the locative singular, shows overabundance. However, the unexpected gender agreement has also been established

to be hybrid – it varies according to target. Old Frisian *wīf* ‘woman, wife’ was known to be a hybrid, and the important study cited was designed to show progression of the semantically justified agreement form (the feminine). It proves significant for our story in that the change is proceeding cell by cell – we need a different specification for each cell. The nominative and accusative preserve the old form and the old, syntactic agreement. The dative form is ambiguous as to inflection class, and it allows both agreements. The main interest is in the genitive, where there are two forms available, and syntactic agreement (neuter) is found with one, and semantic agreement (feminine) with the other. Thus the split has to be specified for specific cells. Whenever neuter agreement is possible, this indicates a hybrid, since the personal pronoun takes feminine agreement. These examples are consonant with the following:

Preliminary generalization 2 (to be refined in §7.1.3)

If there is an internal split at the cell level,

and this determines an external split,

then the agreement is hybrid (it varies according to the agreement target).

The surprising Scottish Gaelic example, *muir* ‘sea’ could be classified in two ways: (i) there is a morphomic split (singular nominative and dative, vs singular genitive and the plural); (ii) since gender is not distinguished in the plural, the plural sub-paradigm should be ignored and the split is singular nominative and dative, vs singular genitive. It is tempting to take the first analysis, thereby expanding the typology. But the simpler solution is the second. For this we look at the genitive singular cell, since the plural cells are not relevant to the external split. As we noted, it had been established earlier that this cell had a different gender value, but it had not previously been pointed out that it is hybrid: the gender value varies across targets. This would fit under Preliminary generalization 2, and is the approach we take. If we were to take the other route, the morphomic split, then we would expand the first line of the generalization accordingly.

### 7.1.3 The constraint on featural information

We must be clear about the different types of split. Let us start from a canonical noun: this takes the same agreements to the degree possible; that is, it always takes the same gender value, and furthermore when singular it takes singular agreements, when plural it takes plural agreements, and so on. This holds true for all possible agreement targets, that is to say, it has a consistent agreement pattern (Corbett, 2006:11–12). Here the canonical type is unremarkable and there is no term for it. We need to distinguish two sorts of deviation from this canonical baseline, two sorts of split. First, there can be a split in that not all cells of the paradigm (not all featural specifications) behave alike; for instance, a noun may be neuter in the singular but feminine in the plural (like Serbo-Croat *oko* ‘eye’, as in §3.1). Second there can be a split in that different targets behave differently; a noun may be neuter for attributive modifiers but feminine for the personal pronoun, making it a hybrid (German *Mädchen* ‘girl’ would be an example). And while the canonical noun has neither of these splits, some have both, like Serbo-Croat *gazda* ‘landlord’ noted in §2, which has a singular-plural

split; the singular is masculine, but the plural is masculine / feminine, according to the target. Thus it is a hybrid, just in the plural (Corbett, 2015b:205–207 and references there).

All of our examples below have a split of the first type – there is a restriction to a part of the paradigm. Moreover, given our focus, in all of them I have argued that this split in agreement requirements is linked to their internal morphological split. So the question here is what we find externally: is the agreement consistent (the same for every target) or hybrid (different according to target, also called ‘mixed agreement’)? I shall be cautious, since the examples are relatively few. However, given how difficult they are to find, and how valuable they are, it makes sense to draw tentative conclusions, partly in the hope of stimulating further research. The generalization which covers the available data is that for a lexeme which has externally split requirements, if these are to be for consistent agreement (rather than hybrid), the split must involve a sub-paradigm (it must be motivated). That is to say, it is possible for Serbo-Croat *oko* ‘eye’, Latin *balneum* ‘bath’, Serbo-Croat *dokument* ‘document’, Russian *sani* ‘sledge’ and Asia Minor Greek heteroclitics (also even Σεξε *ko-nεε* ‘hand’, §7.1.1) to have a straightforward external requirement (consistent agreement), for both sides of the split, because the split is motivated (it involves the singular-plural distinction). This is not possible for Scottish Gaelic *muir* ‘sea’, Polish *ręka* ‘hand’ or Old Frisian *wīf* ‘woman, wife’, since here the split is not motivated (it involves individual cells, or arguably a morphomic pattern for Scottish Gaelic).<sup>31</sup> This is evident in Table 6.

**Table 6** The constraint on featural information

example	data	internal split: sub-paradigm (motivated) vs cell (not motivated)	agreement
Serbo-Croat <i>oko</i> ‘eye’	§3.1	sub-paradigm	consistent
Latin <i>balneum</i> ‘bath’	§3.2	sub-paradigm	consistent
Serbo-Croat <i>dokument</i> ‘document’	§4.2	sub-paradigm	consistent
Asia Minor Greek heteroclitics	§5.1	sub-paradigm	consistent
Russian plurale tantum <i>sani</i> ‘sledge’	§6	sub-paradigm	consistent
Serbo-Croat <i>deca</i> ‘children’	§3.5	sub-paradigm	hybrid
Finnish <i>Yhdysvallat</i> ‘the United States’	§6	sub-paradigm	hybrid
Scottish Gaelic <i>muir</i> ‘sea’	§3.4	cell <sup>a</sup>	hybrid
Polish <i>ręka</i> ‘hand’	§4.1	cell	hybrid
Old Frisian <i>wīf</i> ‘woman, wife’	§5.2	cell	hybrid

<sup>a</sup>Or arguably a morphomic pattern, see §7.1.2

Table 6 reveals a pattern: examples of splits with consistent agreement all involve a motivated internal split (sub-paradigm); those with an internal split involving a sin-

<sup>31</sup>The Italo-Romance evidence (see Appendix 1) cannot take us further here, since the paradigms involved are typically minimal, consisting only of a number distinction.

gle cell all have hybrid agreement. What then of the remarkable Serbo-Croat *deca* ‘children’? Here we have, arguably, a singular-plural split, and yet the noun is a hybrid. In brief (see §3.5, the detail is in Corbett, 1983:76–81), in attributive position we find feminine singular agreement in all cases, except in the nominative where it is ambiguous between feminine singular and neuter plural. The predicate requires plural agreement, and the gender agreement there is best analysed as neuter plural. The relative pronoun takes feminine singular or neuter plural (according to case), and the personal pronoun takes masculine, feminine or neuter plural (see Corbett, 1983:84–85, Wechsler & Zlatić, 2003:51).

We find a one-way implication: only lexemes with motivated splits *may* have straightforward external requirements (in this they are like canonical nouns with no split). The generalization is this:

Generalization 3 (replacing preliminary generalizations 1 and 2)  
 If there is an internal split,  
 and this determines an external split,  
 for agreement to be consistent,  
 it is a necessary, but not a sufficient criterion, that the split be a motivated one.

In the case of Serbo-Croat *deca* ‘children’, this noun could theoretically not be a hybrid, but it is. It has highly unusual properties which may underlie its behaviour here: two features (gender and number) are involved in its hybrid behaviour while typically hybrids involve just one. But mainly, the fact that it denotes humans means that its form-meaning mismatch is likely to make it a hybrid.<sup>32</sup> And pluralia tantum nouns illustrate the generalization well. Their split involves a sub-paradigm (the singular is missing), which is a necessary condition for consistent agreement: we find this with the Russian plurale tantum *sani* ‘sledge’. But it is not a sufficient condition, so they may be hybrids, as with Finnish *Yhdysvallat* ‘the United States’ (which is a hybrid so long as the plural attributive remains a possibility).

Several of our examples involve a split in gender agreement conditioned by number, and these data bear on a further issue. For consistent agreement in gender to be possible, the split must be conditioned by values of number (as the examples in Table 6 show). This generalization fits with various feature hierarchies which have been proposed over the years (e.g. Brown & Hoppisley, 2012:57–62). It is also relevant to the claim by Carstairs-McCarthy (1994:771), that gender mixture can be based only on number. By this he means that when there are more controller genders than target genders, the extra value is always based on number (for example, there can be a class of nouns which are masculine in the singular and feminine in the plural, which if large enough gives a third gender value, as with the Romanian neuter). It was suggested by Enger and Corbett (2012:303–306) that this claim does not hold, on the basis of data from Norwegian dialects, where gender agreement is split on the basis of definiteness, and from Scottish Gaelic *muir* ‘sea’, where gender agreement is split on the basis of case (§3.4 above). It was known that the Norwegian instance was a hybrid, but Scottish Gaelic *muir* ‘sea’ was believed to have consistent agreement,

<sup>32</sup>Recall too that there are hybrids which are split in terms of their external requirements but without an internal split. Serbo-Croat nouns like *gazda* ‘landlord’ noted above (§2) are internally consistent (there is no irregularity in their inflection), but they are hybrids, only in the plural.

something we have just shown is incorrect. Provided we restrict ourselves to normal gender values, those with consistent agreement patterns (which may well be what Carstairs-McCarthy intended), then till now his generalization holds.

Saving this generalization leads to a new issue. Scottish Gaelic *muir* ‘sea’, as just discussed, and Polish *ręku* ‘hand’ both have inconsistent agreement. Their external requirements suggest an Agreement Hierarchy effect, but to capture it would require us to extend the scope of the hierarchy, the issue to which we now turn.

## 7.2 Extending the scope of the Agreement Hierarchy

The Agreement Hierarchy consists of these positions, which we have already seen as relevant to some of our examples (see Fig. 1).

attributive > predicate > relative pronoun > personal pronoun

Fig. 1 The Agreement Hierarchy

The constraint is:

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 (that is, with no intervening decrease).  
(Corbett 1979, 2006:206–233)

Can we use the Agreement Hierarchy to cover the more unusual instances we have met? In terms of the target positions it is perfect. The question is whether the constraint applies. The constraint refers to ‘greater semantic justification’, that is a relative ranking of alternative agreement forms. This covers instances like Serbo-Croat *deca* ‘children’ (§3.5), where the possible agreements include feminine singular, neuter plural and masculine plural. We need to be able to say that neuter plural agreement (the ‘right’ number but the ‘wrong’ gender) shows greater semantic justification than feminine singular, but that masculine plural shows even greater semantic justification.

Consider now Polish *ręka* ‘hand’, where we saw that the unexpected locative singular form *ręku* can take masculine / neuter agreement, but only in attributive position. Once discovered, this restriction makes sense, since we do not expect case to ‘survive’ over longer syntactic stretches: the fact that the phrase is in the locative would not be a factor for gender and number outside the phrase. Yet this expectation is equally a reflection of the structural constraint of the Agreement Hierarchy. Inside the nominal phrase, then, we find masculine / neuter agreement, and outside it feminine. Is it reasonable to say that one of these has greater semantic justification? I suggest that the feminine has greater semantic justification, in that it facilitates access to the antecedent, which is of feminine gender (apart from one overabundant form). Turning to Scottish Gaelic *muir* ‘sea’ (§3.4), this is a masculine noun, except when the noun is genitive *and* the agreement target is within the nominal phrase. In previous analyses, hybrids have been discussed based on two types of mismatch: (i) meaning–meaning mismatches (English *committee*, where plural agreement matches

the more general semantic distinction of plurality of items as opposed to the less general notion of ‘collection’, which does not relate directly to the semantics of number in English); or (ii) form-meaning mismatches (for example, *mamma* ‘mum’ in the Nordreisa dialect of Norwegian, where the form does not match the meaning); see Corbett (2015b) for sources and for discussion of both types. Instances like Polish *ręka* ‘hand’ illustrate the third logical possibility, a form-form mismatch. The locative form *ręku* does not fit with the rest of the forms of this noun, and induces a gender value that is less general than that which applies to the rest of the noun.

It is important to recall that ‘greater semantic justification’ is a relative notion; it does not imply a large degree (any more than ‘longer than’ implies long); it runs from complete lack of semantic justification, to full semantic justification. Can we calibrate the agreements with Polish *ręka* ‘hand’ according to this scale? The use of feminine agreements with all forms of Polish *ręka* ‘hand’ has some function, in that this gender value is that of the noun as a whole and it allows *ręka* ‘hand’ to be tracked as the antecedent. The masculine / neuter does not allow this. If we place feminine agreement with *ręka* ‘hand’ on the semantic justification scale, clearly it scores very low, but surely higher than masculine / neuter. To this extent, then, the feminine arguably has greater semantic justification than the masculine or neuter. We can treat Polish *ręka* ‘hand’ as a split hybrid (similarly Scottish Gaelic *muir* ‘sea’); it falls under the constraint of the Agreement Hierarchy and we do not need to introduce additional machinery for it.

### 7.3 The role of frequency

The evidence has consisted mainly of small classes (in terms of types); thus Serbo-Croat *oko* ‘eye’ has just *uxo* ‘ear’ showing similar behaviour. But even here we must be careful to avoid some reasonable false connections. We start from the observation that we have seen the “usual suspects” among the evidence, for example certain paired body parts, as in the Serbo-Croat examples just mentioned. An obvious observation here would be that we are dealing with frequent nouns. But we need to distinguish between absolute and relative frequency (e.g. Corbett et al., 2001:202–203, Hippisley et al., 2014:392–393). It is not primarily that *oko* ‘eye’ is frequent as a lexeme (there are many items which are more frequent), rather its plural is frequent relative to its singular; and originally its dual was frequent relative to singular and plural. It is this relative frequency, I suggest, that has allowed it to maintain its heteroclitic status and its split in agreement for over 1000 years (Corbett, 2019:67–69). Having suggested that relative frequency is the more important factor here, we must also note the next potential source of confusion, namely that there is a correlation between relative and absolute frequency (Hay, 2001): in our instance, items with high relative frequency of the plural will also tend to be absolutely frequent.

Hay (2001) suggests that different meanings may also be associated with high relative frequency; she is mainly concerned with derivation, but the same argument should be considered for number. Here, again, we must keep the different factors apart: we may link relative frequency to semantic irregularity, and we may link it to morphological irregularity, but we should recall that singular and plural may diverge semantically, irrespective of any morphological irregularity (as with, for instance,

*ground* ~ *grounds*, *fund* ~ *funds*, and so on). These singular-plural effects have led some to suggest that plural nouns are distinct lexemes (for instance, Baayen et al., 1997:866–868); these authors were concerned with Dutch, and the arguments are less convincing when we consider larger paradigms (as with *oko* ‘eye’). While their arguments accommodate nouns like *oko* ‘eye’ with different gender values in singular and plural, they work less well in explaining why it is so very often the case that nouns have the same gender value across numbers. This issue again illustrates the importance of looking at larger systems, in this instance items with larger paradigms.

Hence we need to keep apart relative and absolute frequency, the possibility of non-compositional use of number, and the potentially confusing properties of small paradigms. If we take account of all these factors, the examples we have examined provide telling evidence bearing on the location of featural specifications.

## 7.4 Lexicalism and Morphology-free Syntax

Our data involve the specific behaviour of a limited number of lexemes, in some instances a single lexeme. Each one of these can be taken as an argument for lexicalist models. The issues, including the different combinations of principles involved in lexicalism, are carefully laid out by O’Neill (2016). In particular, we have seen quite specific morphology-internal mechanisms at work, which are different in kind from the workings of syntax (for another particularly cogent example see Feist & Palancar, 2021). However, we have gone to some lengths to show a linkage between internal and external splits. Furthermore, some examples have provided evidence that it is the internal split that determines the external one. We must therefore clarify how such instances are consonant with the Principle of Morphology-free Syntax (Zwicky, 1992:354–356). This principle ‘prohibits any rule of syntax from making reference to the internal structure of a word or to purely morphological features’ (O’Neill, 2016:244).

As a representative example, we take specifically *oko* ‘eye’ in Serbo-Croat (§3.1). We saw that this lexeme combines material from two inflection classes and that this internal split determines its external split in its gender requirement. As we saw in §3.1, the language has principles of gender assignment, which depend: (i) on semantics (nouns denoting females are feminine and those denoting males are masculine); and (ii) when the first do not apply, on inflection class (inflection class I implies masculine, II and III feminine, and IV neuter). Thus the morphological information required as part of the lexeme allows prediction of its gender, which is what the syntax has access to (the suggestion that the syntax might have access to inflection class in Serbo-Croat has been shown to be unnecessary, see Corbett, 2009b). The particular noun *oko* ‘eye’ follows exactly the same principles. Its morphology is split, this leads to two outcomes from the gender assignment principles, and these are what the syntax has access to. The interface is a featural one, as is shown particularly clearly by Serbo-Croat *dokument* ‘document’, where the key data involved the relative pronoun in a different case from its antecedent and with forms which had no phonological resemblance to it (§4.3). Specifically in terms of the Principle of Morphology-free Syntax as given above, the syntax does not need to make reference to the internal structure of a word; of course, different cells may provide different information, as is trivially



found in number agreement, but that does not imply reference to the structure of the word. Nor does the syntax need to reference purely morphological features, such as inflection class. These determine the values of gender, a morphosyntactic feature, to which the syntax always has access. This line argument holds for the other examples discussed, hence our data are in accord with the Principle of Morphology-free Syntax.

## 8 Conclusion

Internal and external splits had been documented previously; the new challenge was to build a typology of how they are linked. We set up the Extended Lexeme Consistency Principle as a baseline from which to calibrate a range of largely unreported phenomena. There are clear examples where internal and external splits co-vary, and the evidence in some instances was sufficient to show that other interpretations were highly implausible. More specific versions of the plausibility argument involved lexemes showing overabundance, lexemes showing variation over time and space, and the different types of pluralia tantum nouns: in all of these types of evidence the co-variation of internal and external splits argued strongly for the linkage. Moreover, wherever there was evidence bearing on directionality, it pointed to the internal linkage determining the external one. Since morphosyntactic features (gender and number) were involved, this result is consonant with the Principle of Morphology-free syntax. A further specific finding was that the range of the internal split (sub-paradigm or cell) determines the nature of the external split (consistent or hybrid). This result demonstrates that featural information is associated with the lexeme in a default hierarchy, namely at the level of the lexeme by default, unless overridden at the sub-paradigm level, unless in turn overridden at the level of individual cells (but compare the discussion in Stump, 2016:92–95). The examples involving hybrids led to an extension of the scope of the Agreement Hierarchy. Previous controllers subject to the Agreement Hierarchy had involved meaning-meaning mismatches or form-meaning mismatches. Our new examples provide the missing type, namely form-form mismatches. This means that they can be accommodated within the constraint of the Agreement Hierarchy without the need to propose a new mechanism for them. Often the key evidence is found with few lexemes, or even a single one. In the era of big data, this investigation shows that small data too can have great value.

## Appendix 1: Further evidence from overabundance

These examples complement the data in §4.

**Polish *tajdak* ‘scoundrel’ and similar nouns:** here just one cell is affected (in this instance the nominative plural). Polish nouns denoting male humans have distinctive nominative plural inflection (the ‘masculine personal’) and potentially a mutation of the stem-final consonant. The expected nominative plural for *tajdak* ‘scoundrel’ is *tajdacy*. In addition, there is also a ‘depreciative’ form, *tajdaki*, found with this



and other nouns with typically pejorative overtones (the previous term was ‘devirilized’, see Swan, 2015:93–94, following Saloni). For data and discussion, see Brown (1998:206–209), Brown and Hippisley (2012:103–105), Corbett (2006:216–217), Swan (2015:93–96), and references there. We turn now to the external split. The masculine personal form *tajdacy* ‘scoundrels’ takes masculine personal agreement forms, like normal nouns denoting male persons. The depreciative form *tajdaki* is a hybrid, which takes both non-masculine personal agreement forms and masculine personal agreement forms. There is great variation between speakers here: a summary picture is that we find non-masculine personal agreement in attributive position, both types of agreement in the predicate, and masculine personal agreement for the relative and personal pronoun (Corbett, 2006:216–217). Thus *tajdak* ‘scoundrel’ is similar to *ręka* ‘hand’ in involving a single cell of the paradigm and in being overabundant in that cell. (Note, however, that to a more restricted degree, a second cell, the accusative plural, may be involved for ‘especially pejorative’ use, for which see Swan (2015:93) and Brown and Hippisley (2012:102–103).) On the other hand, *tajdak* ‘scoundrel’ differs from *ręka* ‘hand’, in that it is not unique (there are several similar nouns), and in denoting humans, so that the agreement options are semantically charged.

**Italian and Italo-Romance:** we find careful discussion of the relation of overabundance to inqorate genders in Thornton (2010–2011:459–460); she gives detailed information on Italian *ciglio* ‘eyelash, side/edge (of road)’. The most common options are *ciglio* (M) ‘eyelash’, plural *ciglia* (F), and *ciglio* (M) ‘roadside’, plural *cigli* (M). Like *braccio* (M) ‘arm’, with plurals *braccia* (F) and *bracci* (M), the agreements link to the particular plural form, and neither noun is a hybrid (Anna Thornton, personal communication, 23 Nov. 2020). For discussion see Loporcaro and Paciaroni (2011:401–404), Loporcaro (2018:81–87), and for the interesting history of this gender see Loporcaro et al. (2014). For Italo-Romance more generally, see Paciaroni et al. (2013); Loporcaro produces spectacular evidence linking internal and external splits, with hybrid nouns resulting, in Loporcaro (2018: 142–145 for Agnonese, and 2018:271–272 for Sangiovese).

## Appendix 2: Further evidence from diachrony: Slovenian *pot* ‘way, path’

This example complements the data in §5.

In contemporary Slovenian (slv), *pot* ‘way, path’ (and a range of associated meanings) shows interesting remnants from its chequered past. It was one of the minority of *i*-stems (inflection class III) that were masculine in Common Slavonic (more were feminine). The reflexes in the different Slavonic languages present a varied and interesting picture. The Slovenian data are mainly from Tom Priestly and Don Reindl (numerous personal communications, September–October 2020). Dictionaries and grammars suggest that the noun is a heteroclite, and that there are different forms (internal split) which induce in turn masculine, feminine and neuter gender agreement (external split). However, the contemporary picture is rather simpler. Don Reindl contacted

over 300 former students, and for them the noun is now simply feminine, with just two saying they had knowledge of masculine usage. But while the noun follows inflection class III (as in Priestly, 1993:406) and is indeed typically feminine, there are interesting survivals in the language, preserved especially in names. And what is key for us is that these remnants preserve the unusual inflection and the unusual agreement together. For example, there is a hikers' magazine *Naša pota* 'Our Paths' (probably chosen to have a ring to it), where *pot-a* is nominative plural of inflection class IV, and *naš-a* is neuter plural (as would be expected for a noun in this inflection class). But the neuter is highly restricted, in that examples which could be unambiguously recognized as having neuter agreement are found for this noun only in the nominative and accusative plural. When we look at usage within the magazine for 2018, we find examples such as:<sup>33</sup>

Slovenian

- (i) Pot je bi-l-a strm-a in dolg-a  
 path[SG.NOM] AUX.3SG be-PST.PTCP-SG.F steep-SG.F and long-SG.F  
 'the path was steep and long'

In (i) we find the expected contemporary feminine agreement. (Note that the nominative singular form *pot* could belong to inflection class I or III, which plays a major role in its change of inflection class and gender.) As already noted, there are remnant masculines, preserved in names. There is *križev pot*, literally 'cross's way' hence 'Way of the Cross', with the derived adjective *križev* in the masculine. If we take another case value, the appropriate morphology (class I) and gender (masculine) are retained:

- (ii) postaj-e križ-ev-ega pot-a  
 station-PL.NOM cross-ADJ-SG.GEN.M way-SG.GEN  
 'the Stations of the Way of the Cross' (the Stations of the Cross)

The noun *pot* is found in several Ljubljana street names: 65 are feminine, and three masculine (Don Reindl, personal communication 30 Jan. 2021). An example of the masculine is *Lepi pot* 'Fine Street'. Again, the appropriate inflection class and gender is used in oblique cases:

- (iii) na Lep-em pot-u  
 on fine-SG.LOC.M street-SG.LOC  
 'on Fine Street'

The point then is that although *pot* is largely regularized as an inflection class III noun, and hence feminine, other forms (as in various dictionaries and grammars) live on in different types of name, preserving the older form and the older gender appropriate to the form. Before leaving Slovenian we should note that in the Sele dialect, spoken in southern Austria and famous through Priestly's descriptions (1983, 1984, 2006), *pot* has adopted almost entirely the forms of inflection class I (remaining irregular forms are restricted to certain plural cases where agreement targets do not

<sup>33</sup><http://www.pdhrastnik.si/nasa-pota-2018/nasa-pota-2018/>.

distinguish gender). Consequently it is fully masculine in gender in the Sele dialect. The Slovenian data are a clear illustration of the remnants of overabundance, retaining the linkage from internal to external split.

**Acknowledgements** Versions of this paper were read at the workshop ‘How to fill a cell: computational approaches to inflectional morphology’, University of York, 16 September 2020, at the Indiana University Linguistics Department Colloquium, 11 December 2020, and at the workshop ‘Internationalisms in Slavic as a window into the architecture of grammar’, Department of Slavic Studies, University of Graz, Austria, 24–26 February 2021. I thank participants at all three (online) events for their helpful comments. For general discussion and helpful suggestions on a draft version, I am grateful to Matthew Baerman, Patricia Cabredo Hofherr, Hans-Olav Enger, Michael Franjeh, Steven Kaye, Erich Round, Anna Thornton and Nigel Vincent. For expert help primarily on specific languages I am indebted to: Yvonne Agbetsoamedo (Sele), Wayles Browne and Ljubomir Popović (Serbo-Croat), Wolfgang de Melo (Latin), Jürg Fleischer (Old Frisian), Petros Karatsareas (Asia Minor Greek), William Lamb (Scottish Gaelic), Michele Loporcaro (Italo-Romance languages), and Tom Priestly and Donald Reindl (Slovenian). Greg Stump kindly gave me access to his heteroclis database. Penny Everson and Lisa Mack provided valuable help in the preparation of the paper. Special thanks to Olivier Bonami as Editor, and two *Morphology* referees, for their time and expertise. Part of this research was carried out while I was a guest at the Max Planck Institute for Evolutionary Anthropology (Leipzig), Department of Human Behavior, Ecology and Culture; I am thankful for this hospitality. The support of the Arts and Humanities Research Council UK (grant: AH/N006887/1: Lexical splits: a novel perspective on the structure of words) is gratefully acknowledged.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agbetsoamedo, Y. (2014). Noun classes in Sele. *Journal of West African Languages*, 41, 95–124.
- Arkadiev, P., & Kozhanov, K. (2021, forthcoming). Borrowing of morphology. P. Ackema, S. Bendjaballah, E. Bonet, & A. Fábregas (Eds.) *The Wiley Blackwell companion to morphology*. Hoboken: John Wiley.
- Baayen, R. H., Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics*, 35, 861–877.
- Babou, C. A., & Loporcaro, M. (2016). Noun classes and grammatical gender in Wolof. *Journal of African Languages and Linguistics*, 37, 1–57. <https://doi.org/10.1515/jall-2016-0001>.
- Bach, X. (2018). *The origins of inflectional classes*. PhD thesis, University of Oxford.
- Baerman, M. (2007). Morphological typology of deponency. In M. Baerman, G. G. Corbett, D. Brown, & A. Hippisley (Eds.), *Deponency and morphological mismatches, Proceedings of the British Academy* (Vol. 145, pp. 1–19). Oxford: British Academy and Oxford University Press.
- Baerman, M., Brown, D., & Corbett, G. G. (2017). *Morphological complexity*. Cambridge: Cambridge University Press.
- Bailly, J. F. (2010). To what degree are Croatian and Serbian the same language? Evidence from a translation study. *Journal of Slavic Linguistics*, 18, 181–219.
- Benson M. (1971). *SerboCroatian-English dictionary*. Philadelphia: University of Pennsylvania Press.
- Bremmer, R. H. Jr. (2009). *An introduction to Old Frisian: History, grammar, reader, glossary*. Amsterdam: John Benjamins.
- Brown, D. (1998). Defining ‘subgender’: Virile and devirilized nouns in Polish. *Lingua*, 104, 187–233.
- Brown, D., & Hippisley, A. (2012). *Network morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press.

- Browne, W. (1993). Serbo-Croat. In B. Comrie, & G. G. Corbett (Eds.), *The Slavonic languages* (pp. 306–387). London: Routledge.
- Browne, W. (2021). Adjacent vs. separated placement of preposition and noun as a factor in noun inflection: the cases of Bosnian-Croatian-Serbian *pazuho* ‘armpit’. In P. Arkadiev, J. Pakerys, I. Šeškauskienė, & V. Žeimantienė (Eds.), *Studies in Baltic and other Languages. A Festschrift for Axel Holvoet on the occasion of his 65th birthday*, Vilnius: Vilnius University Press. <https://www.journals.vu.lt/open-series/article/view/24484/23748>.
- Browne, W., & Alt, T. (2004). *A handbook of Bosnian, Serbian, and Croatian*. SEELRC: Duke University. [www.seelrc.org:8080/grammar/mainframe.jsp?nLanguageID=1](http://www.seelrc.org:8080/grammar/mainframe.jsp?nLanguageID=1).
- Bugariski, R. (2012). Language, identity and borders in the former Serbo-Croatian area. *Journal of Multilingual and Multicultural Development*, 33, 219–235.
- Bugariski, R. (2019). Past and current developments involving pluricentric Serbo-Croatian and its official heirs. In T. Schoonheim & J. Van Hoorde (Eds.), *Language variation: A factor of increasing complexity and a challenge for language policy within Europe: Contributions to the EFNIL conference 2018 in Amsterdam* (pp. 105–114). Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences. <http://www.efnil.org/documents/conference-publications/amsterdam-2018>.
- Carstairs-McCarthy, A. (1994). Inflection class, gender and the principle of contrast. *Language*, 70, 737–788.
- Chumakina, M. E., Hippiusley, A., & Corbett, G. G. (2004). Istoričeskie izmenenija v ruskoj leksike: Slučaj čeredujuščegosja suppletivizma [Historical changes in the Russian lexicon: a case of alternating suppletion]. *Russian Linguistics*, 28, 281–315.
- Comrie, B. (2001). How independent is Tsezic morphology? In M. Andronis, C. Ball, H. Elston, & S. Neuvel (Eds.), *CLS 37: The panels: 2001: Proceedings from the parasessions of the 37th meeting of the Chicago Linguistic Society* (Vol. 37–2, pp. 367–383). Chicago: Chicago: Linguistic Society.
- Corbett, G. G. (1979). The agreement hierarchy. *Journal of Linguistics*, 15, 203–224. <https://doi.org/10.1017/S002226700016352>.
- Corbett, G. G. (1983). *Hierarchies, targets and controllers: Agreement patterns in Slavic*. London: Croom Helm.
- Corbett, G. G. (2000). *Features*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2007a). Canonical typology, suppletion and possible words. *Language*, 83, 8–42. <https://doi.org/10.1353/lan.2007.0006>.
- Corbett, G. G. (2007b). Deponency, syncretism and what lies between. In M. Baerman, G. G. Corbett, D. Brown, & A. Hippiusley (Eds.), *Deponency and morphological mismatches, Proceedings of the British Academy* (Vol. 145, pp. 21–43). Oxford: British Academy and Oxford University Press. <https://doi.org/10.5871/bacad/9780197264102.003.0002>.
- Corbett, G. G. (2009a). Canonical inflectional classes. In F. Montermini, G. Boyé, & J. Tseng (Eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux* (pp. 1–11). Somerville: Cascadilla Proceedings Project. <http://www.lingref.com/cpp/decemb/6/abstract2231.html>.
- Corbett, G. G. (2009b). Morphology-free syntax: two potential counter-examples from Serbo-Croat. In S. Franks, V. Chidambaram, & B. Joseph (Eds.), *A linguist's linguist: studies in South Slavic linguistics in honor of E. Wayles Browne*, (pp. 149–166). Bloomington, Indiana: Slavica.
- Corbett, G. G. (2012). *Features*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2015a). Morphosyntactic complexity: A typology of lexical splits. *Language*, 91, 145–193. <https://muse.jhu.edu/article/576997>.
- Corbett, G. G. (2015b). Hybrid nouns and their complexity. In J. Fleischer, E. Rieken, & P. Widmer (Eds.), *Agreement from a diachronic perspective* (pp. 191–214). Berlin: De Gruyter Mouton.
- Corbett, G. G. (2019). Pluralia tantum nouns and the theory of features: A typology of nouns with non-canonical number properties. *Morphology*, 29, 51–108. <https://doi.org/10.1007/s11525-018-9336-0>.
- Corbett, G. G. (under review). The typology of external splits.
- Corbett, G. G., Hippiusley, A., Brown, D., & Marriott, P. (2001). Frequency, regularity and the paradigm: A perspective from Russian on a complex relation. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 201–226). Amsterdam: John Benjamins.
- Corbett, G. G., & Browne, W. (2018). Serbo-Croat: Bosnian, Croatian, Montenegrin, Serbian. In B. Comrie (Ed.), *The world's major languages* (3rd ed., pp. 339–356). London: Routledge.
- de Melo, W. D. C. (2019). Varro: *De lingua Latina: Introduction, text, translation, and commentary: I & II*. Oxford: Oxford University Press.
- Despić, M. (2017). Investigations on mixed agreement: polite plurals, hybrid nouns and coordinate structures. *Morphology*, 27, 253–310. <https://doi.org/10.1007/s11525-017-9301-3>.

- Di Garbo, F., & Agbetsoamedo, Y. (2018). Non-canonical gender in African languages: A typological survey of interactions between gender and number, and between gender and evaluative morphology. In S. Fedden, J. Audring, & G. G. Corbett (Eds.), *Non-canonical gender systems* (pp. 176–210). Oxford: Oxford University Press.
- Drettas, G. (1997). *Aspects pontiques*. Paris: Association de recherches pluridisciplinaires.
- Dvořák, B., & Sauerland, U. (2006). The semantics of the Slovenian dual. In J. Lavine, S. Franks, M. Tasseva-Kurkchieva, & H. Filip (Eds.), *Proceedings of FASL 14: The Princeton meeting* (pp. 98–112). Ann Arbor: Michigan Slavic Publications.
- Dwelly, E. (1902–1911). *Faclair Gaidhlig: A Gaelic Dictionary: specially designed for beginners and for use in schools*. Herne Bay: E. MacDonald and Co. at the Gaelic Press. <http://www.dwelly.info>. Accessed 10 October 2020.
- Enger, H.-O., & Corbett, G. G. (2012). Definiteness, gender and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics*, 24, 287–324. <https://doi.org/10.1017/S1470542712000098>.
- Feist, T., & Palancar, E. L. (2021). Paradigmatic restructuring and the diachrony of stem alternations in Chichimec. *Language*, 97, 1–41. <https://doi.org/10.1353/lan.2021.0000>.
- Fleischer, J. (2012). Grammatische und semantische Kongruenz in der Geschichte des Deutschen: eine diachrone Studie zu den Kongruenzformen von ahd. *wīb*, nhd. *Weib*. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 134, 163–203.
- Fleischer, J., & Widmer, P. (2016). When lexical hybrids become feminine: The declension and agreement behavior of OFr *wīf* ‘woman; wife’ in Old Frisian and modern Frisian varieties. *Philologia Frisica*, 14, 219–239.
- Franks, S. (2020). *Microvariation in the South Slavic noun phrase*. Bloomington IN: Slavica.
- Gardani, F. (2020). Borrowing matter and pattern in morphology: An overview. *Morphology*, 30, 263–282. <https://doi.org/10.1007/s11525-020-09371-5>.
- Harrison, A. J. (2009). *The production of subject-verb agreement in Slovene and English*. PhD thesis, University of Edinburgh.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, 39, 1041–1070. <https://doi.org/10.1515/ling.2001.041>.
- Herrity, P. (2000). *Slovene: A comprehensive grammar*. London: Routledge.
- Hippisley, A., Chumakina, M., Corbett, G. G., & Brown, D. (2014). Suppletion: Frequency, categories and distribution of stems. *Studies in Languages*, 28, 387–418. <https://doi.org/10.1075/sl.28.2.05hip>.
- Hristov, B. (2013). Defacing agreement. In M. Butt & T. Holloway King (Eds.), *Proceedings of the LFG13 conference* (pp. 335–355). Stanford: CSLI Publications. <http://web.stanford.edu/group/cslipublications/cslipublications/LFG18/lfg13.html>.
- Hyman, L. M. (1981). *Noni grammatical structure: with special reference to verb morphology (southern California Occasional Papers in Linguistics, 9)*. Los Angeles: Department of Linguistics, University of Southern California.
- Ignjatović, D. (1963). Novine u morfologiji unesene najnovijim imenicama stranog porekla [Novelties in morphology introduced by the most recent nouns of foreign origin]. *Naš jezik (nova serija)*, 13, 208–224.
- Iosad, P., & Lamb, W. (2020). Dialect variation in Scottish Gaelic nominal morphology: A quantitative study. *Glossa: A Journal of General Linguistics*, 5(1), 1–31. <https://doi.org/10.5334/gjgl.1023>.
- Karatsareas, P. (2011a). *A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective*. PhD dissertation, University of Cambridge.
- Karatsareas, P. (2011b). Neuter heteroclisis in Asia Minor Greek: Origin and development. *Neoelliniki Dialektologia [Modern Greek Dialectology]*, 6, 111–135.
- Karlsson, G. (1968). Eras suomen inkongruenssi-ilmio [An incongruence phenomenon in Finnish]. *Fennougrica: Juhlakirja Lauri Postin kuusikymmenvuotispäiväksi 17. 3. 1968. Mémoires de la Société Finno-Ougrienne*, 145, 117–126. Helsinki: Suomalais-ugrilainen seura.
- Kaye, S. J. (2015). *Conjugation class from Latin to Romance: heteroclisis in diachrony and synchrony*. PhD thesis, University of Oxford.
- Klajn, I. (1998). Vrste romanizama u savremenom srpskohrvatskom jeziku i putevi njihovog dolaska [Types of Romance borrowings in modern Serbo-Croat and the ways in which they enter the language]. *Zbornik Matice srpske za filologiju i lingvistiku*, 41, 69–89.
- Klajn, I. (2001). Neologisms in present-day Serbian. *International Journal of the Sociology of Language*, 151, 89–110.
- Kornfilt, J. (1997). *Turkish*. London: Routledge.
- Lamb, W. (2008). *Scottish Gaelic speech and writing: Register variation in an endangered language*. (Belfast Studies in Language, Culture and Politics 16). Belfast: Cló Ollscoil na Banríona.

- Landau, I. (2016). DP-internal semantic agreement: A configurational analysis. *Natural Language and Linguistic Theory*, 34, 975–1020. <https://doi.org/10.1007/s11049-015-9319-3>.
- Lazarsfeld, P. F. (1937). Some remarks on the typological procedures in social research. *Zeitschrift für Sozialforschung*, 6, 119–139.
- Lewis, C. T., & Short, C. (1879). *A Latin dictionary: founded on Andrews' edition of Freund's Latin dictionary: Revised, enlarged, and in great part rewritten by Charlton T. Lewis, Ph.D. and Charles Short, L.L.D.* Oxford: Clarendon Press.
- Loporcaro, M. (2018). *Gender from Latin to Romance: History, geography, typology*. Oxford: Oxford University Press.
- Loporcaro, M., Faraoni, V., & Gardani, F. (2014). The third gender of Old Italian. *Diachronica*, 31, 1–22. <https://doi.org/10.1075/dia.31.1.01gar>.
- Loporcaro, M., & Paciaroni, T. (2011). Four-gender systems in Indo-European. *Folia Linguistica*, 45, 389–433. <https://doi.org/10.1515/flin.2011.015>.
- Mackinnon, D. (1910). Traces of the neuter gender in modern Scottish Gaelic. *Celtic Review*, 6(24), 296–304. <https://doi.org/10.2307/30070227>.
- Maiden, M. (2004). When lexemes become allomorphs — On the genesis of suppletion. *Folia Linguistica*, 38, 227–256. <https://doi.org/10.1515/flin.2004.38.3-4.227>.
- Maiden, M. (2009). Where does heteroclisism come from? Evidence from Romanian dialects. *Morphology*, 19, 59–86. <https://doi.org/10.1007/S11525-009-9130-0>.
- Maretić, T. (1931). *Gramatika i stilistika hrvatskoga ili srpskoga književnog jezika [Grammar and stylistics of the Croatian or Serbian literary language]* (2nd ed.). Zagreb: Obnova.
- Mark, C. (2003). *The Gaelic-English dictionary*. London: Routledge.
- Marković, S. V. (1954). O kolebljivosti slaganja u rodu kod imenica čiji se prirodni i gramatički rod ne slažu (i o rodu ovih imenica) [Variability of gender agreement in nouns whose natural and grammatical gender do not match (and on the gender of these nouns)]. *Pitanja književnosti i jezika (Sarajevo)*, 1, 87–110.
- Matešić, J. (1965–1967). *Rückläufige Wörterbuch des Serbokroatischen*. Wiesbaden: Otto Harrassowitz.
- Matthews, P. H. (2007). *The concise Oxford Dictionary of Linguistics* (2nd edn). Oxford: Oxford University Press.
- McGill, S. J. (2009). *Gender and person agreement in Cicipo discourse*. PhD thesis, SOAS, University of London.
- Muravenko, E. V. (2014). Izmenenie upravljenija predloga *po* v ruskom jazyke [The change in government of the preposition *po* in Russian]. *Acta Linguistica Petropolitana: Trudy instituta lingvističeskix issledovanij*, 2, 643–672. <https://cyberleninka.ru/article/v/izmenenie-upravljeniya-predloga-po-v-ruskom-yazyke>.
- Noyer, R. (2005). A constraint on interclass syncretism. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2004* (pp. 273–315). Dordrecht: Springer.
- O'Neill, P. (2016). Lexicalism, the principle of morphology-free syntax and the principle of syntax-free morphology. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 237–271). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139814720.010>.
- Oftedal, M. (1956). *A linguistic survey of the Gaelic dialects of Scotland: Vol. III: The Gaelic of Leurbost, Isle of Lewis*. Oslo: Aschehoug & Co.
- Olander, T. (2015). *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill.
- Paciaroni, T., Nolè, G., & Loporcaro, M. (2013). Persistenza del neutro nell'italo-romanzo centro-meridionale [Persistence of the neuter in central-southern Italo-Romance dialects]. *Vox Romanica*, 72, 88–137.
- Piškočić, T. (2011). *Gramatika roda [The grammar of gender]*. Zagreb: Disput.
- Priestly, T. M. S. (1983). On 'drift' in Indo-European gender systems. *Journal of Indo-European Studies*, 11, 339–363.
- Priestly, T. M. S. (1984). O popolni izgubi srednjega spola v selščini: enodobni opis [On the complete loss of the neuter gender in Sele: a synchronic description]. *Slavistična revija*, 32, 37–47.
- Priestly, T. M. S. (1993). Slovene. In B. Comrie, & G. G. Corbett (Eds.), *The Slavonic languages* (pp. 388–451). London: Routledge.
- Priestly, T. (2006). Gender change, local markedness, and speculations on the dual. In G. Wiebe, G. Libben, T. Priestly, R. Smyth, & S. Wang (Eds.), *Phonology, morphology and the empirical imperative: Papers in honour of Bruce L. Derwing* (pp. 409–423). Taipei City, Taipei: The Crane Publishing Co.
- Risch, E. (1977). Das System der lateinischen Deklinationen. *Cahiers Ferdinand de Saussure*, 31, 229–245.



- Rothstein, R. A. (1993). Polish. In B. Comrie, & G. G. Corbett (Eds.), *The Slavonic languages* (pp. 686–758). London: Routledge.
- Round, E. R., & Corbett, G. G. (2020). Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology*, 24, 489–525. <https://doi.org/10.1515/lingty-2020-2060>.
- Sag, I. A., Chaves, R. P., Abeillé, A., Estigarribia, B., Flickinger, D., Kay, P., Michaelis, L. A., Müller, S., Pullum, G. K., Van Eynde, F., & Wasow, T. (2020). Lessons from the English auxiliary system. *Journal of Linguistics*, 56, 87–155. <https://doi.org/10.1017/S002222671800052X>.
- Sagna, S. (2019). Syntactic and semantic agreement in Eegimaa (Banjal): An account of lexical hybrids in an African noun class system. *Studies in Languages*, 43, 585–627. <https://doi.org/10.1075/sl.14023.sag>.
- SANU (Srpska akademija nauka i umetnosti) (1959). *Rečnik srpskohrvatskog književnog i narodnog jezika* [Dictionary of standard and vernacular Serbo-Croat]. Belgrade: SANU.
- Sapsford, R. (2006). Extracting and presenting statistics. In R. Sapsford & V. Jupp (Eds.), *Data collection and analysis* (2nd ed.). London: Sage.
- Spencer, A. (2013). *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.
- Steller, W. (1928). *Abriß der altfriesischen Grammatik: mit Berücksichtigung der westgermanischen Dialecte des altenglischen, altsächsischen und althochdeutschen: mit Lesestücken und Wortverzeichnis*. Halle: Niemeyer.
- Stevanović, M. (1970). *Savremeni srpskohrvatski jezik (Gramatički sistemi i književnojezička norma): I: Uvod, fonetika, morfologija* [Contemporary Serbo-Croat (Grammatical systems and literary norm): I: Introduction, phonetics, morphology]. Belgrade: Naučna knjiga.
- Stump, G. T. (2006). Heteroclis and paradigm linkage. *Language*, 82, 279–322. <https://doi.org/10.1353/lan.2006.0110>.
- Stump, G. (2016). *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.
- Swan, O. (2015). Polish gender, subgender, and quasi-gender. *Journal of Slavic Linguistics*, 23, 83–122. <https://www.jstor.org/stable/24602179>.
- Thornton, A. M. (2010–2011). La non canonicità del tipo it, braccio // braccia / bracci: Sovrabbondanza, difettività o iperdifferenziazione? [Non-canonicity of Italian braccio // braccia / bracci: Overabundance, defectiveness or overdifferentiation?]. *Studi di grammatica italiana*, 29–30, 419–477 [volume year 2010–2011, published 2013].
- Thornton, A. M. (2019a). Overabundance in morphology. In *Oxford research encyclopedia of linguistics*, Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.554>.
- Thornton, A. M. (2019b). Overabundance: A canonical typology. In F. Rainer, F. Gardani, W. U. Dressler, & H. C. Luschützky (Eds.), *Competition in inflection and word-formation (studies in morphology)* (Vol. 5, pp. 223–258). Cham: Springer. [https://doi.org/10.1007/978-3-030-02550-2\\_9](https://doi.org/10.1007/978-3-030-02550-2_9).
- Tošović, B. (2016). *Korelaciona morfologija srpskog, hrvatskog i bošnjačkog jezika: Imenice: Konstrukcijsko izdanje* [Correlational analysis of the morphology of Serbian, Croatian and Bosnian: Nouns: Construction edition]. Graz: Institut für Slawistik der Karl-Franzens-Universität Graz.
- Vaillant, A. (1964). *Manuel du vieux slave: I: Grammaire* (2nd ed., revised: first edition 1948). Paris: Institut d'Études slaves.
- Vujaklija, M. (1980). In S. Ristić & R. Aleksić (Eds.), *Leksikon stranih reči i izraza* [Lexicon of foreign words and expressions] (3rd ed.). Belgrade: Prosveta.
- Wechsler, S., & Zlatić, L. (2003). *The many faces of agreement*. Stanford: CSLI.
- Wechsler, S., & Zlatić, L. (2012). The wrong two faces. *Language*, 88, 380–387. <https://www.jstor.org/stable/23251836>.
- Zaliznjak, A. A. (1967/2002). *Russkoe imennoe slovoizmenenie* [Russian nominal inflection]. Moscow: Nauka. [Reprinted In Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*, 1–370. Moscow: Jazyki slavjanskoj kul'tury.] [Page references to 2002 edition].
- Zwicky, A. M. (1992). Some choices in the theory of morphology. In R. Levine (Ed.), *Formal grammar: Theory and implementation* (pp. 327–371). New York: Oxford University Press.