

Pluralia tantum nouns and the theory of features: a typology of nouns with non-canonical number properties

Greville G. Corbett¹

Received: 27 April 2018 / Accepted: 9 November 2018
© The Author(s) 2018

Abstract Many languages have an interesting class of nouns, the pluralia tantum, which have restricted number possibilities when, in some sense, they should not. Thus English *binoculars* has no singular, which is worth noting (that is, it is not predictable). True, there are other nouns denoting items consisting of two significant parts which behave similarly (*spectacles, trousers...*); indeed they are subject to ‘middle-size generalizations’ (Koenig 1999). But there are two reasons to note such nouns. First there are many English nouns equally denoting items consisting of two significant parts which are unremarkable in this respect: *bicycle, bigraph, Bactrian camel, couple, duo...* And second, there are languages with number systems roughly comparable to that of English in which the equivalents of *binoculars* and *trousers* are normal count nouns: Russian *binokl’*, French *pantalon*. While pluralia tantum are of continuing interest, it is typically only the English type which is considered. But these familiar examples offer an entry point to a collection of lexical items, some with much stranger behaviour, lurking between the semi-predictable and the unexpectedly defective. In particular, some instances demonstrate that we cannot maintain the general assumption that the ‘internal’ morphosemantic specification of a paradigm cell and its ‘external’ morphosyntactic requirement are necessarily identical. I therefore set out a full typology of these fascinating nouns, so that their significance can be more fully appreciated. I start from the notion of canonical noun, and calibrate the different non-canonical properties according to a set of orthogonal criteria.

Keywords Grammatical number · Canonical typology · Pluralia tantum · Morphosemantic features · Morphosyntactic features · Agreement Hierarchy

✉ G.G. Corbett
g.corbett@surrey.ac.uk

¹ Surrey Morphology Group, Literature and Languages, Faculty of Arts and Social Sciences, University of Surrey, Guildford, Surrey, GU7 2XH, UK.

1 Introduction

Pluralia tantum are a type of noun that ought not to exist. Like many other linguistic terms, the term ‘pluralia tantum’ carries the implication that something is not quite right. Thus forms which are syncretic contradict the expectation that every grammatical function has a unique form, periphrastic forms involve multi-word expressions where elsewhere the system involves synthetic exponence, and so on. Pluralia tantum is similarly a label for nouns which have only a plural when, in some sense, this is not expected. English *binoculars* has no singular; this is worth noting, since it is not predictable, given that *binoculars* can denote one item or more than one. True, there are other nouns denoting items consisting of two significant parts which behave similarly (*spectacles, trousers, . . .*). But there are two reasons to focus on such nouns. First there are many English nouns equally denoting items consisting of two significant parts which are normal in respect of number: *bicycle, bigraph, Bactrian camel, couple, duo*... And second, there are languages with number systems roughly comparable to that of English in which the equivalent of *binoculars* is a normal count noun: Russian *binokl’* (SG) and *binokli* (PL). Conversely, Russian *sani* ‘sledge’ is a pluralia tantum noun, unlike its English equivalent. (Which prompts the question of how we talk about one sledge in Russian.) As we shall see, these examples lead to a wide variety of nouns, which lack the full range of number behaviour.

For investigating these different items, considering possible definitions will prove thought-provoking and helpful. It may seem obvious what pluralia tantum are, but while there are linguists with clear intuitions about what should count as pluralia tantum, these intuitions are not necessarily shared. The differences of view can be understood when we examine carefully the nature of the features involved. We shall move between definitions and data, homing in on a fuller typology, a more satisfying definition, and a better account of morphosyntactic features. Let us start with two definitions from dictionaries of linguistic terms:

“**pluralia tantum**. Latin ‘plurals only’: i.e. nouns, like *oats* or *trousers*, which appear only in a plural form.” (Matthews 1997:284)

This is a simple definition, and better than most. Clearly it covers *binoculars* and the other examples we have noted so far. While this definition is a good starting point, much rests on how we interpret ‘plural form’.

In the same year, another dictionary of linguistic terms offered this definition:

“**plurale tantum** A noun which is plural in form but singular in meaning, such as *scissors, pants* or *binoculars*. The plural of this term is *pluralia tantum*.” (Trask 1997:172)

This definition already illustrates some of the pitfalls. *Scissors* is not necessarily singular in meaning (see Sect. 2.1).¹

¹Wikipedia offers a traditional definition:

A **plurale tantum** (Latin for “plural only”, plural form: **pluralia tantum**) is a noun that appears only in the plural form and does not have a singular variant for referring to a single object. (Wikipedia, accessed 13.11.2016)

It is worth asking why pluralia tantum and comparable nouns deserve a new study. Part of the continuing interest is that pluralia tantum are clearly exceptional, yet in some instances partly predictable too (Sect. 5). They are missing something, but they are typically unlike other defectives (Sect. 5, Sect. 8.6). They are problematic for Full Interpretation (at least for strong versions), see Sect. 4.6. While morphologists and syntacticians have done most research on pluralia tantum, formal semanticists are becoming curious, as are others in cognitive science (Wisniewski 2009) and psycholinguistics (Bock et al. 2001; Nenonen and Niemi 2010; and Nickels et al. 2015). Most work in these wider areas, however, is limited to examples like *scissors*. This is a missed opportunity, given the variety of pluralia tantum. But this is part of a more general issue, the limited focus of much experimental research (see the recent survey by Acuña-Fariña 2016). Research on pluralia tantum has recently been extended to sign languages (Börstell et al. 2016).

The current paper analyses data from a range of languages, and includes pluralia tantum nouns within a broader landscape of nouns which have some sort of limitation in terms of the number feature. This entails both showing the rich variety of types of number-deficient nouns, of which *scissors* is the tip of a fascinating iceberg, and giving the wider context of limitations within number systems. Some include pluralia tantum under the umbrella ‘lexical plurals’ (Sect. 4.3); pluralia tantum require a specification in their lexical entry, and are ‘lexical’ in that sense. But they are to be strictly delineated from what is generally covered under the lexical plural heading by the characteristic of deficiency, as will be clear as our definitions are refined. I first introduce enough data to show that there is indeed a challenging set of questions here (Sect. 2). Then I show how these instances fit into a bigger picture, by taking a canonical approach. Useful previous work using this approach is laid out in Sect. 3, which leads to an analysis of what we might expect from a noun which is absolutely straightforward in terms of number (Sect. 4). This serves as a baseline against which we calibrate the real examples we find. The related issues of animacy and motivation are examined in Sect. 5; this takes us to the specific motivation which has been claimed for bipartites (like *scissors*), and from there it is a natural step to look at larger systems, those with duals and other number values (Sect. 6). In Sect. 7 we examine the ways in which pluralia tantum nouns are accommodated (how their lack of number form(s) is managed). Finally we bring together the typological range of pluralia tantum (Sect. 8), consider pluralia tantum constructions (Sect. 9) and draw the main conclusions (Sect. 10).

This definition covers instances like English *binoculars and scissors*, where there is only a plural form, which is used for one object or more than one. It starts and finishes with the morphology; the syntax (agreement) is not mentioned. Glottopedia (accessed 2.2.2017), basing itself on the Utrecht Lexicon of Linguistics, has:

A **plurale tantum** is a noun that only occurs in the plural. The term is a traditional term used for words which (a) end in a plural affix, (b) have a plural meaning, and (c) do not have a singular counterpart.

It gives *trousers, scissors, scales* as examples, though it is not clear that these have a plural meaning. Hence the Wikipedia version is a better starting point.

2 Pluralia tantum: an initial typology

To identify key types of pluralia tantum nouns, let us look at the basic relations between their semantics, syntax and morphology. What we might expect (and this will be the basis of our typology in Sect. 4) is that semantics, syntax and morphology would line up. Thus *books* denotes more than one entity, it is syntactically plural in that it takes plural agreement, and it is morphologically plural having the plural marker *-s*. With pluralia tantum nouns the three components typically do not line up neatly, which is the source of their theoretical importance. And this lack of alignment means that we must be careful in our use of terms, in particular what exactly we mean when we say a noun is ‘plural’. This need for care holds equally, whatever our theoretical position.

An approach using the traditional morpheme faces great challenges since, typically, pluralia tantum nouns involve different realizations of the plural (see Sect. 4.6), hence the special factor is not to be located in a plural morpheme but is a featural problem. We shall need to be clear about the type of feature involved, since it may be morphosemantic or morphosyntactic (Corbett 2012:49–50), a distinction examined carefully in Sect. 2.4. I shall therefore take an inferential-realizational approach (Stump 2001:1–30), and present the data as inflectional paradigms. There are good theoretical reasons for this (recapitulated in Corbett 2015a:147–148). Equally, for those who believe paradigms are merely a useful way of presenting data, we shall see that in this particular case they prove a very useful exploratory as well as presentational device.

2.1 *Scissors*: semantics vs (syntax and morphology)

Consider this example:

- (1) These scissors are blunt

The noun *scissors* is what most think of as a plurale tantum noun. It fits the definition given by Matthews since it is plural in form. It also takes plural agreement. Hence its morphology and syntax line up. It is a plurale tantum noun because on the obvious reading of examples like (1) it denotes a single entity, so that its semantics is out of step with its syntax and morphology. However, besides denoting one item, it can denote more than one, as is clear in (2):

- (2) All these scissors are blunt.

In both (1) and (2), syntax and morphology are aligned. In the instances where one item is denoted (as can be the case in (1)), the semantics does not align with the syntax and morphology.

While the English examples are well known,² there are many comparable but less familiar instances. Thus in Cicipu, a Benue-Congo language of northwest Nigeria,

²Payne and Huddleston (2002:340–349) give a fine account of English, including good lists, though forms and agreement are not always fully differentiated. They refer (2002:340) to the *scissors* type (those ‘denoting objects made up of two like parts’) as ‘bipartites’ and note that these are sometimes called ‘summation plurals’. This is the usage in Quirk et al. (1985:300–304), who also give extensive

with about 20,000 speakers (in 1995), there is one pluralia tantum noun. The noun *à-húlá* ‘name’ has a plural form, plural agreements, and this is so whether it denotes one name or more than one (McGill 2007:55, 2009:244); for examples see McGill (2009:145, 173, 187).³

Definitions: Not surprisingly *scissors* fits Matthews’ definition given above, but instances like (2) do not fit within Trask’s definition. However, while *scissors* is the most familiar type of example, there are other possibilities, which are more challenging for the definition, as we shall see.⁴

2.2 Tsez *xex-bi* ‘child(ren)’: (semantics and syntax) vs morphology

The noun *xex-bi* ‘child(ren)’, from the Dagestanian language Tsez, provides an illuminating contrast with nouns like English *scissors*, as seen in Comrie (2001). For the essential background, consider first a regular Tsez count noun:

(3) Regular Tsez noun *besuro* ‘fish’ (Comrie et al. 1998:6–7)

	SINGULAR	PLURAL
ABSOLUTE	besuro	besuro-bi
ERGATIVE	besur-ā	besuro-z-ā
GENITIVE 1	besuro-s	besuro-za-s
DATIVE ^a	besuro-r	besuro-za-r
.	.	.
.	.	.
.	.	.

^aFor the relation of dative to (al)lative in Tsez see Comrie and Polinsky (1998:104).

Note: a vowel is dropped before a following vowel.

lists. For a model-theoretic approach, see Ojeda (2001). I would still draw attention to two entertaining blog posts: Simon Thomas “12 nouns that are always plurals”: <http://blog.oxforddictionaries.com/2015/05/12-nouns-that-are-always-plurals/>, and Arika Okrent “11 nouns that only have a plural form” *Mental Floss* <http://mentalfloss.com/article/52672/11-nouns-only-have-plural-form> September 12, 2013.

³McGill (2009:253) treats this noun as belonging to an inqorate gender, but I see it rather as a number problem (it lacks a singular form).

⁴Crystal (1995:201), writing about English, states that ‘Many nouns do not show a contrast between singular and plural: the *invariable nouns*. These are usually classified into two types: those used only in the singular, and those used only in the plural.’ He provides useful lists, including ‘A few nouns which look singular but are always plural: *vermin, livestock, cattle, poultry, people, folk* and *police*.’ However, ‘invariable’ works for English, but not for languages with more extensive nominal morphology. In languages like Russian (Sect. 2.4), pluralia tantum do not show a contrast between singular and plural, but they typically have a full set of case forms in the plural and hence are not invariable. Compare also Nickels et al. (2015:293), talking specifically about nouns like *scissors*: ‘*Scissors* is an example of a class of words sometimes known as *pluralia tantum* (e.g. *pliers, goggles, trousers, etc.*). These words are morphologically and grammatically plural whilst conceptually singular.’ Again, they can be conceptually singular, but do not need to be, as example (2) demonstrates.

These are just some of the case values in Tsez (Comrie and Polinsky 1998); the many further case values are indicated by the dots in (3) and (4). It is against that background that we should consider *xexbi* ‘child(ren)’ (Comrie 2001:381–383, discussed in Corbett 2007b:31–38).

(4) Paradigm of Tsez *xexbi* ‘child(ren)’

	SINGULAR	PLURAL
ABSOLUTIVE	xex-bi	xex-bi
ERGATIVE	xex-z-ā	xex-z-ā
GENITIVE I	xex-za-s	xex-za-s
DATIVE	xex-za-r	xex-za-r
.	.	.
.	.	.
.	.	.

Comparing with *besuro* ‘fish’ in (3) above, we see that *xexbi* ‘child(ren)’ is plural in form (*-bi* and *-za-* are clear plural markers); it has a full set of plural case forms. The paradigm as presented in (4) suggests that *xexbi* ‘child(ren)’ is both singular and plural. The evidence comes both from semantics: *xexbi* ‘child(ren)’ may denote one or more children, and from syntax: this noun takes the appropriate agreements, singular for one and plural for more than one. (And at least in terms of number the noun γ^5 *anabi* ‘woman/women’ behaves similarly.) To confirm this, we look at the agreement system of Tsez: agreement involves four gender values as well as two number values. Assignment of these gender values is by a combination of semantic and formal criteria. The main semantic assignment rules are included in (5), from Polinsky and Comrie (1999:110); more detail can be found in Plaster et al. (2013). Here then are the agreement forms for Tsez verbs:

(5) Gender and number agreement markers in Tsez verbs

GENDER	SINGULAR	PLURAL
I (male humans)	∅ ^a	b-
II (female humans, and some inanimates)	y-	
III (animals, and some inanimates)	b-	r-
IV (residue)	r-	

^aSince (5) gives agreement affixes, the contrast with the overt affixes elsewhere in this paradigm means that using a ∅ here makes good sense. Where full word forms are given, these will be glossed as bare stems (using [] to indicate the relevant values).

The singular-plural syncretisms in this system are tricky for our purposes. However, there is in addition the demonstrative, which distinguishes singular and plural, and Comrie (2001) shows that there are therefore clear diagnostics for singular versus plural.

(6) The Tsez demonstrative *howdu* ‘this’ (Comrie 2001:380)

GENDER	SINGULAR	PLURAL
I (male humans)	howda	
II-IV all others	howdu	howziri

Examples (7)–(9) include the forms for singular nouns (gender values III and I) and for the plural of gender value I.

Tsez (Comrie 2001:381–383)

- (7) howdu k’et’u b-ik’i-s.
 this.II/III/IV.SG.ABS cat(III)[SG.ABS] III.SG-go-PST.WIT
 ‘This cat went.’
- (8) howda uži ik’i-s.
 this.I.SG.ABS boy(I)[SG.ABS] [I.SG]go-PST.WIT
 ‘This boy went.’
- (9) howziri uži-bi b-ik’i-s.
 this.PL.ABS boy(I)-PL.ABS I.PL-go-PST.WIT
 ‘These boys went.’

While the verb forms are the same in (7) and (9), the demonstrative distinguishes the two situations. Given these diagnostic environments we turn to *xexbi* ‘child(ren)’. Example (10) is exactly as we might expect:

- (10) howziri xex-bi b-ik’i-s.
 this.PL.ABS child-PL.ABS I.PL-go-PST.WIT
 ‘These children went.’

Here more than one child is referred to and the agreements are plural (specifically gender I plural, which includes groups consisting only of males or of a mixture of males and females). Now consider what happens for one child. There are two possibilities. Traditional usage is as follows:

- (11) traditional usage (III in the singular)
 howdu xex-bi b-ik’i-s.
 this.II/III/IV.SG.ABS child-PL.ABS III.SG-go-PST.WIT
 ‘This child went.’

The combination of agreements shows that we are dealing with gender III singular: *howdu* ‘this’ is singular, the *b-* marker on the verb could indicate gender III singular or gender I plural, so the only consistent specification is gender III singular.⁵ This is understandable when we note that cross-linguistically, ‘child’ is often treated as

⁵There are, of course, nouns which control inconsistent agreements (discussed in Sect. 4.4). For *xexbi* ‘child(ren)’, however, there is an obvious feature specification; the agreements in (11) are exactly as for a regular noun of gender III, as in (7). The combination of gender III (singular) and gender I plural is less usual, and the syncretism of these agreement forms may have had a role in the development of the use of plural forms for denoting a single individual. Agreements with γ^5 *anabi* ‘woman/women’ are more straightforward: gender II (singular) and gender II–IV plural (Comrie 2001:382).

not quite male human or female human. The noun *xexbi* ‘child’ itself is unchanged. Younger speakers, on the other hand, have gender I.⁶

- (12) younger speaker’s version (I in the singular)
 howda xex-bi ik’i-s.
 this.I.SG.ABS child-PL.ABS [I.SG]go-PST.WIT
 ‘This child went.’

In (12) there is again one child in question, and the agreements are singular. More generally, the agreements are singular or plural as appropriate, but the forms of the noun stay the same (the noun takes plural inflectional forms and the appropriate case value). Given the substantial inventory of case values in Tsez, *xexbi* ‘child(ren)’ has an appropriate inventory of case forms, but with just one number form for each of them; the inflectional forms it has are recognizably plural. This is important: the absolutive has the unmistakable plural marker *-bi*, and all the oblique forms have the plural augment *-za-*. This noun has therefore more evidence for its plurality than, say, *scissors* or *oats*. However, unlike nouns like English *scissors*, it takes singular and plural agreements according to the meaning.

As pointed out earlier, γ^5 *anabi* ‘woman/women’ has the same behaviour in terms of number; but importantly its gender agreements are different. For single referents it takes gender II (like all female human nouns), and when used as a plural it takes non-male-human agreement (again like all female human nouns). Thus its gender agreements are as expected. This demonstrates, therefore, that the number problem which this noun shares with *xexbi* ‘child(ren)’ cross-cuts the gender issue.⁷

The Tsez data are significant, because in addition to their relevance for number, they are hard to classify in terms of their morphological behaviour. In Corbett (2007b:31–38) it is argued that *xexbi* ‘child(ren)’ falls mid-way between canonical syncretism and canonical deponency, and hence there is no established term for such examples.

Definitions: the Tsez examples fit Matthews’ definition well (once it is extended from ‘plural form’ to ‘plural forms’), in that they have only plural forms, in the morphological sense. Yet this is only half the story, since the forms we are focussing on (those denoting a single entity), fail to control plural agreement.

⁶This innovative use presumably reflects the wider range of gender I as opposed to gender II; if sex mattered, one would use *uži* ‘boy’ or *kid* ‘girl’ rather than *xex-bi* ‘child’ (Bernard Comrie, personal communication, 6 April 2018).

⁷These two nouns *xexbi* ‘child(ren)’ and γ^5 *anabi* ‘woman/women’ are high on the Animacy Hierarchy (Sect. 5), unlike most pluralia tantum nouns. I suggest that this is possible, at least in part, because they are of the type where syntax is aligned with semantics (though we need further examples to corroborate this). As we shall see (Sect. 5), there is an implicational relation between syntactic and morphological properties in respect of the Animacy Hierarchy. We should ask, then, how other pluralia tantum nouns behave in Tsez. According to Bernard Comrie and Maria Polinsky (personal communications, January 2017), there are no pluralia tantum of the *scissors* type. There are compounds of the type *hawa-baq*, literally ‘air sun’, which means ‘weather’ (Polinsky 2015); such compounds take plural agreement, of the gender predicted from the two nouns involved (*hawa* is gender IV, and *baq* gender III), and the agreement is gender II–IV plural). Such compounds may involve animates or inanimates, but the agreement is that predicted from the combination of the two items.

We return in Sect. 2.4 to the way in which the paradigms are presented. A related phenomenon, but more complex than the Tsez type, is found in Archi (analysed in Sect. 4.5).

2.3 Russian *galife* ‘riding breeches’: semantics vs syntax vs morphology

Russian nouns distinguish six indisputable case values, found in both singular and plural. (Further, more contentious values are discussed in Corbett 2012:200–222.) Nouns fall into four main inflection classes; I illustrate with nouns of inflection class II, since this class most readily demonstrates the need for six case values.

(13) A normal and a plurale tantum noun in Russian

	<i>bol'nica</i> ‘hospital’		<i>nožnicy</i> ‘scissors’	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOMINATIVE	bol'nic-a	bol'nic-y	–	nožnic-y
ACCUSATIVE	bol'nic-u	bol'nic-y	–	nožnic-y
GENITIVE	bol'nic-y	bol'nic	–	nožnic
DATIVE	bol'nic-e	bol'nic-am	–	nožnic-am
INSTRUMENTAL	bol'nic-ej	bol'nic-ami	–	nožnic-ami
LOCATIVE	bol'nic-e	bol'nic-ax	–	nožnic-ax

The noun *bol'nica* ‘hospital’ has the expected singular-plural opposition, with the plural forms used to refer to more than one entity. There are various syncretisms, but it is clear which is the set of plural forms. Note that for our definitions we need to refer to ‘plural forms’ here, unlike definitions framed for English. Clearly *nožnicy* ‘scissors’, has plural forms (comparable to those of *bol'nica* ‘hospital’) but no singular forms. It is similar to English *scissors*, in that it can be used of one entity or more than one, and it takes plural agreements.

Russian also offers a less common type of plurale tantum. Consider these two nouns:

(14) Uninflecting nouns in Russian

	<i>pal'to</i> ‘coat’		<i>galife</i> ‘riding breeches’	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOMINATIVE	pal'to	pal'to	–	galife
ACCUSATIVE	pal'to	pal'to	–	galife
GENITIVE	pal'to	pal'to	–	galife
DATIVE	pal'to	pal'to	–	galife
INSTRUMENTAL	pal'to	pal'to	–	galife
LOCATIVE	pal'to	pal'to	–	galife

These are unusual paradigms. *Pal'to* ‘coat’, like many similar nouns, does not inflect. Yet it is not defective. It can stand in all syntactic environments appropriate for nouns. And where agreement is required, the agreement targets all agree appropriately. We may treat nouns like *pal'to* ‘coat’ as constituting a separate inflection class

(as in Corbett and Fraser 2000:308), or we may treat them as different from inflecting nouns. The essential point is that their inflectional behaviour (rather their lack of it) is a morphology-internal matter. It has no repercussions in syntax.

We can now see the significance of *galife* ‘riding breeches’. Like *pal'to* ‘coat’, *galife* ‘riding breeches’ can appear in syntactic environments requiring each of the case values. Yet it is also a plurale tantum noun, in the sense that it takes only plural agreements (Isačenko 1962:77), as here:

- (15) Russian
 v ser-yx galife
 in grey-PL.LOC riding.breeches
 ‘in grey riding breeches’ (From Ju. Trifonov, *Dom na naberežnoj*)

This fits a pattern: several other nouns in Russian, denoting bipartites, are pluralia tantum nouns of the *scissors* type (having plural morphology and syntax, as in Sect. 2.2). The effect of the pattern can be seen in the treatment of the borrowing *džins-y* ‘jeans’, where Russian plural morphology is added to the original plural form (-y is the nominative/accusative plural, and there is a full plural paradigm), making it a plurale tantum noun.⁸ This means that *galife* has the common semantics-syntax mismatch (it can denote a single object, but takes plural agreement) and in addition its morphology does not match the syntax (it does not show plural inflection, indeed any inflection).

Definitions: the key point is that *galife* ‘riding breeches’ indeed has the plural only, but this relates to its agreement requirement and not to morphological form.

This restricted application of the definition is reasonable, since we typically understand that definitions apply to the extent that is possible. Given that *galife* ‘riding breeches’ does not inflect, it is a plurale tantum noun to the extent that this is possible (in terms of agreement).⁹ While in Russian it is exceptional to be uninflecting, it can be regular to be uninflecting (as we shall see in Sect. 4.5), and then there can still be pluralia tantum nouns.

2.4 Types of feature: morphosemantic and morphosyntactic

Let us take stock. We have encountered three types of pluralia tantum nouns, summarized in Table 1.

These examples have already presented difficulties of definition. While it seemed straightforward to say that pluralia tantum nouns have only the plural form, with that

⁸Ora Matushansky points out (personal communication, 27 March 2018) that slang *lavè* ‘money, dough’ is also an uninflecting noun which can take plural agreements (following the basic *den'gi* ‘money’ which is an inflecting plurale tantum); there are also some instances of its taking singular agreement, however.

⁹Similarly, animate nouns in Russian are defined morphosyntactically in terms of the syncretism of case values (accusative syncretic with genitive) both in their inflection and in their agreements (see Corbett 2012:160–162 for details). Uninflecting nouns like *kenguru* ‘kangaroo’ are treated as animate since they fit the definition to the extent that is possible, that is, in terms of agreement. They provide a good comparison for *galife* ‘riding breeches’, which is as much a plurale tantum noun as an uninflecting noun can be, namely in terms of agreement.

Table 1 Summary of the types of pluralia tantum nouns in Sect. 2

Example	Section	Mapping
<i>scissors</i>	2.1	semantics \neq (syntax = morphology)
Tsez <i>xexbi</i>	2.2	(semantics = syntax) \neq morphology
Russian <i>galife</i>	2.3	semantics \neq syntax \neq morphology

raising questions for expected count nouns like *scissors*, less so for non-count nouns like *oats*, it is not self-evident how to demonstrate that *scissors* is indeed the plural. Linguists point to the agreements required, though this does not relate directly to the form. And then, in some of the data presented above, the layout of the paradigms was surprising. It is time to confront these issues.

Let us take up a Russian example, since the existence of several case forms makes it easy to recognize singular and plural forms.¹⁰ The Russian *sani* ‘sledge’ has only forms which are recognizably plural in their morphology. However, it is not a problem to refer to a single sledge, as with the phrase *odn-i san-i* one-PL sledge-PL ‘one sledge’ (well attested in the Russian National Corpus). Note that the numeral stands in the plural (see Sect. 4.6). There are two ways we might represent this noun:

(16) Representations of Russian *sani* ‘sledge’

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

If we ask how this noun realizes the grammatical meanings singular and plural, that is, how the semantics and morphology are linked, the answer is the morphosemantic representation on the left of (16). Each possibility can be realized, but the outcome is the same for singular and plural. So long as we consider only the “internal” picture, this is an appropriate representation. The phonological shape of the inflections, compared with regular nouns, makes it clear that the forms are what would elsewhere be used only for the plural. If we turn to the external requirements of *sani* ‘sledge’, then the morphosemantic representation is insufficient. Even when denoting a single object, this noun requires plural agreement, as mentioned above: *odn-i san-i* one-PL sledge-PL ‘one sledge’. For the external or morphosyntactic nature of *sani* ‘sledge’, the representation on the right is appropriate.

We tend to assume that the morphosemantic and morphosyntactic feature values associated with any paradigm cell are the same. In the canonical world they are. In many actual instances they are too, but there are also numerous examples where they are not. This possible discrepancy is what leads to some of the confusion with plu-

¹⁰The example is famous from Zaliznjak’s insightful discussion (1967/2002:57–61, 75–80).

ralia tantum nouns, and we shall see other examples where the distinction is needed below (Miya in Sect. 5 and Bayso in Sect. 6.4). For previous discussion of morphosemantic versus morphosyntactic features, see Corbett (2012:49–50) and Spencer (2013:219–232), and on their relation to the contextual-inherent distinction, see Corbett (2012:67). Note that in terms of Stump’s paradigm types, the difference between morphosemantic and morphosyntactic features would require splitting his content paradigm, which is the ‘interface with syntax and semantics’ (Stump 2016:104; see also Stump 2007).

In the earlier examples, I gave morphosyntactic representations. These are striking both in the case of Tsez *xexbi* ‘child(ren)’, where the difference compared with Russian *sani* ‘sledge’ in (16) was shown clearly, and in the case of Russian *galife* ‘riding breeches’ (14), where an uninflecting noun has specific morphosyntactic requirements. I continue to give morphosyntactic representations below, except where specified otherwise.

2.5 Prerequisites for a fuller typology

The examples so far demonstrate that the typology of pluralia tantum nouns is more extensive than most researchers have allowed for. Here I preview what will be required for the fuller account below.

2.5.1 Further criteria required

In brief, we need to lay out the criteria which allow us to calibrate the examples already discussed, and see whether these criteria allow for further types. To keep the scope manageable, I will concentrate on instances which are clear in terms of the lexical entries involved. There are monographs to be written on the lexicographical problems of related lexical entries, where an arguable instance of a pluralia tantum noun may be related to a noun with a full paradigm. For instance, in Russian we find *vybory* ‘election’ (Soboleva 1984:67), which has only the plural in this sense. It is related to the noun *vybor* ‘choice’, which has a full paradigm. And there are many others, some harder to recognize, and the choices are not clear-cut. The issue was recognized by Wackernagel (1920:86–88), and numerous following researchers;¹¹ thus Payne and Huddleston (2002:334–338) include relevant examples from English where polysemy is the appropriate analysis. Where such an analysis is implausible, my approach would be to say that in a default inheritance lexicon there can be two entries, *vybor1* and *vybory2* in the case above. These inherit some of the same information (including the stem and the inflectional type) but they have different lexical indices.¹² In this paper, I concentrate on setting out the extremes, particularly the clear pluralia tantum, and so will not go further on this point here.

¹¹ See Saas (1965) for this issue related to Classical Greek, in the light of Moorhouse (1966).

¹² Canonically a noun has a single sense. (Recall that canonical is to be distinguished from frequent and from prototypical, Corbett 2012:154–155.) Less canonical is to have more than one. In which case, it is more canonical if the relations between the senses are principled (e.g. animal and meat of animal, Apresjan 1974:23) and if the rest of the lexical entry is the same for all senses (most relevant for us is that the same number values are available for all senses, realized by the same inflectional morphology).

2.5.2 *The typology of number*

There has been progress on the typology of number, and some of this research will be cited as we work through the different aspects of pluralia tantum. For a general survey of number see Corbett (2000), with an extensive bibliography. For a helpful annotated bibliography see Acquaviva (2014), and for fine surveys see Acquaviva (2017) and Moravcsik (2017).

There are two key components of the typology. First there is the featural part, the values of number. Here we must remember that there are more number values than just singular and plural, and that these other values (of which the most common is the dual) can illuminate what is going on with pluralia tantum. And second, we must be clear about the part of the noun inventory involved in any generalization. There is typically a distinction between nouns for which the number system is clearly relevant, the count nouns, and those which do not make number distinctions, the non-count or mass nouns. Cross-linguistically, they are distributed according to the Animacy Hierarchy, which we examine in Sect. 5. As we noted in Sect. 1, nouns of both types can be pluralia tantum, count nouns like *trousers* and non-count like *oats*.

2.5.3 *The notions “canonical number” and “canonical noun”*

From what we have already seen, it is evident that different criteria are in play. We need a typology of features and of nouns (both individual nouns and groups of nouns), and even beyond nouns (Sect. 9). It will be helpful to have baselines from which to calibrate the variation, and here the notions “canonical number” and “canonical noun” will prove invaluable (Sect. 3).

3 A canonical approach to pluralia tantum

The central strategy of Canonical Typology has been summarized by Stump (2016:31):

- (a) to identify the dimensions of possible cross-linguistic variation in a given phenomenon and the logical endpoints of these dimensions; and
- (b) to situate canonical types at the extremes of these dimensions, calibrating attested phenomena according to the degree and direction of their deviation from these canonical types.

A valuable effect of our strategy is that: ‘the canonical approach breaks down complex concepts in a way that clarifies where disagreements may lie between different linguists and theoretical frameworks.’ (Nikolaeva 2013:100). Canonical instances, those that are clear and indisputable, must match a full set of criteria. It follows that such instances may well be infrequent, hence canonical is to be clearly distinguished from prototypical. Canonical instances can even be non-existent; this is not a problem: we are adopting an axiomatic approach, designed to ensure that we recognize and investigate the full range of the phenomena we wish to account for, and that we have a metalanguage for describing them. This canonical approach is justified by utility and results. Recent research within this framework includes Kwon and

Round (2015), Forker (2016a), Corbett and Fedden (2016), Audring (2017), Corbett et al. (2017), Fedden and Corbett (2017, 2018), Round and Corbett (2017), Stump (2016, 2017), Kwon (2017), Fedden et al. (2018) and Evans et al. (2018); a working bibliography can be found on the Surrey Morphology Group website.¹³ A good deal has already been done within this approach, that we can use directly. We take the research on canonical features that is relevant to number (Sect. 3.1), the work on canonical parts of speech, as relevant to nouns (Sect. 3.2), and integrate them (Sect. 3.3).

3.1 Canonical number

First consider canonical morphosyntactic features in general. These have been described in terms of two overarching principles, and these principles in turn cover ten converging criteria (Corbett 2012:156–199):

Canonical morphosyntactic features—Principle I (evidence from form)

Features and their values are clearly distinguished by formal means (and the clearer the formal means by which a feature or value is distinguished, the more canonical that feature or value).

This principle is obviously relevant to our needs; a feature which is realized through various formal means (such as number in Russian) is more canonical than one where it is limited to agreement, as we shall see in Sect. 4.5.1.

Canonical morphosyntactic features—Principle II (syntax)

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

This principle is less central for our concerns, but it will still prove useful; simple syntax includes consistent agreement, for which see Sect. 4.4 and Sect. 8.4.

Besides these two, for the realization of morphosyntactic features there is an additional general principle of inflection (Corbett 2012:197–198):

Canonical morphosyntactic features—Principle III (morphological realization)

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

This principle too is of use, in distinguishing some less obvious examples; we noted the non-canonical morphological behaviour of Russian *galife* ‘riding breeches’, which fails to inflect (Sect. 2.3).

It has been argued that number is the morphosyntactic feature that comes closest to the canonical idealization of a morphosyntactic feature (Corbett 2013:57). The argument depends, however, on the interaction with parts of speech, to which we now turn.

¹³<http://www.smg.surrey.ac.uk/approaches/canonical-typology/bibliography/>.

3.2 Canonical noun

We need the notion of a canonical part of speech (or ‘word class’ or ‘lexical category’). We may think of a part of speech primarily in terms of its syntax, and also in terms of its semantics and its morphology. In the canonical situation, equivalent to what Spencer calls the ‘morphologically coherent lexicon’, these three specifications are fully aligned (Spencer 2005:102). Specifically for nouns, this means that canonical nouns denote referential entities, head an appropriate syntactic phrase and inflect for the relevant features.

3.3 Interaction: canonical number and canonical noun

Intuitively, the greater the degree of orthogonality shown by a feature, with parts of speech and with other features, the clearer the argument for recognizing a feature, and the more canonical it is. The arguments are laid out in Corbett (2013:52–57) and discussed in Corbett and Fedden (2016:499–503); they are reported briefly here.

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

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

This is a general principle, whose implications can be spelled out in four criteria (Corbett 2013:52–57):

Canonical morphosyntactic features—Principle IV: Criterion 1: EXCLUSIVENESS

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

This criterion will be useful; in particular, it leads us to be wary of analyses which would treat pluralia tantum nouns as being specified for gender rather than number (discussed in Sect. 4.6).

Canonical morphosyntactic features—Principle IV: Criterion 2: EXHAUSTIVENESS

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

In the canonical instance, a clear motivation for postulating a morphosyntactic feature is that it generalizes across a large number of items. If this is the case, without the feature we would need to multiply the number of lexical items according to the values of the feature, and would miss generalizations. This criterion is vital for calibrating potential pluralia tantum nouns. A canonical noun “should” have all values of number available to it; as we shall see, there are different ways of being non-canonical here (see particularly Sect. 6.3 on Yup’ik).

Canonical morphosyntactic features—Principle IV: Criterion 3: OPEN AND CLOSED CLASSES

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

This idealization is necessary for the logic of the system, but it will not figure large in our discussion. It underlies the intuitive distinction between part of speech and morphosyntactic feature. In the canonical instance, features, values and parts of speech are all closed classes, and only lexical items constitute an open class.

Canonical morphosyntactic features—Principle IV: Criterion 4: COMPOSITIONALITY

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

We separate out features and their values because of the regularity involved. In the canonical instance, if we can specify what a noun means, and what plural means, we know what the plural of that noun means. This principle is vital for delineating our problem. As well as being relevant for pluralia tantum nouns, it distinguishes various types of lexical plurals from semantically regular plurals.

4 The number behaviour of a canonical noun

We can put together two previously established notions: that of canonical number (as an instance of a canonical morphosyntactic feature) and that of canonical noun (as an instance of a canonical part of speech). This will provide what we require for a typology of pluralia tantum nouns. Given different deviations from canonicity, we can situate pluralia tantum nouns (of different types) within an overall scheme. We discuss the five criteria (Sects. 4.1–4.5) and then contrast canonical number with canonical gender (Sect. 4.6).

4.1 Exhaustiveness

Every lexical item of every part of speech has available to it all values of all features. (alternatively: every feature value applies to all lexical items)
(Criterion IV.2 above, Corbett 2013:54)

This gives a canonical point from which to calibrate. Items can be non-canonical by lacking one or more number values, and this in two main ways. First, the missing values may be predictable or not. A predictable gap (for instance, one consonant with the Animacy Hierarchy, Sect. 5) is closer to canonical than an idiosyncratic gap. And second, we should ask how many values are missing. In a two-value system (singular and plural), one missing is equivalent to one present, but in larger systems (for instance, including a dual), there can be different degrees of deviation from exhaustiveness.

4.2 Semantic-syntactic-morphological alignment

In the canonical world, an item's semantic, syntactic and morphological properties all line up (Corbett 2013:52, Sect. 3.2 above). This alignment holds also at the level of their feature values. The importance of this criterion to our typology should already be clear, from our initial examples in Sects. 2.1–2.3. The matching of features is often assumed, but we noted significant mismatches between morphosemantic and morphosyntactic feature values in Sect. 2.4.

4.3 Semantic compositionality

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

(Criterion IV.4 above, Corbett 2013:56–57)

Since the canonical use of number is compositional, as in this criterion, the various types of lexical plurals are all non-canonical; see Acquaviva (2008, 2016a), Gardelle (2016), Lammert (2016) and Lauwers and Lammert (2016), among others. This criterion gives no reason to expect that the values should be anything other than “equal”, and hence equally distributed. Of course, we know that this is not what we find. We need to look at the various situations where there is an imbalance, to see whether there is less than totally free choice of number, and hence the combination of number value and lexical meaning is not fully compositional.¹⁴

4.3.1 Dominance

The notion of dominance, and its importance for psycholinguistic research, was introduced in Baayen et al. (1997:97). They refer to contrasting distributions where ‘Either the singulars were much more frequent than their plurals (singular-dominant) or the plurals had a much higher surface frequency than their singulars (plural-dominant pairs)’.¹⁵ An illustration of such distributions is given in (17). In the CELEX database (Baayen et al. 1995), the lexemes *neck* and *lip* are approximately matched: each occurs around 80 times per million words. Yet their distribution with respect to number is dramatically different:

(17) English examples per million words (Davis et al. 2003:430)

	SINGULAR	PLURAL
neck	72	7
lip	17	61

Like most nouns, *neck* occurs more frequently in the singular. The frequencies for *lip* are different: it occurs much more frequently in the plural; it is said to be ‘plural dominant’. Note the wording ‘much more frequent’ in the paper cited. We must bear in mind that overall, the singular occurs more frequently than the plural, in language after language. Thus for a noun to appear in the plural more often than in the singular, by whatever degree, is worthy of note, while for one to appear more often in the singular is of interest only when it is in excess of the general preference for singular. Evidence for this general pattern is found in Greenberg (1966:31–32), who gives data on French, Russian, Latin and Sanskrit, and the list is expanded in Corbett (2000:281–282) to include Slovene and Upper Sorbian. For those (Indo-European) languages

¹⁴For further discussion of compositionality in inflection see Stump (2015). And see Panevová and Ševčíková (2011) for discussion of Czech nouns whose plural typically denotes a pair or small group, such as *sírky* ‘matches’.

¹⁵This research was taken further, notably in Baayen et al. (2003); and more recently in Beyersmann et al. (2015).

the singular is used typically in over 70% of the instances in running text (somewhat lower in Upper Sorbian); 70% is also the approximate figure reported for English (Haskell et al. 2003:125), while 79% is reported for Finnish texts (Räsänen 1979:24), and 91% for the Dagestanian language Hinuq (Forcker 2016b:93).

We should look beyond basic systems, and think of dominance with regard to values besides the plural. Here we consider dual dominance. The status of the dual varies considerably cross-linguistically; in languages with a facultative dual, like Slovene, the number of instances in running text is very low (Corbett 2000:281–282).¹⁶ A language with an excellent claim to having an obligatory dual is Sanskrit.¹⁷ The statistics in (18), derived from Lanman (1872:583), give a clear picture of the distribution of the number values (note that Lanman includes nouns, adjectives, participles and nominal forms from pronominal stems):¹⁸

(18) Distribution of number on nominals in Sanskrit (based on Lanman 1872:583)

	N	SINGULAR (%)	DUAL (%)	PLURAL (%)
MASCULINE	57,950	72.9	5.7	21.5
FEMININE	15,909	55.2	6.0	38.8
NEUTER	19,418	75.2	0.6	24.3
total	93,277	70.3	4.7	25.0

Note: percentage figures are rounded, and may not sum to 100

The total number of examples is substantial. Yet even though the dual is obligatory, it is relatively infrequent. Lanman's remarkable corpus work sets a baseline against which we can measure the anomalous behaviour of particular nouns. Two intriguing examples are given in (19), this time with data from the Digital Corpus of Sanskrit (comprising some four million words, tagged manually).¹⁹

(19) Two Sanskrit nouns with interesting frequency distributions for number

	N	SINGULAR %	DUAL %	PLURAL %
baseline (Lanman)	93,277	70.3	4.7	25.0
<i>karna</i> 'ear'	2,173	89.2	7.9	2.9
<i>upānah</i> 'sandal'	50	12.0	86.0	2.0

Compared with the baseline, *karna* 'ear' has a surprising distribution, but not what we might have expected. It does indeed have a raised frequency of dual use, but this is overshadowed by its very low plural frequency. On the other hand, *upānah* 'sandal' is clearly dual dominant. These examples demonstrate that there is still much to be

¹⁶And for Homeric Greek, Diver (1987) gives useful statistics; the dual is less used with natural pairs than with other references to two items.

¹⁷'The dual number is in regular use and of strict application, the plural practically never referring to two objects.' (MacDonell 1927:180).

¹⁸I am very grateful to Greg Stump for expert help with the Sanskrit data.

¹⁹<http://kjc-sv013.kjc.uni-heidelberg.de/dcs/index.php> accessed 27.7.2017.

learned about individual distributions, particularly in systems with more than just a singular-plural opposition.

Besides its importance for psycholinguistic research, value dominance has more obvious effects, of which we discuss three.

The first is **resistance to change**. Following Greenberg (1966:29), Tiersma (1982:841) showed that plural dominant nouns can maintain irregularities in the plural, when these are regularized with other nouns. He discussed the phenomenon in terms of ‘local unmarkedness’. Consider the native irregular plurals of English. Here we find *men, women, children, oxen, geese, mice, lice, feet* and *teeth*. It is well known that frequent items preserve irregularity; here, however, it is not the frequency of the lexeme, rather the high frequency of the plural as opposed to the singular.²⁰

It is not just individual exceptions that are involved in resistance to change. In Serbo-Croat,²¹ there is an ongoing change in one inflection class affecting many hundreds of nouns. Nouns which otherwise inflect like *prozor* ‘window’ (plural *prozori*) have an augment in the plural, according to various conditions, a main one being that being monosyllabic favours having the augment: *grad* ‘city’, plural: *grad-ov-i*, *film* ‘film’, plural: *film-ov-i*. There are various complications, but one set of exceptions to the generalization is relevant here. Ethnonyms retain the short plural (without augment), for instance *Kurd* ‘Kurd’, plural *Kurdi*. Units of measurement behave similarly, as in *volt* ‘volt’, plural *volti*, as do a few isolated nouns like *zub* ‘tooth’ plural *zubi*. All of these share the property that they are more likely to appear in the plural. For details and sources see Baerman et al. (2017:93–98).

The second effect is the **tendency for the dominant form to be borrowed into other languages**, also pointed out by Tiersma (1982). Thus Russian has borrowed English *rail* (for railways) in the plural as *rel’s*, but treats this as the singular and gives it the regular plural *rel’s-y*. The role of dominance, among other factors, is demonstrated in code mixing by Hakimov (2016); his case study is the plural marking of German nouns inserted into Russian speech.

And third, there is the **pattern of marking**. Consider these examples from the Cushitic language, Arbore:

(20) Arbore (Hayward 1984:159–183, Corbett 2000:17)

SINGULATIVE	GENERAL	PLURAL
–	keléh ‘gelded goat(s)’	keleh-mé ‘gelded goats’
–	garlá ‘needle(s)’	garlá-n ‘needles’
tiis-in ‘a maize cob’	tíise ‘maize cob(s)’	–
lassa-n ‘a loaf’	lássa ‘bread’	–
nebel-inté ‘a hen ostrich’	nebel ‘ostrich(es)’	–

²⁰We expect this to apply to other number values too; for instance, the impact of dual dominant nouns in diachrony, as the dual category is lost, is shown by Belić (1932) for the Slavonic family, and by Žolobov and Kryš’ko (2001:62–67) and Maruyama (2011, 2014) for Old Russian.

²¹Following the 2017 ‘Deklaracija o zajedničkom jeziku’ (<http://jezicinalizmi.com/deklaracija/>), I treat Serbo-Croat as one polycentric language, with four standards, Bosnian, Croatian, Montenegrin, and Serbian. See Corbett and Browne (2018) for a linguistic outline and Bugarski (2012) for the sociolinguistic background.

For some nouns, Arbore has general number forms (the number value is unspecified, see Corbett 2000:9–19), and this may be contrasted with a specified number value, singular or plural, depending on the noun. And even in more straightforward singular-plural systems, we may find examples where there is a base form and then one of the number values has a marker. Haspelmath and Karjus (2017) argue that there is a crosslinguistic tendency for such asymmetries to follow usage frequency. In our terms, plural dominant nouns are more likely to have a singulative (a singular with overt morphologically marking), while singular dominant nouns (the usual type) are more likely to have the normal plural marking. For statistical data on Middle Welsh, consistent with this view, see Nurmio (2017:70–71), and for Lopit, an Eastern Nilotic language, see Moodie (2016).

Given the variation in the proportions of number values we have seen, some quite extreme, we should ask whether full compositionality is maintained. Within the variation, it is important to retain the distinction between pluralia tantum, for which the singular is not available, and plural dominant nouns, for which the singular is rare, but fully available. For instance, for real world reasons Italian *spaghetti* appears frequently in the plural. In appropriate circumstances, say discussing how one strand of spaghetti breaks, the singular *spaghetto* is current and normal.²²

4.3.2 Value orientedness

Value orientedness is a semantic restriction, pointed out for Russian by Polivanova (1983) and Kulikov (2004:127), following work by Igor' Mel'čuk. There are nouns which have all the appropriate number forms, but do not use them in an unconstrained way. Rather, one value is used by default, while the other is found only 'under pressure', so to speak. Note that the default value, the one to which the particular noun is oriented, varies from noun to noun:

- (21) Russian (Kulikov 2004:127)
 rep-a doroža-et
 turnip.SG.NOM get.dearer-3SG
 'turnips are getting dearer'
- (22) ogurc-y dorožaj-ut
 cucumber-PL.NOM get.dearer-3PL
 'cucumbers are getting dearer'

Each of these nouns has forms for both values; however, unless there is specific contextual pressure (what Polivanova terms an 'arithmetical context'), *repa* 'turnip' appears in the singular, and *ogurcy* 'cucumbers' in the plural, as in the same neutral context in (21) and (22). It is not just that one value is more likely (as with dominance), but rather that there are contextual limitations on one value (which is not the

²²See http://bressanini-lescienze.blogautore.espresso.repubblica.it/2008/03/26/lequazione-degli-spaghetti/?refresh_ce, with thanks to Anna Thornton (personal communication).

case with the items in (17) showing number dominance). For further discussion and additional data see Mel'čuk (1979) and Ljaševskaja (2004:28–30, 50, 227–236).

To return to the general criterion of compositionality, these examples are some way from canonical, while pluralia tantum represent greater non-canonicity. Examples of value orientedness are not fully compositional; rather it is as though 'general number' (see discussion of (20)) intrudes for some senses of these nouns. Note again that these nouns argue for a featural account (not a morphemic one) since the same considerations apply across the different morphological realizations (Corbett 2015a:148). These nouns are on the border of number differentiability, showing that this count-mass border can be gradient.

4.4 Syntactic consistency

Syntactic consistency falls under the general scope of Principle II (see Sect. 3.1), namely that the use of canonical morphosyntactic features and their values is determined by simple syntactic rules. But there is a more specific criterion that is particularly relevant, namely that in agreement it is canonical for features to have matching values rather than non-matching values (Corbett 2006:24). Given the external requirements of the noun, in the canonical world that is all that is required by the syntax: a feminine plural noun is feminine plural "whoever is asking". However, there are various types of noun for which this is not the case: their agreement specification varies according to the agreement target (which is not simple syntax). Such nouns are termed hybrids. Familiar examples are English *committee* and *family*, which take only singular attributive modifiers but may take singular or plural for other agreement targets (Corbett 2012:94–105, 2015b; Smith 2017).

We might expect that pluralia tantum would be consistently plural. But there are instances which are hybrids. Thus in Finnish, according to Karlsson (1968) a singular predicate is normal with plural proper nouns like *Yhdysvallat* 'the United States', *Filippiinit* 'the Philippines'. But in attributive position, he reported both singular and plural adjectives, with singular on the increase. Hannu Tommola (personal communications, 5 June 2017) states that the singular is now normal, though occasional plurals are found in attributive position.²³ This shows the path, the Agreement Hierarchy (Corbett 2006:206–237), along which a hybrid plurale tantum can be regularized in terms of its agreements, and become singular in all but morphological form. We expect first the anaphoric pronoun to take the singular, then the relative pronoun, next the verbal predicate, and only finally the attributive modifier (the stage being reached by Finnish).²⁴

A truly remarkable example of an item with inconsistent agreements is Modern Hebrew *be'alim* 'owner(s)', as analysed by Landau (2016). The singular *ba'al* is ambiguous between 'husband' and 'owner'. In the plural, if we focus on the 'owner' meaning, the form *be'alim* 'owner(s)' can have a singular or plural referent.

²³For discussion of (these) *United States* in English see <http://languageog.ldc.upenn.edu/nll/?p=1794> and <http://languagehat.com/these-united-states/>.

²⁴For more on pluralia tantum nouns in Uralic languages see Bayırlı (2017).

Modern Hebrew (Landau 2016:984)

- (23) ha-be'al-im ha-kodem maxar et ha-makom lifney šana.
 the-owner-PL the-previous.SG sold.3SG ACC the-place before year
 'The previous owner sold the place a year ago.'
- (24) ha-be'al-im ha-kodm-im maxru et ha-makom lifney šana.
 the-owner-PL the-previous-PL sold.3PL ACC the-place before year
 'The previous owners sold the place a year ago.'

The plural agreements in (24) are as expected. Yet in (23) we find singular agreement, for a single referent, even though the noun is plural in form. So far, this is similar to the Tsez situation, except that the lexeme in question is split: there is a singular form *ba'al*, but with rather different properties.²⁵ In addition, however, *be'alim* 'owner' is a hybrid: it does not take consistent agreements, rather its agreements vary according to the agreement target. In what follows, for simplicity, I present only those of Landau's examples which are relevant for a single referent. The first examples show agreement of different targets, namely attributive modifier and verb:

Modern Hebrew (Landau 2016:985)

- (25) ?ha-be'al-im ha-kodm-im maxar et ha-makom lifney šana.
 the-owner-PL the-previous.PL sold.3SG ACC the-place before year
 'The previous owner sold the place a year ago.'

Example (25), with syntactic agreement of the attributive modifier and semantic agreement of the verb, was considered 'slightly off' by some speakers. Yet it is much better than (26), with the reverse pattern of agreement:

Modern Hebrew (Landau 2016:985)

- (26) *ha-be'al-im ha-kodem maxru et ha-makom lifney šana.
 the-owner-PL the-previous.SG sold.3PL ACC the-place before year
 'The previous owner sold the place a year ago.'

This relation between examples like (25) and (26) was confirmed in a web search of examples. The pattern is consistent with the Agreement Hierarchy, since we find syntactic agreement within the noun phrase and semantic agreement of the predicate. We should check the personal pronoun:

Modern Hebrew (Landau 2016:987)

- (27) hine ha-be'al-im ha-xadaš. hu/*hem kvar al koc-im
 here the-owner-PL the-new.SG. He/They already on thorn-PL
 'Here is the new owner. He's/*They're already on edge.'

Example (27) shows that the personal pronoun must show semantically justified agreement, and this too is in accord with the Agreement Hierarchy. Finally we find stacking of modifiers:

²⁵Having a split paradigm, with one part being a hybrid, is unusual; it is significant for showing how broad the definition of 'possible lexeme' needs to be. For other split hybrids, see Corbett (2015b).

Modern Hebrew (Landau 2016:1005)

- (28) ha-be'al-im ha-pratiy-im ha-axaron šel ha-tmuna haya
 the-owner-PL the-private.PL the-last.SG of the-painting was.3SG
 ha-psixo'analitika'i Jacques Lacan.
 the-psychoanalyst Jacques Lacan
 'The last private owner of the painting was the psychoanalyst Jacques Lacan'
- (29) *ha-be'al-im ha-prati ha-axron-im šel ha-tmuna
 the-owner-PL the-private.SG the-last.PL of the-painting
 haya/hayu ha-psixo'analitika'i Jacques Lacan.
 was.3SG/was.3PL the-psychoanalyst Jacques Lacan
 'The last private owner of the painting was the psychoanalyst Jacques Lacan'

Example (28) is a textual example, with syntactic agreement inside semantic agreement. The reverse, as in (29), is not acceptable.

Be'alim 'owner' shows a combination of non-canonical behaviours. I have stressed its status as a hybrid; it does not take consistent agreements, rather the agreement specification depends in part on the target, and the pattern of possible agreements is in accord with the Agreement Hierarchy. It is not a straightforward plurale tantum noun, in that the singular exists, albeit with rather different properties.

4.5 Morphological distinctiveness

Recall from Sect. 3.1 'Canonical morphosyntactic features—Principle I', according to which morphosyntactic features and their values are clearly distinguished by formal means. They have dedicated forms, they are distinguished across other features and their values, furthermore they are distinguished consistently across parts of speech and across lexemes within the parts of speech (Corbett 2012:155–167). This is clearly relevant to some of the examples we have discussed. It allows us to differentiate typologically within the pluralia tantum, and to situate other non-canonical behaviours in respect of number. We separate out those items that simply fail to inflect (Sect. 4.5.1), those which do inflect but in a non-canonical (non-distinctive) way (Sect. 4.5.2), and those with split paradigms (Sect. 4.5.3).

4.5.1 Uninflecting nouns

We have already seen uninflecting nouns as pluralia tantum, for instance, Russian *galife* 'riding breeches' (Sect. 2.3). This noun lacks any sign of number morphology, but is morphosyntactically plural, requiring plural agreement. This is a lexical exception, in a language where nouns typically inflect for number. We will see similar examples in Kiowa (Sect. 6.2). We now examine the numerous pluralia tantum nouns in Walman, a Torricelli language of Papua New Guinea. In Walman few nouns mark number, and there are many nouns which behave like Russian *galife* 'riding breeches'. The data are from Dryer and Brown (2015, and personal communications). Consider first what we find with ordinary count nouns:

Walman (Dryer and Brown 2015)

- (30) Pelen n-aykiri.
dog 3SG.M-bark
'The male dog barked.'
- (31) Pelen w-aykiri.
dog 3SG.F-bark
'The female dog barked.'
- (32) Pelen l-aykiri.
dog 3SG.DIMIN-bark
'The puppy barked.'
- (33) Pelen y-aykiri.
dog 3PL-bark
'The dogs barked.'

There are four clearly distinct agreement markers; these appear on various targets, including personal pronouns, sometimes as prefixes, sometimes as suffixes, even as infixes. Note that in these examples there is no inflection on the noun; this is fully expected in Walman, since only a minority of nouns (mainly denoting humans) have plural forms. However, the agreement markers distinguish singular and plural. Given this background, the following are the significant examples:

- (34) Apar pa<y>ten y-o rachi.
bed that<PL> 3PL-be strong
'That bed is strong.'
- (35) Tim lapo-y y-anan kon.
dew big-PL 3PL-go.down night
'A heavy dew fell last night.'

In (34) and (35) we see no noun inflection (as expected) and we find plural agreement on two agreement targets, including subject agreement on the verb. The nouns *apar* 'bed' and *tim* 'dew' are pluralia tantum, just in terms of agreement. The following examples demonstrate object agreement on the verb:

- (36) Kipin k-ao-y nyi ngo-ny
We 1PL-shoot-OBJ.3PL fire one-PL
'We lit a fire.'
- (37) Kum m-aro kaintuen m-a-y wi alpa-ny.
I 1SG-take[OBJ.3SG.F] stick 1SG-use-OBJ.3PL hand one-PL
'I grabbed the stick with one hand.'

Again the agreements are all plural, including those of the numeral (see Sect. 7.1 on plural agreeing numerals). Note that *wi* 'hand' (like similar body parts) is a pluralia tantum noun even when denoting one item. Dryer and Brown (2015) have identified more than 80 pluralia tantum nouns, more indeed than they have found nouns of masculine gender.

Definitions: The situation in Walman and similar languages is important for rebalancing the definition of pluralia tantum. The Walman examples are pluralia tantum to the extent that is possible, that is, morphosyntactically (for agreement). They are similar to Russian *galife* ‘riding breeches’, but while that is exceptional within Russian morphology, the Walman examples are morphologically regular, since Walman nouns typically do not inflect for number.

4.5.2 Non-distinctive inflection

An extreme type here is Tsez *xexbi* ‘child(ren)’. We noted in Sect. 2.2 that each of its inflectional forms is uniquely identifiable as a plural form, within the morphological system of Tsez. And yet each of its forms can be used to denote a single entity, and take singular agreement. Thus for this noun, singular and plural are not distinguished inflectionally.

In a canonical system, then, plural inflection is uniquely for plural. Given an inflected form, it can be identified as plural, or as not plural. In a less canonical situation, given a form, it can be identified as plural, but only within the *set* of forms. Thus in a language with inflection classes, like Russian, a single inflected form may not be diagnostic, but the set of forms is uniquely identifiable (in (16) the nominative plural form *sani* ‘sledge’ could in principle be genitive singular, but the full set of forms is clearly plural, and makes that interpretation impossible). In a language with limited inflection, there are examples which are even less canonical. In English, given a noun ending in *-s*, there is a probability that it is plural.²⁶ But there are nouns like *economics*, *linguistics* (Jul 1975:18–24; Payne and Huddleston 2002:347), for which the model could be *sticks* vs *stick* or else *mix* vs *mixes*. Typically these nouns are treated as singular (despite the existence of the adjectives *economic* and *linguistic* which suggest that the noun has an *-s* affix). The system would be more canonical if either there were no nouns like *mix* or *linguistics*, or else there were larger paradigms, so that other forms of the paradigm were sufficient to make the distinction clear, and establish whether a given item has a stem-ending divide.²⁷

4.5.3 Split paradigms

Let us turn to further examples which fill out the typology of pluralia tantum, and then briefly consider other number effects which are non-canonical according to this criterion. In the canonical world, paradigms are consistent, but of course we find many instances of paradigms that are split (Corbett 2015a). In a sense, all pluralia tantum nouns have a split paradigm, in that a part of it is lacking. But here we look

²⁶An intriguing question is whether there is any phonetic distinctiveness. Since Plag et al. (2017) have shown that *-s* inflections assumed to be the same actually have different phonetic properties, we can wonder whether the final segments of regular plural nouns like *sticks*, nouns like *economics*, and pluralia tantum like *pants* are distinguishable phonetically.

²⁷Since part of the aim is to show the typology of non-canonical number, into which pluralia tantum nouns fit, we should consider the Amharic situation, described by Kramer (2016:529–530). Every noun which has an irregular plural also has a regular one, with no change in meaning: for instance, *k'al* ‘word’, *k'al-at* (irregular plural), *k'al-ot/ftf* (regular plural). Moreover, the two can be combined (though not all speakers accept all possibilities), again with no change in meaning according to Kramer. Thus *k'al-at-ot/ftf* ‘words’. The relevance is that nouns which have (irregular) plural marking can attract further number marking, as though the original marking were not sufficiently distinctive.

at the intersection of pluralia tantum with splits. That is, there are instances of split paradigms, where a noun is partially but not fully a pluralia tantum noun. Note first that place names are a rich source of pluralia tantum nouns, often with a relatively clear origin. Thus in English we have the impressive *Himalayas* and the rather smaller *Cotswolds*. However, certain Finnish place names like *Parainen* add a new twist. They were pointed out by Karlsson (2000); the information here is from Tommola (2018 and personal communications). There are several dozen such place names, including *Jokioinen*, *Kaipainen* and *Oravainen*. The clear facts are that the nominative is singular in form (the hypothetical plural **Paraiset* is not accepted), while most of the remaining cases have plural markers:

Finnish (Karlsson 2000)

- | | | | |
|------|--|----|---|
| (38) | a. Parais-i-lla
Parainen-PL-ADESSIVE
‘in Parainen’ | b. | Parais-i-lle
Parainen-PL-ALLATIVE
‘to Parainen’ |
|------|--|----|---|

Karlsson states that the partitive is singular (as in *kohti Parais-ta* towards Parainen-PARTITIVE ‘towards Parainen’), which for Tommola sounds strange, and he points to examples on the web with the plural: *kohti Parais-i-a* towards Parainen-PL-PARTITIVE ‘towards Parainen’. In the accusative we find the singular (*Paraisen*). Unfortunately, agreeing modifiers appear to be avoided so we cannot investigate the morphosyntax of these split paradigms.

The split can be more complex, leading to a remarkable combination of non-canonical behaviours in respect of distinctiveness, as in the Archi noun *χáli* ‘family’. The data come originally from Kibrik (1977:28–29n18).²⁸ Archi nouns have a singular-plural distinction, indicated by different stems. There are many case values, with a clear divide between the absolutive and the oblique case values. By default, the absolutive singular consists of the bare stem, and the plural is marked with *-mul* (consonant final stems) and *-tu* (vowel final stems). The ergative serves as the stem for the remaining oblique case values (for all but certain kinship terms with a different oblique stem, Kibrik 1977:21). While there are several possible oblique markers, the defaults are: singular: *-li* (or, for substantivized adjectives, *-mu* or *-mi*); plural: *-čaj* (consonant-final bases), *-taj* (vowel-final bases), *-maj* (substantivized adjectives). The important point is that, given a form, we have a clear prediction as to its number. Consider then the key cells of the paradigm of *χáli* ‘family’.

(39) Archi *χáli* ‘family’ morphosemantic paradigm (partial)

	SINGULAR	PLURAL
ABSOLUTIVE	<i>χáli</i>	<i>χáli-tu</i>
ERGATIVE	<i>χáliə-maj</i>	<i>χáli-taj</i>

Note: the forms are as in the Archi dictionary (<http://www.smg.surrey.ac.uk/archi-dictionary/?LI=78>), with a schwa in the ergative singular as in the sound file. This is not included in Kibrik’s example below.

²⁸I thank Marina Chumakina for help here. The data are discussed in Baerman et al. (2006), and Corbett (2007b:39–41); audio files are available at: <http://www.smg.surrey.ac.uk/archi-dictionary/?LE=102&WE=3146>.

Given the information about defaults in the Archi noun paradigm, we have expectations about the forms here. *Xáli* is the morphologically unmarked form, and so should be the absolutive singular (as it is). *Xáli-tu* is a regular absolutive plural. The form *χáli-t:aj* is the ergative plural (and the base for the oblique plural case values) which we would expect for a noun with a vowel-final stem. But what of *χálə-maj* (the shaded cell)? This has an ergative plural marker (albeit one expected of a substantivized adjective). Note that it is unproblematic to add the default ergative marker *-li* to a stem in *-li* with non-final stress; for instance, *dáli* ‘stick’ with ergative *dállli*, so *χáli* would be expected to have the ergative **χállli*, rather than *χálə-maj*. When we turn to agreement, we find that the first three cells discussed control the expected agreement, singular or plural, as shown by modifiers within the noun phrase. What then of the ergative singular *χálə-maj*? This takes plural agreement, even though it denotes a single family. Here are Kibrik’s examples:

- (40) Archi *χali* ‘family’ in the oblique (Kibrik 1977:28–29n18)
 misginn-ib χal-me-n lo
 poor-PL family(III)-OBL.PL-GEN son[SG.ABS]
 ‘son of a poor family’
- (41) Archi *χali* ‘family’ in the absolutive
 misginnu-b χali [not *misginn-ib χali]
 poor-III.SG family(III)[SG.ABS]
 ‘poor family’

This pattern recalls that discussed in Tsez (Sect. 2.2), but there are two differences to note. First that there is a split in the singular, with the absolutive behaving normally and the ergative (and the remaining oblique cases) showing unexpected behaviour. And second, the form in the ergative singular cell, while plural in form, is not the same as that in the ergative plural (as was the case in Tsez). Still, this plural form (although singular in meaning) controls plural agreement. Thus in terms of agreement, this noun is normal in the absolutive, but it takes only plural agreement in the oblique cases. It has a split paradigm, and is a plurale tantum noun in the oblique cases (in terms of agreement). The oblique singular cells (see the shaded cell in (42)) are morphosemantically singular, but morphosyntactically plural (for this reason, exceptionally, (39) above is a morphosemantic paradigm, while generally I give morphosyntactic paradigms). The morphosyntactic requirements are as in (42):

- (42) Archi *χáli* ‘family’ morphosyntactic requirements (partial)

	SINGULAR	PLURAL
ABSOLUTIVE	SINGULAR	PLURAL
ERGATIVE (and other oblique)	PLURAL	PLURAL

As (42) makes clear, *χáli* ‘family’ is a plurale tantum noun, in its forms and in its agreements, but only in the numerous oblique cases; its paradigm is split, and the absolutive is normal.

- (45) Jedn-a kol-a su stig-l-a.
 one-N.PL.NOM carriage(N)-PL.NOM AUX.3PL arrived-PST-PL.N
 ‘One carriage arrived.’

Serbo-Croat has three gender values, with distinct agreements in the singular and the plural. Examples (43)–(45) show a pluralia tantum noun of each gender, with the appropriate gender agreement in the plural (including on the numeral *jedan* ‘one’, compare Sect. 7.1). Thus the issue of pluralia tantum nouns cross-cuts gender. Furthermore, each noun illustrated belongs to an inflection class containing many other nouns (which have singular and plural); there is nothing inflectionally different about the example nouns, except their lack of morphologically singular forms, of course.³⁰

5 Count–mass and the issue of motivation

A key point for discussing whether pluralia tantum nouns are motivated or not is the expected distribution of number marking in a given language. There is considerable cross-linguistic variation in the balance between count and non-count nouns. This variation is constrained by the Animacy Hierarchy (Smith-Stark 1974, following earlier proposals). The modified version here is from Corbett (2000:54–88) where there is a good deal of relevant data, extended in Daniel (2005) and Haspelmath (2005); for values apart from the plural see Corbett (2000:89–132).

- (46) The Animacy Hierarchy

speaker > addressee > 3rd person > kin > human > animate > inanimate (1st person (2nd person pronouns) pronouns)
--

Smith-Stark proposed that plurality can be ‘a significant opposition for certain categories but irrelevant for others’ (1974:657). Where this occurs, it will affect some top segment of the hierarchy. He adduced two types of evidence: marking of the noun phrase for number (on the noun itself) and agreement in number.

There are languages where the threshold for number-differentiability is much higher than in languages like English. In others, countability can extend right down the scale, well below the English system, as for instance in Yudja (Juruna family), a Tupi language of Brazil (Lima 2014). There can be intra-language variation too, as shown by the fact that the cut-off point for countability varies across the different varieties of English (Ziegeler 2010; Schmidtke and Kuperman 2017), and that there are degrees of countability (Allan 1980:562–563).³¹ While the Animacy Hierarchy constrains the possibility of number differentiability, there are also differences in the frequency of number marking according to the hierarchy (Brown et al. 2013). And

³⁰There is unusual data on pluralia tantum nouns in Serbo-Croat, in Kostić (1987:74–77). Taking a corpus of newspaper texts and poetry (almost two million words of running text), Kostić gives frequencies for noun occurrences by gender, number and case. The general picture, with few exceptions, is that nouns in the singular are most frequent, then come ‘ordinary’ plurals, and finally pluralia tantum nouns.

³¹Allan identifies eight classes of noun, including pluralia tantum; examples, ranked by level of countability from highest to lowest: *car, oak, cattle, Himalayas, scissors, mankind, admiration, equipment*.

see Farmer (2015:80–109) for discussion of its relevance, for number marking and other phenomena, especially with respect to the Western Tukanoan language Máíhĩki.

For pluralia tantum nouns, the issue of the two types of evidence is particularly relevant. On the evidence of morphological marking, English *sheep* does not fit the constraint of the Animacy Hierarchy; on the other hand, in terms of agreement it is in accord with the Hierarchy. And this is a general picture. Hence the following claim (Corbett 2000:67):

Lexical items may be irregular in terms of number marking with respect to the Animacy Hierarchy and regular in terms of agreement, but not vice versa.

Consider again Tsez *xexbi* ‘child(ren)’ from Sect. 2.2. In terms of number marking this noun does not conform, since singular and plural are identical. However, in terms of agreement it is fully regular, following the claim above.³²

Besides lexical examples, like *sheep* and *xexbi* ‘child(ren)’, there are principled sets of examples where the split on the Animacy Hierarchy differs, according to marking vs agreement. These are highly relevant to the point that internal (morphosemantic) and external (morphosyntactic) feature values can differ (as discussed in Sect. 2.4). A telling example is provided by the West Chadic language Miya. Outline data are given here, from Schuh (1989, 1998), discussed in Corbett (2006:177–179). Miya has two gender values, masculine and feminine, and two number values, singular and plural. In terms of number marking, there is an animacy distinction: animate nouns are those which denote ‘all humans, most, if not all, domestic animals and fowl, and some large wild animals.’ There is a grey area consisting of nouns denoting large wild animals, while the remaining nouns are inanimate (Schuh 1989:175). For animate nouns marking of plurality is obligatory.

There are several different agreement targets, and these have three agreement forms: masculine singular, feminine singular and plural. We can see this with a demonstrative:

(47) The demonstrative ‘this’ in Miya (Schuh 1998:243)³³

	SINGULAR	PLURAL
MASCULINE	nákən	níykin
FEMININE	tákən	

With animate nouns, agreement is as we would expect; when plural they take plural targets, like *níykin* ‘these’ in (47). With inanimate nouns, however, the situation is rather different. First, marking is optional (Schuh 1989:175), which is most evident according to Schuh in numeral phrases. Here singular and plural are both possible, but examples like (48b) are more common in texts (Schuh 1998:258):

³²Indeed pluralia tantum denoting animates are rare (recall footnote 7), but see Landau (2016:987n6) on the Hebrew word for God, and Zaliznjak (1967/2002:59–60, 99–100) for Russian examples.

³³˘ indicates low tone, and ˊ high tone; ə is a high central vowel, and tl is a lateral fricative. The basic analysis is in Schuh (1989); the presentation of forms is changed slightly in Schuh (1998) and we follow the later version here.

Miya (Schuh 1998:198)

- | | | | | |
|------|----|---|----|---|
| (48) | a. | zàkiy-áyàw vaatlə
stone-PL five
'five stones' | b. | zàkiy vaatlə
stone[SG] five
'five stones' |
|------|----|---|----|---|

Second, and more important for us, even when marked as plural, inanimate nouns do not take plural agreement. But agreement is not blocked. The target shows different agreement forms, tracking the noun's gender in the singular:

Miya (Schuh 1998:193n6, 197)

- (49) nákən víyayúw-awàw
this.M.SG fireplace(M)-PL
'these fireplaces'
- (50) tákən tlərkáy-ayàw
this.F.SG calabash(F)-PL
'these calabashes'

Thus, agreement in number with animates is obligatory, plural agreement with inanimates is impossible, whether or not they are morphologically marked as plural. Agreement with inanimate plurals does occur, but in gender not number. This means that in examples (49) and (50) the noun is clearly plural for internal (morphosemantic) purposes, while for agreement purposes (its morphosyntax) it requires singular agreement, and hence the target can show agreement in gender.

Returning to the central point of the section, we note that there is a substantial semantics literature on the count–mass distinction including: Link (1983), Jackendoff (1991), Borer (2005:86–135), Chierchia (2010), Landman (2011, 2016), Massam (2012), Nicolas (2016), Rothstein (2017:82–116) and Grimm (2018); Bale and Barner (2011) is a helpful bibliography, and Lasersohn (2011) includes clear discussion of the varying use of terms (including the issue of whether mass nouns are necessarily singular, as held by some). A valuable resource for English is described in Kiss et al. (2016). In a recent contribution, Sutton and Filip (2017) account for the variability in the treatment of particular nouns in terms of the competing pressures of individuation and reliability. While the discussion has been primarily in terms of familiar languages, the substantial cross-linguistic variation is beginning to be addressed: Doetjes (2012) extends the discussion, Pelletier (2011, 2012) emphasises the language-specific nature of the count-mass distinction, and Lin et al. (2018), and the papers in their collection, target typologically diverse languages.³⁴

For nouns which are below the threshold for number-differentiability on the Animacy Hierarchy there are different possibilities (we should not assume that they will be singular). Some languages do treat them as singular; thus the Finno-Ugric language Mansi, which has a singular- dual-plural system, makes all its non-count nouns singular (Rombandeeva 1973:40). Others, like Manam, have them all as plural (Lichtenberk 1983:269). And yet others have some singular and some plural, as in Turkana (Dimmendaal 1983:224). There may be regularities in such systems; for

³⁴Further from our concerns in this paper, we should note research on what happens when singular mass nouns are pluralized; see, for instance, Alexiadou (2011).

instance, Andersen (2014:253–254) points out that in Nilotic languages it is common for at least some nouns denoting liquids to be plural (he gives examples from Dinka). If all non-count nouns are plural in a given language, there is no more to be said, but if there is an option, then those non-count nouns which are plural are of interest. And while plural mass nouns are not our primary concern, they fall under Matthews' definition, and indeed he includes *oats* as an example. Ojeda (2005:392) includes a list of English mass plurals, following Jespersen (1954:122–124), and Lauwers (2014) and Lammert (2015) discuss the situation in French; Smith (2016) analyses plural mass nouns in Telugu. The question of the motivation of singular vs plural, within the non-count nouns, has been the subject of lively discussion, starting from the opposition of *oats* and *wheat* (Wierzbicka 1988:499–560, and 1991a, 1991b; Palmer 1990:226–229; Moravcsik 1991:136–139), taken up again in Ljaševskaja (2004:147–150), Goddard (2009) and Wisniewski (2009:181–184).

From mass nouns, let us return to those pluralia tantum which would be expected to be count nouns. The question of motivation here, that is, the issue of which expected count nouns are pluralia tantum, is truly fascinating. For a given language, we expect countability to depend on the Animacy Hierarchy, as just discussed, and the threshold for number-differentiability varies from language to language. Once that threshold is established, the existence of pluralia tantum nouns may be surprising. To be specific, take English *trousers*. We might expect it to be a normal count noun, like *shirt*, *jacket* or *hat*. Or indeed like contemporary French *pantalon* 'trousers'. Yet *trousers* is not a random exception. Williams (1994:13) points out that its behaviour is shared by all nouns denoting items of clothing 'worn on the legs in such and such a way'. This exemplifies what Koenig (1999:1–2) calls a 'medium size generalization'. The generalization is specific to English and yet, of course, equivalents of *trousers* recur cross-linguistically in the sets of pluralia tantum nouns.

Related languages may differ dramatically; thus English has many pluralia tantum nouns (Payne and Huddleston 2002:40–348) while there are few in German. It has been suggested that the prevalence of pluralia tantum nouns is an areal phenomenon, with Circum-Baltic languages having substantial inventories (Vraciu 1976; Koptjevskaja-Tamm and Wälchli 2001:629–637). That research on Circum-Baltic languages produced a useful table of the "usual suspects". The method was first drawing up two lists of 30 items which were pluralia tantum in Baltic languages or in Russian; consolidating those lists gave a list of 56 items. These were translated into 41 languages of Europe. A fraction indicates the proportion of pluralia tantum when the dictionary consulted gave more than one translation. In addition the data are represented as a neighbour net in Wälchli (2011:327). Sadly the raw data are no longer available for further analysis. However, there is new work with the same areal perspective (Tommola 2018).

Table 2 shows clearly that there are items which are frequently pluralia tantum; at the same time, seeing them as naturally pluralia tantum is unwarranted because they are often not in this category. It is helpful to think of such phenomena in terms of two dimensions (cf. Bye 2015:107): first they can be predictable/rule governed vs unpredictable; and second, they can be functionally motivated/natural vs arbitrary.

Consider first the opposition predictable or rule governed versus unpredictable. There are instances where there appears to be an exceptionless or almost exceptionless generalization, as in the claim cited earlier, that items of clothing worn like

Table 2 Frequency count of pluralia tantum in 41 languages of Europe (Koptjevskaja-Tamm and Wälchli 2001:631)

Meaning	Languages with plurale tantum equivalent
'glasses'	24
'trousers'	22.5
'tongs'	20
'scissors'	19.5
'firewood'	14.5
'measles', 'Christmas'	14
'horse-race'	13.5
'money', 'slops'	13
'chicken-pox'	11.5
'gate', (major) place names	11
'christening'	10.5
'wedding', 'hide-and-seek'	10
'bran', 'sleigh', 'brain', 'door'	9.5
'organ'	9
'yeast', 'troubles'	8.5
'car(t)', 'sorcery'	8
'balance/scales', 'funeral', 'rake', 'childbirth'	7.5
'debate', 'jungle'	6.5
'mane,' 'thickets'	6
'ashes', 'straw'	5.5
'flour', 'oat(s)'	5
'twilight', 'clock/watch'	4.5
'salad', 'twenty-four hours', 'ceiling'	4
'lunch'	3.5
'smoke', 'foam', 'wheat', 'ink'	3
'rye', 'saliva' 'cream'	2.5
'dung/manure', 'fat', 'shirt', 'south', 'laughter'	2
'anger/wrath'	1.5
'year'	1

trousers will be pluralia tantum in English (Williams 1994:13). There are many areas, however, where predictability is less clear, and the degree of predictability is rather in the eye of the linguist. This leads to the question of whether pluralia tantum nouns are defective. Let us start from Matthews' (1997:89) definition of a defective lexical item as one 'whose paradigm is incomplete in comparison with others of the major class that it belongs to.' Baerman and Corbett (2010:2) start from this definition, and point out that '... the more idiosyncratic and lexically restricted the gap, the more canonically defective it is, and the more canonically defective the gap is, the greater the analytical challenge.' Thus the less predictable a plurale tantum noun is, the closer it comes to being defective. In the canonical instances of defectivity there is a resulting gap (as with the lack of the first singular of the verb *pobedit* 'conquer' in Russian).

However, pluralia tantum nouns tend to be less than canonically defective, since there may well be a standard ‘patch’ for what would otherwise be missing. For instance, there are different possibilities for combining pluralia tantum nouns with numerals to obviate constructions that would otherwise be impossible (Sects. 7.1–7.2).

Turning now to the opposition between functionally motivated or natural versus arbitrary, the first observation is that pluralia tantum nouns often seem not to fit the Animacy Hierarchy, which is another possible motivation for linguists to look for semantic explanations. The fit depends on the version of the Animacy Hierarchy used; thus the version proposed by Sasse (1993:659) leads to more problem cases than that given as (46). Table 2 suggests possible arguments for motivation. There can be greater or smaller sub-regularities: so for instance, various languages of the Baltic region have large numbers of pluralia tantum, falling into semantic groups like names of meetings and festivals, and paired cutting tools. There are various lists of semantic groupings including Braun (1930:1–15), Karlsson (2000:648–649), Koptjevskaja-Tamm and Wälchli (2001:630), Kibrik (2003:92–97), Ljaševskaja (2004:92–108) and Acquaviva (2008:19–21). The difficulty is that one can point to the “usual suspects” that appear in several different languages, and equally one can stress the differences.³⁵

While the degree of synchronic motivation for particular instances of pluralia tantum needs careful, dispassionate study, their earlier motivation is often clearer. Take the Serbo-Croat plurale tantum noun *kola* ‘carriage, car’ (as in (45) above), compared with the singular *kolo* ‘wheel’. And Lezgian *surar* ‘cemetery’, which is originally the plural of *sur* ‘grave’. *Surar* ‘cemetery’ is on a trajectory towards losing its plurale tantum status. It can take a numeral (*sa surar* ‘one cemetery’) and it has alternative agreement possibilities (Haspelmath 1993:81–82):

- (51) Lezgian (Haspelmath 1993:82)
 Či šeher-di-n surar jeke-bur / jeke-di ja.
 1 PL.GEN town-OBL-GEN cemetery big-SBST.PL / big-SBST.SG COP
 ‘Our town’s cemetery is large.’
 Note: SBST = substantivizer

Pluralia tantum place names often betray their origin as denoting a significant group of landmarks. (For more on the origins of pluralia tantum, see Degtjarev 1982.)

Motivated exceptions are a lesser departure from canonicity than lexical exceptions. At the most general level we have motivated exceptions to systems of number, which are the items which fall below the threshold for number differentiability, the area on the Animacy Hierarchy in a given language. Then we have Koenig’s medium size generalizations, where all items in a restricted class follow a (perhaps surprising) generalization.³⁶ And least canonical are individual lexical stipulations. For each deviation from canonicity, we can ask whether it is predictable or not, and motivated or not. And specifically, since bipartites have figured large in discussions of motivation, we could ask how they fare in larger systems, including those with a dedicated number value for two referents—the dual. We turn to these larger systems.

³⁵See Corbett (2018) for a Slavonic case study.

³⁶These have been discussed extensively also with regard to gender; see, for instance, Enger (2009) and references there.

6 Reduced expression of number in larger systems

We look at larger number systems, to see how the reduction of possibilities for particular nouns plays out here. The immediate link is that bipartites are often cited as being motivated pluralia tantum. This would lead to expectations about how they will behave in systems with a dual (Sect. 6.1). And more generally, larger systems help clarify the issues discussed earlier. The definitions given in Sect. 1 take the perspective that the items under consideration have only the plural. And this is what pluralia tantum means—having only the plural. But what if they lack the singular? In languages like English the point is moot. However, in languages with more than two values for number there is a difference: one value or more might be lacking. The distinctions we have discussed can now be applied to values other than singular and plural. We look at morphological reduction in a singular-dual-plural system in Sect. 6.2, and to morphological and syntactic reduction in Sect. 6.3. Finally we tackle a four-value system, one with a clear mismatch between morphosemantic and morphosyntactic values (Sect. 6.4).

6.1 The issue of bipartites and the dual

Linguists who stress the motivation of apparent irregularities sometimes point to bipartites, like *trousers*, as being motivated pluralia tantum. We might ask, then, what might happen to such nouns in a language with a dual. If ‘two-ness’ is the motivating factor, we might expect such nouns to occur as dualia tantum. Yet what we often find (in the relatively few instances where there are data) is that the usual suspects turn up, as pluralia tantum. Thus in Slovene, which has a singular-dual-plural system, nouns like *hlače* ‘trousers’ are pluralia tantum and not dualia tantum. Priestly (2006) investigates the issue specifically and states that there are no dualia tantum nouns in Slovene. In the other contemporary Slavonic languages with a dual, Upper and Lower Sorbian, there are numerous pluralia tantum nouns, including bipartites. Thus for Lower Sorbian, Janaš (1976/1984:72) gives examples of different types of pluralia tantum, pointing out that they can be used for one, two, or more than two referents. For Upper Sorbian, which also has numerous pluralia tantum, Faßke (1981: 417) suggests two dualia tantum, *staršej* ‘parents’ and *dwójnikaj* ‘twins’. However, the dictionary by Völkel (1981:87, 405) has the first as a duale tantum, but gives *dwójnik* (singular) ‘twin’. Gerald Stone (personal communication, 16 April 2018) states that the plural *starši* ‘parents’ is in regular use, but that this noun has no singular (like German Eltern ‘parents’, Sect. 8.7), while *dwójnik* ‘twin’ has a full number paradigm (see also Menzel 2018). And, examining the oldest Slavonic texts, Moszyński (1985) concludes that there were no dualia tantum nouns in Old Church Slavonic.

Now Slovene and Sorbian have ‘weak’ duals. What of languages with robust duals? These are more interesting, but for other reasons, as we shall see in Sect. 6.2 and Sect. 6.3. The Papuan language Lavukaleve has pluralia tantum nouns and no dualia tantum (Terrill 2003:130). Yet there are tantalizing notes in the literature, suggesting the existence of dualia tantum, but little detail; see, for example, Boas (1893:59) on Chinook.

6.2 Missing dual forms: Lavukaleve

Lavukaleve is a Papuan language spoken in the Solomons; the data are all due to Terrill (2003). It has a singular-dual-plural system, with obligatory marking of number. There are some pluralia tantum nouns (under 5% in a count of over 800 nouns), both count nouns such as *neo* ‘teeth’ and mass nouns like *ura* ‘steam’ (Terrill 2003:130–132). Such nouns are pluralia tantum in terms of their agreements, they do not necessarily have any morphological marking of plurality. As we noted in Sect. 6.1 Lavukaleve has no dualia tantum. Lavukaleve is of greater importance for us, however. It shows the coverage of our typology, demonstrating how in larger systems there are types of non-canonical behaviour which often slip under the radar. This can be seen in the representative nouns in Table 3.

Table 3 Number marking in Lavukaleve nouns (Terrill 2003:105–130), discussed in Baerman et al. (2017:35)

SINGULAR	DUAL	PLURAL	gloss
taol	taol-al	taol-kal	‘towel’
buti	buti	buti-kal	‘shoe’
fakas	fakas-al	fakas-aul	‘shoulder’
planteson	planteson	planteson-aul	‘plantation’
hoho	hoho-l	hoho-n	‘rafter’
hai	hai	hai-n	‘forehead’
belo	belo-l	belo-vil	‘bell’
kua	kua	kua-vil	‘moon’
kua	kua-ol	kua	‘eye of greensnail shell’
wo	wo	wo	‘war’

Various plural markers are illustrated in Table 3, and for each of them two nouns are given. One with a distinct dual marker and one without (this may have to be specified lexically, as shown by the two nouns with the stem *kua* which behave differently). Each of the nouns without a unique dual is non-canonical (Principle I, see Sect. 4.5). Where there is no unique dual form, the form matches the singular (it is the bare stem).³⁷

We see again that in larger systems, there may be one form lacking (here the dual) rather than only one being available. And in Lavukaleve only the morphological form is lacking: in terms of agreement, the noun with the dual form matching the singular still takes the appropriate agreements.

The non-canonical behaviour of Lavukaleve is taken a stage further in languages with ‘inverse’ systems, as shown for instance by Kiowa (a member of the Kiowa-Tanoan family, spoken in south-western Oklahoma). Its number system has rightly

³⁷The final two items are noteworthy in more general terms, showing that Lavukaleve has nouns which do not inflect for number at all, and that there can be a distinct form for the dual, while singular and plural are syncretic.

attracted considerable research: Early work was done by Wonderly et al. (1954), who introduced the term ‘inverse’, and Merrifield (1959). There is a valuable section in the grammar of Kiowa (Watkins 1984:78–100), taken up in several sources including Noyer (1992:228–236), Watkins (1995), Corbett (2000:159–162), and most recently in Harbour (2007, 2011) and Sutton (2010). Of all the interesting facets of the system, let us concentrate on the inverse marker $-g\grave{o}$ (and its variants, including $-d\grave{o}$). The following two nouns demonstrate its use:

(52) Noun number marking in Kiowa (Watkins 1984:84, 86–87)

	SINGULAR	DUAL	PLURAL
‘horse’	$c\acute{e}:$	$c\acute{e}:$	$c\acute{e}:-g\grave{o}$
‘pole’	$\acute{a}:-d\grave{o}$	$\acute{a}:$	$\acute{a}:$

Nouns like $c\acute{e}:$ ‘horse’, are comparable to those Lavukaleve nouns which do not have a unique dual. And, like Lavukaleve, the Kiowa verb distinguishes singular, dual and plural. Nouns like $\acute{a}:$ ‘pole’, are the opposite, in the sense that there is again no unique dual, but now the dual is syncretic with the plural. And what is striking is that it is the same marker, namely $-g\grave{o}/-d\grave{o}$, which marks the distinct value (hence the term ‘inverse’). This inverse marker makes Kiowa even less canonical in terms of number marking, since the marker $-g\grave{o}/-d\grave{o}$ does not uniquely mark any number values.

The nouns in (52) are representative in that animates behave broadly like $c\acute{e}:$ ‘horse’, in marking the plural. (The noun $t'\acute{a}p$ ‘deer’ does not inflect, but its agreements are like those of $c\acute{e}:$ ‘horse’, which makes it comparable to English *sheep* in a three-value system.) Most nouns denoting inanimates behave like $\acute{a}:$ ‘pole, stick’. We noted in Sect. 5 that there can be a mismatch between the evidence of marking on the noun and the evidence of agreement: the latter is the more regular. We have a similar situation here, in that the marking of number on the animates gives an unexpected plural versus singular / dual system; however, the verb is quite regular in showing singular-dual-plural. Having seen the importance of Kiowa for the issue of limited number marking, we should also note that Kiowa is noteworthy for its pluralia tantum nouns. It has uninflecting nouns which always take plural agreement (irrespective of the number of referents): these nouns include $h\acute{o}ld\grave{a}$ ‘dress, shirt’, $k\acute{u}t$ ‘book, letter, school’, $k^h\grave{o}:\acute{d}\acute{e}$ ‘trousers’ and $t\acute{o}:$ ‘tepee’ (Watkins 1984:90–91).

6.3 Missing singulars: Central Alaskan Yup’ik

Central Alaskan Yup’ik has a more robust dual³⁸ than that of the languages analysed in Sect. 6.1, and it proves particularly significant (for earlier discussion of nouns with reduced number possibilities see Corbett (2000:174) and Mithun (2010:132–135)). Recall our question about what it is that pluralia tantum lack, which is more pressing when there are three number values (singular, dual and plural). Central Alaskan Yup’ik has pluralia tantum nouns like $niicugnissuute-t$ ‘radio(s)’ (Miyaoka 2012:716). Such nouns have a clear inflectional marker of plurality (the $-t$ inflection

³⁸Even this relatively robust dual has ‘leaks’: for one person’s hands or feet it is more usual to use the plural rather than the dual (Jacobson 1995:59), recalling the similar effect in Slovene (Priestly 1993:441).

here), they can refer to one item or more, and they take plural agreement. These are of the type discussed in Sect. 2.1, since their morphology and syntax line up, but their semantics can fail to match (when they are used of a single object).

There are other nouns, however, which have no singular, but do have a dual (and at least some are arguably bipartites (Jacobson 1995:434); thus ‘sled’ as in (53) has two runners):

- (53) Central Alaskan Yup’ik (Miyaoaka 2012:716)³⁹
 Tan’gurraa-m ikamra-k atura-k.
 boy-ERG.SG sled-ABS.DU use-IND.3SG>3DU
 ‘The boy is using the sled.’

The key point is that the boy is using one sled, but the noun stands in the dual and the verb agrees with it in the dual. ‘Box’ is a similar noun:

- (54) Central Alaskan Yup’ik (Miyaoaka 2012:716)
 uku-k yaassiige-k
 DEM-ABS.DU box-ABS.DU
 ‘this box’ or ‘these two boxes’

Miyaoaka is quite explicit that the dual form can denote one or two boxes. If there are more, the plural is used:

- (55) Central Alaskan Yup’ik (Osahito Miyaoaka, personal communication 23.3.2018)
 uku-t yaassiige-t
 DEM-ABS.PL box-ABS.PL
 ‘these boxes’

This matches the view (using a different noun) of Elizabeth Ali and Marianne Mithun, reported in Corbett (2000:174). This means that these nouns are not dualia tantum. Rather they lack the singular. There is extensive data on the related Central Siberian Yupik in de Reuse (2000). He is primarily concerned with the question of motivation. Willem de Reuse confirms (personal communication, May 2012) that he believes the items he cites in the dual can also appear in the plural for denoting more than two referents.⁴⁰

6.4 A larger system: Bayso

The Cushitic language Bayso is significant for our typology, since it has a four-value number system, and this larger system gives novel possibilities for missing

³⁹I use ergative for the traditional ‘relative’ case; > is the Leipzig Glossing Rules indicator of an agent-like argument acting on a patient-like argument.

⁴⁰There is a fully motivated type of noun with dual and plural only, namely dyadic kin terms. Oksapmin (Ok-Oksapmin family, Papua New Guinea) has instances like *almd* ‘grandparent and grandchild’, which takes a dual pronoun, and has the plural *almdil* ‘three or more people in a grandparent-grandchild relationship’ (Loughnane 2009:99–101, and references there to the wider literature).

forms. Moreover there is a dramatic mismatch between the morphosemantic and morphosyntactic systems of number, which raises the issue of definition sharply. Bayso is a Cushitic outlier of Ethiopia, spoken on Gidicco island in Lake Abaya, and in certain villages on the shores of the lake (see Savà 2011 for information on its status). The data are originally from Hayward (1979), discussed in Corbett and Hayward (1987) and in Corbett (2012:224–233), with personal communications from Dick Hayward. The account here develops that work.

Bayso pronouns have a simple system:

(56) Third person pronouns in Bayso (Corbett and Hayward 1987:12)

	SINGULAR	PLURAL
MASCULINE	úsu	íso
FEMININE	ése	

This three-way opposition maps straightforwardly onto all agreement targets, namely the verbal predicate, the demonstrative and the associative particle.⁴¹ Since personal pronouns are at the top of the Animacy Hierarchy (Sect. 5) we use them on principle to determine the appropriate glossing: thus our agreement targets have three forms, masculine singular, feminine singular and plural. Bayso nouns, however, distinguish four values, general, singular, paucal and plural. General number allows expression of the lexical meaning of a noun without reference to number (it may refer to one or more than one); that is, the speaker need not specify the actual number of referents. The singular is used of a particular individual only. The paucal is for a few individuals, from two to about six, and plural is for larger numbers. We now ask how the four values available for nouns map onto agreements. Here are relevant examples for a regular masculine noun (*lúban* ‘lion’) and a regular feminine noun (*kimbír* ‘bird’), each representing many nouns:

Bayso regular nouns (Hayward 1979 and personal communications, Corbett and Hayward 1987)

- | | | | | | | |
|------|----|----------------------------------|------------------|----|----------------------------------|------------------|
| (57) | a. | <i>lúban</i> | <i>hudure</i> | b. | <i>kimbír</i> | <i>hudurte</i> |
| | | lion[GENERAL] | slept.M.SG | | bird[GENERAL] | slept.F.SG |
| | | ‘lion(s) slept’ | (one or more) | | ‘bird(s) slept’ | (one or more) |
| (58) | a. | <i>lubán-titi</i> | <i>hudure</i> | b. | <i>kimbír-titi</i> | <i>hudurte</i> |
| | | lion-SG | slept.M.SG | | bird-SG | slept.F.SG |
| | | ‘a single/particular lion slept’ | | | ‘a single/particular bird slept’ | |
| (59) | a. | <i>luban-jaa</i> | <i>hudureene</i> | b. | <i>kimbír-jaa</i> | <i>hudureene</i> |
| | | lion-PAUCAL | slept.PL | | bird-PAUCAL | slept.PL |
| | | ‘a few lions slept’ | | | ‘a few birds slept’ | |
| (60) | a. | <i>luban-jool</i> | <i>hudure</i> | b. | <i>kimbir-jool</i> | <i>hudure</i> |
| | | lion-PL | slept.M.SG | | bird-PL | slept.M.SG |
| | | ‘lions slept’ | | | ‘birds slept’ | |

⁴¹The mismatch between the pronominal and the noun system fits into a wider typological picture of number systems (Corbett 2000:127–129).

The agreements for general number and singular are the same, for masculines and feminines, and this is a general regularity. Then, nouns of both genders take the masculine singular form when plural. This type of pattern is widespread in Cushitic languages: many have the feminine singular here, and a minority like Bayso use the masculine singular. Bayso is unusual in innovating a paucal number, for which plural agreement is used.

Examples (57)–(60) demonstrate again that the morphosemantic specification of a paradigm cell may not match its morphosyntactic requirement. Thus the morphosemantic plural takes a singular agreement target for all regular nouns. The usual patterns of correspondence between the morphosemantics of the noun paradigm and its external requirements are given in (61):

(61) The system of Bayso regular nouns

Morphosyntax: agreement

		MASCULINE SINGULAR	FEMININE SINGULAR	PLURAL
Morphosemantics: noun number	GENERAL	lúban ‘lion(s)’	kimbír ‘bird(s)’	
	SINGULAR	lubantiti	kimbírtiti	
	PAUCAL			lubanjaa kimbirjaa
	PLURAL	lubanjool kimbirjool		

The shading shows the regular patterns, for masculine and feminine nouns. The left column gives the noun’s number value and the columns specify the agreement it takes; thus *lúban* ‘lion(s)’ is general number, and it takes masculine singular agreements.

Bayso has some non-count nouns, of two types. Some are like *ees* ‘grass’, which takes masculine singular agreement, and its form is best analysed as general number. A small group of others, namely *eenoo* ‘milk’, *ogorroo* ‘hair’, *soo* ‘meat’ and *udú* ‘faeces’ take plural agreement. These are in a sense like English *oats*. But this being Bayso things are not that simple. Taking plural agreement is not the same as being morphosemantically plural. Three of the nouns end in *-oo*, which reconstructs as a plural marker in Proto-Omo-Tana, from which Bayso descends (Corbett and Hayward 1987:19–26); nouns in *-oo* took plural agreement. In the modern language, however, being morphosemantically plural and taking plural agreement are different things. In order to specify the irregularity of *eenoo* ‘milk’, *ogorroo* ‘hair’, *soo* ‘meat’ and *udú* ‘faeces’, the logical way is to mark them as paucal. The regular rules of agreement will ensure that they then take only plural agreement. They are, then, paucal tantum.

A few nouns take irregular agreement patterns, which we mention briefly. There are four nouns which, when plural, take feminine singular agreement instead of the expected masculine singular (of them, three are masculine, like *aar* ‘ox/bull’, and one is feminine, *abba* ‘sister’). Then six nouns take plural agreement when plural (which is aberrant in Bayso); these include *baal* ‘feather/leaf’, which is masculine, with the plural *baalallo*, and *nébe* ‘ear’, feminine, with the plural *nebebboo*.

We now turn to the key examples: those count nouns which are missing a form, within this four-value system (we have discussed the non-count nouns already). There are two types.

(62) Bayso nouns with a missing value—type 1

Morphosyntax: agreement

Morphosemantics: noun number		MASCULINE SINGULAR	FEMININE SINGULAR	PLURAL
	GENERAL			saé ‘cattle’
	SINGULAR			
	PAUCAL			saejaa
	PLURAL	saejool		

Type 1 nouns lack the singular. Their general number form takes plural agreement. The nouns of this type again contain some usual suspects: *kalaljaa* ‘kidney(s)’ has the shape of a paucal (though it has not been recorded without the paucal ending), presumably because kidneys typically occur in pairs; similarly for feet (*lukkaa*), sandals (*keferoo*), eyes (*iloo*) and hips (*moo*). These nouns, then, lack a singular, and require exceptional marking of their general number form as taking plural agreement. Arguably the second fact explains the first: since singulars are formed from general number forms, and retain the same gender, there is no regular general number form from which a singular could be derived. Thus specifying the irregular agreement of the general form is sufficient. Finally, then, our most intriguing example (another usual suspect):

(63) Bayso noun with a missing value—type 2

Morphosyntax: agreement

Morphosemantics: noun number		MASCULINE SINGULAR	FEMININE SINGULAR	PLURAL
	GENERAL			ilkoo ‘tooth/teeth’
	SINGULAR			
	PAUCAL			ilkoojaa
	PLURAL			ilkool

Ilkoo ‘tooth/teeth’ is remarkable. Like type 1 nouns it lacks a singular, and its general form takes plural agreement. Its paucal form, *ilkoojaa*, takes plural agreement as expected. And then its plural form, irregularly, takes plural agreement. This is a noun which lacks only one form, the singular. But because of its irregularities, the effect is that it takes only plural agreements (defined earlier as those which agree with the plural pronoun). In a sense, this is indeed a plurale tantum noun, in terms of agreement rather than form.

Definitions: the definitions given earlier do not specify what is intended by ‘form’. Naturally we expect that (i) cells labelled “plural” in a paradigm (the internal spec-

ification) are indeed plural in terms of their semantics (grammatical meaning), and (ii) control plural agreement (their external specification). But these specifications, internal and external, need not line up. The regular nouns of Bayso demonstrate this clearly. (If there was a total mismatch for all instances, we would assume that the forms were simply mislabelled, but this is not the case for Bayso.) And as we can have pluralia tantum which have only plural forms (but more than one type of number agreement), Bayso *ilkoo* ‘tooth/teeth’ is a plurale tantum noun in the sense that it takes only plural agreement, for all three of its morphosemantic number values.

7 Repercussions of pluralia tantum

The existence of pluralia tantum nouns, of different types, has repercussions through the system. The most striking arise when there is a mismatch of semantics versus syntax and morphology, in particular where a noun which is semantically a count noun, gives syntactic problems when it is to be counted. This may involve special numerals or forms of numerals (Sect. 7.1) or it may require a classifier-like construction (Sect. 7.2). In the opposite situation, where the plurale tantum noun is a mass noun, there may be problems induced by semantic recategorization (Sect. 7.3). Finally, with respect to morphology, compound formation may be difficult if a plurale tantum form is not appropriate within a compound (Sect. 7.4).

7.1 Numerals

Pluralia tantum nouns, when they would be expected to be count nouns, can create problems for numeral constructions, notably in instances where a singular would be expected. Let us start with the numeral ‘one’. In the Slavonic languages we find that this numeral agrees in number with its head, and hence is plural (as noted in Sect. 2.4 above):

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

There is a trickier problem, however. As a result of the loss of the dual in most of the Slavonic languages, some very specific forms are found in numeral phrases. In particular, forms may be used which, for the majority of nouns, are synchronically equivalent to a singular form. In Russian, for instance, we find the noun in the genitive singular, with the numerals *dvaldve* ‘two’, *tri* ‘three’, *četyre* ‘four’ (and with some other items including *obalobe* ‘both’). Besides the restriction to these numerals, the phrase must be in the nominative (or the accusative syncretic with the nominative); see Corbett (2012:209–210) for details and sources. The problem is now clear: numerals may require a form (genitive singular) that a plurale tantum noun cannot supply; specifically, there is no grammatical combination of *dvaldve* ‘two’ and *sani* ‘sledge’ for a phrase in the nominative. However, there is another set of numerals, the

collective numerals.⁴² These have distinct morphology and take the genitive plural; hence they can be used with pluralia tantum nouns:

- (65) dvo-e san-ej
two(COLL)-NOM sledge.PL.GEN
'two sledges'

This usage is largely restricted to the situation where it is needed (the lower numerals in a direct case); elsewhere the normal numerals are much the more frequent. This has been confirmed with corpus and questionnaire research by Nikunlassi (2000:235–239).⁴³

There are further complications, which I will mention briefly (see also Pesetsky 2013:55–56 for discussion). First, not all pluralia tantum nouns behave alike. Mel'čuk (1985:379) and Nikunlassi (2006) show that for bipartites, like *brjuki* 'trousers', the classifier construction (Sect. 7.2) is preferred to the collective numeral. A part of the reason for this distinction may date from the time when collective numerals were used with paired objects, like boots, to enumerate the pairs, a usage which has now largely disappeared (Mel'čuk 1985:385). And second, in complex numeral phrases, it is the last element of the complex numeral which determines the featural specification of the noun. Thus *dvadcat' odin* 'twenty-one' takes the nominative singular, and *dvadcat' dva/dve* 'twenty-two' the genitive singular. But collective numerals cannot appear in complex numerals. As Mel'čuk (1985:378) points out, this means there are written forms, involving complex numerals and pluralia tantum nouns, which cannot be pronounced (also discussed in Pesetsky 2013:141n5). The classifier construction again saves the day, because the classifier can appear in the singular, and then take the plurale tantum noun in the plural.

While I have concentrated on the complexities of Slavonic, Ojeda (1997) gives detailed attention to the collective numerals of Latin, including their use with pluralia tantum (compare also Löfstedt 1958:100–109, 218); Ojeda also considers Icelandic collective numerals in this function. And Miyaoka (2012:716–717) illustrates the use of group numerals in such instances in Central Alaskan Yup'ik.

7.2 The 'classifier' solution

The problem of combining a plurale tantum noun with a numeral is solved in many languages by the use of a classifier, as in this example:

⁴²The conditions on the use of the collective numerals when in competition with the ordinary numerals (that is, with ordinary nouns) are complex; see Suprun (1959:80–83, 1969:105–110), Mel'čuk (1985:384–388), Nikunlassi (2000), Janko (2002:172–180), Timberlake (2004:195–196), Dobrušina and Panteleeva (2008) and Rjabuškina (2015).

⁴³This usage developed during the 19th century (Suprun 1959:81, 1969:105; Nikunlassi 2000:243n22); earlier there were plural collective numerals used with pluralia tantum nouns. This situation is partially retained elsewhere in Slavonic, for instance in Czech (Short 1993:521), and in Serbo-Croat (Stevanović 1975:315; Lučić 2015:5), though being replaced by the classifier construction (Popović 1979:18). See Stefanović (2014) for a detailed account of the quantification of different types of pluralia tantum nouns. Comrie (1992:807–811) discusses Balto-Slavonic more generally, and Koptjevskaja-Tamm and Wälchli (2001:635) demonstrate the prevalence of such numerals in the Baltic area.

Russian

(66) dv-e par-y nožnic
 two-F.NOM pair-SG.GEN scissor[PL.GEN]
 ‘two pairs of scissors’

The lower numerals of Russian require a form that is for almost all nouns identical to the genitive singular, and so the construction with the classifier element *para* ‘pair’ (here in the singular) solves the incompatibility of numeral with plurale tantum noun. However, as the English translation shows, a classifier may be used where there is no obvious syntactic problem. In general, not just in English, there may be restrictions on how readily the classifier construction is accepted, according to how much lexical meaning is retained by the classifier and hence how compatible this is with the particular noun.

7.3 The problem of recategorization

English permits recategorization of mass nouns like *coffee* to count nouns (*coffee* with plural *coffees*) with a portion reading (*a coffee and two teas please*) and a sort reading (*the coffees in this shop are really something*). English is particularly permissive in this regard; the possibilities and the range of meanings of such recategorizations vary from language to language. With pluralia tantum, for mass nouns, it appears that recategorization involving the use of the singular is less readily available.⁴⁴

7.4 Derivational morphology

The canonical plurale tantum noun, having plural inflectional morphology only, would also have only the plural form available for derivation. And indeed, we often find this situation: derivational morphology in English uses the plural form for compounding, as with *almsgiving*, *clothesbrush* and so on (Kiparsky 1982:137–138). However, the singular is found in compound formation with bipartites, as in *trouser pockets*, *spectacle case* and others (Kiparsky 1982:174n3; Arregi and Nevins 2014:316–318). Haskell et al. (2003:132–134) investigated modifiers in compounds, obtaining acceptability ratings for singulars, regular plurals, and pluralia tantum nouns as modifiers (*hammer rack / hammers rack / pliers rack*). The pluralia tantum nouns fell between the singulars (most acceptable) and regular plurals (least acceptable), and were significantly less acceptable than the singulars. This, they claim, is not what would be predicted from a level-ordering account. Note that the potential to use apparent singular forms (bare stems) of pluralia tantum nouns for derivational purposes is a particular fact about English: it is not general.⁴⁵ Thus in Russian the singular **nožnica* from *nožnicy* ‘scissors’ is impossible for any purpose (Zaliznjak 1967/2002:58).

⁴⁴The issues are complex; see Lauwers (2016) for discussion of French, including nouns like *rillettes* ‘rilletes’.

⁴⁵Though it is also found in Finnish (Nenonen and Niemi 2010:107n7); for example, *häät* ‘wedding’ is a plurale tantum noun, but the derivative *häätäpäivä* ‘wedding day’ includes the apparent singular form *hää* (Silva Nurmio, personal communication 8 April 2018).

8 The range of possibilities: defining pluralia tantum

Since different criteria apply to particular examples, it is now worth drawing out the range of variation we have identified.

8.1 Exhaustiveness

A canonical noun “should” have available all values of the number feature. Pluralia tantum nouns are non-canonical against this criterion. But in larger systems we see the range of contrast. In Central Alaskan Yup’ik (Sect. 6.3), there are full pluralia tantum nouns like *niicugnissuutet* ‘radio(s)’, which have only the plural, as opposed to nouns like *ikamrak* ‘sled(s)’, which lack the singular but have dual and plural.

8.2 Semantic-syntactic-morphological alignment

In the canonical world, a noun’s semantic, syntactic and morphological properties are aligned. We have documented pluralia tantum nouns for which this does not hold: *scissors* has the semantics of a count noun, which is out of alignment with its syntax and morphology (Sect. 2.1); Tsez *xexbi* ‘child(ren)’ has its (plural-only) morphology out of line with its semantics and syntax (Sect. 2.2); Russian *galife* ‘riding breeches’ has semantics, syntax and morphology out of alignment (Sect. 2.3). We return to a further instance in Sect. 8.7. These alignment issues can be seen in the light of the more regular misalignments we saw in Miya (Sect. 5) and Bayso (Sect. 6.4).

8.3 Semantic compositionality

In the canonical world, lexical and grammatical meaning (here the meaning of the number value) can be combined in a fully compositional way. We saw potential issues with full compositionality, from number dominance and orientedness through to pluralia tantum. It is not self-evident that pluralia tantum nouns are necessarily non-compositional in terms of number; see, for instance, Acquaviva (2016b) for instances where plural items which appear as non-compositional may yield to more nuanced interpretations.

8.4 Syntactic consistency

Many examples of pluralia tantum nouns are canonical in respect of consistency: they require the same (plural) agreement on all targets. However, pluralia tantum may be inconsistent, taking some singular agreements (as we noted for the Hebrew *be’alim* ‘owner(s)’, and Finnish *Filippiinit* ‘the Philippines’ Sect. 4.4). As singular agreement becomes available for agreement targets at successive positions on the Agreement Hierarchy, this is the route whereby pluralia tantum nouns can transition into being ordinary in respect of number. (Recall also the Lezgian example (51) in this respect.)

8.5 Morphological distinctiveness

In the canonical world, features and their values are fully distinguished by formal means. Again we draw a distinction between all values and particular values. Thus just in terms of number marking we can separate out English *sheep* (no distinction in form between singular and plural) and items like Lavukaleve *buti* ‘shoe’ (Table 3),

where the plural is distinct, but dual is not distinguished from the singular. Relating this to pluralia tantum nouns, we have those like Tsez *xexbi* ‘child(ren)’ (Sect. 2.2), where we have robust plural marking, to those like Russian *galife* ‘riding breeches’ (Sect. 2.3) with no morphological distinctions. The latter exists in a language where number is normally marked formally but not on a minority (including this noun) and at the other extreme there is Walman (Sect. 4.5.1) where nouns in the main do not mark number, and so pluralia tantum nouns are pluralia tantum only in respect of agreement.

8.6 Defining pluralia tantum nouns

On the basis of the observations through the paper, and the criteria in Sects. 8.1–8.5, we can now offer a more robust definition.

Pluralia tantum nouns have only the plural.

- **inflectional criterion:** In inflectional terms, having only the plural means that only the plural sub-paradigm is available. This plural sub-paradigm may consist of one cell or more than one.
 - For count nouns, the mismatch between semantic countability and the lack of inflectional form(s) for non-plural values may lead to defectiveness (Sect. 5); or there may be special strategies for using pluralia tantum nouns in environments where the number of items is distinguished (notably in constructions with numerals).
 - For non-count nouns, the defectiveness issue does not arise except in instances of recategorization.
 - Establishing that the form(s) available to the noun are indeed plural forms involves comparing with normal count nouns, and considering agreement evidence.
- **syntactic criterion:** In syntactic terms, having only the plural implies that only plural agreements are possible.
- **relation of the criteria:**
 - the inflectional and the syntactic criteria may be aligned (*scissors* Sect. 2.1)
 - a noun may be a plurale tantum just in terms of inflection, but not in terms of agreement (Tsez *xexbi* ‘child(ren)’ Sect. 2.2)
 - a noun may be a plurale tantum just in terms of agreement, but not in terms of inflection. This may apply to nouns which do not inflect for number (whether this is regular or exceptional), but where other count nouns control different number values in agreement (Russian *galife* ‘riding breeches’ Sect. 2.3, Walman *apar* ‘bed’ Sect. 4.5.1, Bayso *ilkoo* ‘tooth/teeth’ Sect. 6.4).

Thus what it means to be a plurale tantum noun is seen most clearly by calibrating from the baseline of a canonical noun with canonical number properties.

8.7 The canonical plurale tantum noun

In most work within Canonical Typology, it proves logical to anchor one end of a criterion but not the other (rather like the temperature scale anchored to zero Kelvin but without an upper limit). However, there are phenomena where both ends of a

scale have natural end points. Suppletion is one such phenomenon: according to the criteria of maximally regular semantic correlation and maximally irregular formal correlation, we can calibrate from absolutely regular inflection at one end to full suppletion at the other (Corbett 2007a). We should consider pluralia tantum nouns in this way. At one end of the scale we have nouns that are fully canonical for number, and at the other end—arguably—we could situate canonically pluralia tantum nouns.

Intuitively a canonical pluralia tantum noun needs to be lexically specified as such (it would not be part of a medium size generalization for instance). Moreover, in the canonical world, criteria line up, so we would set the limit at nouns whose semantics, syntax and morphology are aligned; that is to say, a noun which has only the plural in semantic, syntactic and morphological terms. (This is a “worse” scenario than those discussed in Sect. 2.) Such a noun would be defective: it would be specified as semantically plural, leaving a gap where the singular would be expected.

Such an item has been proposed (Baerman 2009). Surprisingly it is from German, a language with few pluralia tantum nouns (examples include *Masern* ‘measles’ and *Spesen* ‘expenses’), despite its areal position near the pluralia tantum hotspot of the Circum-Baltic region (Sect. 5). Yet German has the remarkable *Eltern* ‘parents’. This noun has only plural forms, it takes only plural agreements, and it must denote more than one parent. It can be used whenever more than one parent is intended and where there is no specific reference to sex. It does not have a corresponding singular.⁴⁶

8.8 Summary of noun types

We now survey the main noun types discussed, particularly from the perspectives of exhaustiveness (Sect. 8.1) and alignment (Sect. 8.2), see Table 4:

Table 4 Summary: number possibilities of the main types

	Semantic	Syntactic	Morphological	Notes
book (Sect. 2)	ALL	ALL	ALL	Canonical for number
xexbi (Sect. 2.2)	ALL	ALL	ONE	
sheep (Sect. 5)	ALL	ALL	ONE	Uninflecting
pal'to (Sect. 2.3)	ALL	ALL	ONE	Uninflecting
buti (Sect. 6.2)	ALL	ALL	REDUCED	
cê: (Sect. 6.2)	ALL	ALL	REDUCED	
ikamrak (Sect. 6.3)	ALL	REDUCED	REDUCED	
ilkoo (Sect. 6.4)	ALL	ONE	REDUCED	
scissors (Sect. 2.1)	ALL	ONE	ONE	
galife (Sect. 2.3)	ALL	ONE	ONE	Uninflecting
sani (Sect. 2.4)	ALL	ONE	ONE	
apar (Sect. 4.5.1)	ALL	ONE	ONE	Uninflecting
wheat (Sect. 5)	ONE	ONE	ONE	
oats (Sect. 5)	ONE	ONE	ONE	
Eltern (Sect. 8.7)	ONE	ONE	ONE	Canonical plurale tantum

⁴⁶As Baerman points out, there is instead the compound *Elternteil* ‘parental unit’, a more formal term. This has its own plural *Elternteile* (which can mean ‘a pair of parents’, ‘more than one pair of parents’, or ‘more than one parent’). The lack of a singular for *Eltern* ‘parents’ is particularly felt in scientific usage, and here the back-formation *Elter* is attested (besides this technical use it may also be used jocularly, Sebastian Fedden, personal communication).

Book has all number possibilities, for semantics, syntax and morphology. There are increasing restrictions through to the ‘worst’ case, German *Eltern* ‘parents’, which would be expected to be a count noun but which fully lacks the singular. Bold marks items that can be labelled pluralia tantum. REDUCED indicates that not all the potential values are available, but still there is more than one. In the diversity, we note that there is a general trend for greater restriction in syntax (agreement) than in semantics, and greater in morphology than in syntax.

9 Pluralia tantum constructions

The criteria that prove useful for defining pluralia tantum nouns have extensions beyond the simple word. There are set expressions involving pluralia tantum: *give someone the creeps*. Then there are phrases which appear only in the plural, even though the head noun is not a plurale tantum: *United States* (but not *state*), *Olympic Games* (but not *game*). Spanish has an interesting set of these:

- (67) Spanish
- | | | | | | |
|----------------|-----------|--------------------------|-----------------|--------------|-------------|
| buen-os | día-s | buen-as | tarde-s | buen-as | noche-s |
| good-M.PL | day(M)-PL | good-F.PL | afternoon(F)-PL | good-F.PL | night(F)-PL |
| ‘good morning’ | | ‘good afternoon/evening’ | | ‘good night’ | |

While the greeting appears to be pragmatically comparable to that in other languages, including other Romance languages, Spanish uses the plural in this set of expressions.⁴⁷

Moving to pluralia tantum compounds, Russian has *figli-migli* ‘tricks’, in which both parts inflect, but only in the plural (Zaliznjak 1977:283); there are several examples in the Russian National Corpus, with plural attributive modifiers. Recall too the Tsez compounds (plural because they consist of two items) in footnote 7.

We now consider a surprising plurale tantum construction in Russian (Sect. 9.1), and some general uses which are restricted to the plural (affective uses Sect. 9.2).

9.1 The Russian ‘including’ construction *idti v letčiki*

Russian has a construction which is noteworthy for several reasons, including the fact that the nominal element must be plural, even when this lacks motivation. The expression *idti v letčiki* means ‘become a pilot’ (literally ‘go into the pilots’). There are various verbs which can be the first element, and even some nouns. The last slot can be occupied by any animate noun, but typically it is one denoting a profession or social grouping. What is constant is the preposition *v* ‘into’ and the fact that the final noun is plural:

⁴⁷It is an ongoing mystery. For some speculations see the Language Log: <http://languagelog.ldc.upenn.edu/nll/?p=4589>. Yet Turkish also uses the plural for greetings (Matthew Baerman, personal communication), and even English has *best wishes*.

- (68) Russian ‘including’ construction
 kandidat v prezident-y
 candidate into president-PL
 ‘candidate for the presidency’

The point is that the election is for one president, the noun *prezident* ‘president’ has a perfectly good, frequent singular, but in this construction the plural is required (its case form is problematic, see Corbett 2012:210–213). Thus we have a productive, common construction, which involves a plurale tantum slot.⁴⁸ Similarly in French we find expressions like *jouer les héros* ‘play the hero’ (literally ‘the heroes’).

9.2 Affective uses

Number is used in ways beyond its basic function. It is commonly used for honorific purposes, and here all values may be involved, showing different degrees of respect (Corbett 2000:220–228). More relevant for us are various affective uses, such as the exaggerative and the intensificative (Corbett 2000:234–239 provides examples from a range of languages). Here is one from Russian:

- (69) Russian exaggerative (Krasil’nikova 1990:85)
 Vy tam piš-ete na nemeck-ix jazyk-ax
 2PL there write-2PL on German-PL.LOC language-PL.LOC
 ‘You’re there writing in German.’

Here the affective use is clear, since there is no straightforward interpretation of the plural prepositional phrase. Here is an intensificative use from Slovene:

- (70) Slovene intensificative (Janez Orešnik, personal communication)
 Kdo krad-e denarnic-e?
 Who steal-PRS.3SG purse-PL.ACC
 ‘Who’s been nicking my purse?’

Recall that Slovene has a dual, so we can ask whether the dual is possible here (perhaps to indicate less intensity than the plural). In fact *denarici* ‘(two) purses’ can only have the literal meaning.

We can therefore see the same patterning for these uses of number values as we saw for nouns in larger systems in Sect. 6. For honorific use, all values may in principle be available. While for affective use, on the evidence to date, only the plural can be used.

10 Conclusion

The increased interest in pluralia tantum nouns is welcome. However, the topic is broader than most linguists have realized. *Scissors*, Tsez *xexbi* ‘child(ren)’, Russian

⁴⁸The noun *gosti* ‘guests’ has developed from its use in this construction into an event-denoting plurale tantum, as documented by Zaliznjak (2009).

galife ‘riding breeches’, Walman *apar* ‘bed’ and Bayso *ilkoo* ‘tooth/teeth’ are all pluralia tantum, but they are different. We have analysed their differences within a canonical typology, taking in the whole range of limited number, including restrictions within larger systems (even going beyond nouns).

We have seen that confusion over definitions has arisen in part because of the tacit assumption that a noun’s paradigm cells are consistent, in the sense that their “internal” morphosemantic specification and their “external” morphosyntactic specification are identical. While this identity holds in the default case, and indeed in the canonical world, there are important instances involving pluralia tantum nouns where it does not. These demonstrate that we must expand Stump’s content paradigm, in order to meet both sets of requirements. The concern to be fully explicit about the different types of pluralia tantum nouns has led us to this theoretical advance. We also saw that what is required for certain pluralia tantum nouns also holds for the more general, systematic instances in Miya and Bayso. Thus careful typology leads to theoretical advance, and equally formal theory makes explicit the ramifications of our typology.

The value of a specifically canonical approach to typology has also been demonstrated. The paper contributes to that body of research in analysing another phenomenon where we can calibrate from a fully canonical instance of a noun with all number possibilities, through to the “worst” case of a fully canonical plurale tantum noun.

Acknowledgements The support of the AHRC (grant: AH/N006887/1: Lexical splits: a novel perspective on the structure of words) is gratefully acknowledged. Versions of this paper were read at the Linguistics Seminar, University of Queensland, 14 November 2016; the workshop ‘Morphological Typology and Linguistic Cognition’, Linguistics Institute, University of Kentucky, 23 July 2017; the workshop ‘Unlocking new data: Linguistic typology as a key to agreement variation’, University of Zürich, 9 October 2017; CRISSP, KU Leuven, 18 January 2018; the Diversity Linguistics Seminar, University of Leipzig, 25 January 2018; Centre de Linguistique en Sorbonne (CeLiSo), Paris, 23 March 2018; Heinrich Heine University, Düsseldorf, 19 April 2018; 16th International Congress of Slavists, Belgrade, 24 August 2018. I am grateful to those present for helpful suggestions, and to the class at the European Summer School in Linguistic Typology, Porquerolles, 17 September 2016. For sharing their expertise in key languages I thank Lea Brown and Matthew Dryer (Walman), Marina Chumakina (Archi), Bernard Comrie and Maria Polinsky (Tsez), Idan Landau (Modern Hebrew), Ora Matushansky (Russian), Osahito Miyaoka (Central Alaskan Yup’ik), Ljubomir Popović (Serbo-Croat), Gerald Stone (Upper Sorbian), Greg Stump (Sanskrit), Anna Thornton (Italian) and Hannu Tammola (Finnish). And for valuable comments on different versions of the paper, I am thankful to Jenny Audring, Matthew Baerman, Oliver Bond, Patricia Cabredo Hofherr, Matthew Carroll, Marina Chumakina, Bernard Comrie, Sebastian Fedden, Tim Feist, Raphael Finkel, Borja Herce, Nahyun Kwon, Michele Loporcaro, Silva Nurmio, Jeff Parker, Jérémy Pasquereau, Kristian Roncero, Erich Round, Helen Sims-Williams, Anna Thornton and Nigel Vincent. I am grateful to two anonymous referees for constructive suggestions. Thanks too, to Penny Everson and Lisa Mack, for their help in preparing the typescript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Acquaviva, P. (2008). *Lexical plurals*. Oxford: Oxford University Press.

- Acquaviva, P. (2014). Number. In M. Aronoff (Ed.), *Oxford bibliographies in linguistics*. New York: Oxford University Press.
- Acquaviva, P. (2016a). Structures for plurals. *Linguisticae Investigationes*, 39(2), 217–233.
- Acquaviva, P. (2016b). Linguistic plurality and the conceptualization of part structure. In M. Carrara, A. Arapinis, & F. Moltmann (Eds.), *Unity and plurality: Logic, philosophy, and linguistics* (pp. 194–218). Oxford: Oxford University Press.
- Acquaviva, P. (2017). Number in language. In *Oxford research encyclopedia of linguistics*. <http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-61>.
- Acuña-Fariña, C. (2016). (A few) psycholinguistic properties of the NP. *Functions of Language*, 23, 120–141.
- Alexiadou, A. (2011). Plural mass nouns and the morpho-syntax of number. In M. Byram Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 33–41). Somerville: Cascadilla Proceedings Project.
- Allan, K. (1980). Nouns and countability. *Language*, 56, 541–567.
- Andersen, T. (2014). Number in language. In A. Storch & G. J. Dimmendaal (Eds.), *Number—constructions and semantics: Case studies from Africa, Amazonia, India and Oceania* (pp. 221–263). Amsterdam: John Benjamins.
- Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, 12(142), 5–32.
- Arregi, K., & Nevins, A. (2014). A monoradical approach to some cases of disuppletion. *Theoretical Linguistics*, 40, 311–330.
- Audring, J. (2017). Calibrating complexity: How complex is a gender system? *Language Sciences*, 60, 53–68.
- Baayen, R. H., Piepenbrock, R., & Guilikers, L. (1995). *The CELEX lexical database (CD-ROM) (version v2.5)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 37, 94–117.
- Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355–390). Berlin: de Gruyter.
- Baerman, M. (2009). In M. Baerman, D. Brown, & G. G. Corbett (Eds.), *Surrey cross-linguistic database on defectiveness*. University of Surrey. German. <https://doi.org/10.15126/SMG.21/2>.
- Baerman, M., & Corbett, G. G. (2010). Introduction: defectiveness: typology and diachrony. In M. Baerman, G. G. Corbett, & D. Brown (Eds.), *Defective paradigms: missing forms and what they tell us* (pp. 1–18). Oxford: British Academy and Oxford University Press.
- Baerman, M., Corbett, G. G., Brown, D., & Hippisley, A. (2006). *Surrey typological database on dependency*. Surrey: University of Surrey. <https://doi.org/10.15126/SMG.15/1>.
- Baerman, M., Brown, D., & Corbett, G. G. (2017). *Morphological complexity*. Cambridge: Cambridge University Press.
- Bale, A., & Barner, D. (2011). Mass-count distinction. In M. Aronoff (Ed.), *Oxford bibliographies online* (pp. 1–26). Oxford: Oxford University Press.
- Bayrılı, İ.B. (2017). *The Universality of Concord*. PhD dissertation, MIT.
- Belić, A. (1932). *O dvojini u slovenskim jezicima*. Belgrade: Srpska kraljevska Akademija.
- Beyersmann, E., Dutton, E. M., Amer, S., Schiller, N. O., & Biedermann, B. (2015). The production of singular- and plural-dominant nouns in Dutch. *Language, Cognition and Neuroscience*, 30, 867–876.
- Bjorkman, B. M. (2016). Go get, come see: Motion verbs, morphological restrictions, and syncretism. *Natural Language and Linguistic Theory*, 34, 53–91.
- Boas, F. (1893). Notes on the Chinook language. *American Anthropologist*, 6, 55–64.
- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, 43, 83–128.
- Borer, H. (2005). *Structuring sense: I: In name only*. Oxford: Oxford University Press.
- Börstell, C., Lepic, R., & Belsizman, G. (2016). Articulatory plurality is a property of lexical plurals in sign language. *Linguisticae Investigationes*, 19, 391–407.
- Braun, M. (1930). *Das Kollektivum und das Plurale tantum im Russischen: ein bedeutungsgeschichtlicher Versuch*. Inaugural-Dissertation. Frommhold & Wendler, Leipzig.
- Brown, D., Corbett, G. G., Fedden, S., Hippisley, A., & Marriott, P. (2013). Grammatical typology and frequency analysis: Number availability and number use. *Journal of Language Modelling*, 1, 227–241.

- Bugariski, R. (2012). Language, identity and borders in the former Serbo-Croatian area. *Journal of Multilingual and Multicultural Development*, 33, 219–235.
- Bye, P. (2015). The nature of allomorphy and exceptionality: Evidence from Burushaski plurals. In E. Bonet, M.-R. Lloret, & J. Mascará (Eds.), *Understanding allomorphy: Perspectives from optimality theory* (pp. 107–176). London: Equinox.
- Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthese*, 174, 99–149.
- Comrie, B. (1992). Balto-Slavonic. In J. Gvozdanović (Ed.), *Indo-European numerals* (pp. 717–833). Berlin: Mouton de Gruyter.
- 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.
- Comrie, B., & Polinsky, M. (1998). The great Daghestanian case hoax. In A. Siewierska & J. Jung Song (Eds.), *Case, typology and grammar: in honor of Barry J. Blake* (pp. 95–114). Amsterdam: John Benjamins.
- Comrie, B., Polinsky, M., & Rajabov, R. (1998). *Tsezian languages*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Unpublished.
- Corbett, G. G. (1983). *Hierarchies, targets and controllers: Agreement patterns in Slavic*. London: Croom Helm.
- Corbett, G. G. (2000). *Number*. 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.
- Corbett, G. G. (2007b). Deponency, syncretism and what lies between. In M. Baerman, G. G. Corbett, D. Brown, & A. Hippisley (Eds.), *Deponency and morphological mismatches* (pp. 21–43). Oxford: British Academy and Oxford University Press.
- Corbett, G. G. (2012). *Features*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2013). Canonical morphosyntactic features. In D. Brown, M. Chumakina, & G. G. Corbett (Eds.), *Canonical morphology and syntax* (pp. 48–65). Oxford: Oxford University Press.
- Corbett, G. G. (2015a). Morphosyntactic complexity: a typology of lexical splits. *Language*, 91, 145–193.
- 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.
- Corbett, G. G. (2018). *Pluralia tantum nouns in the Slavonic languages*. <http://epubs.surrey.ac.uk/848705/> [Earlier version of part of the current paper, with a case study on ‘violin’ in Slavonic languages.]
- 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.
- Corbett, G. G., & Fedden, S. (2016). Canonical gender. *Journal of Linguistics*, 52, 495–531.
- Corbett, G. G., Fedden, S., & Finkel, R. (2017). Single versus concurrent feature systems: nominal classification in Mian. *Linguistic Typology*, 21, 209–260.
- Corbett, G. G., & Fraser, N. M. (2000). Gender assignment: a typology and a model. In G. Senft (Ed.), *Systems of nominal classification* (pp. 293–325). Cambridge: Cambridge University Press.
- Corbett, G. G., & Hayward, R. J. (1987). Gender and number in Bayso. *Lingua*, 73, 1–28.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Daniel, M. (2005). Plurality in independent personal pronouns. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 146–149). Oxford: Oxford University Press. [Online edition, edited by Matthew S. Dryer & Martin Haspelmath, at: <http://wals.info/>.]
- Davis, M. H., van Casteren, M., & Marslen-Wilson, W. D. (2003). Frequency effects in processing inflected Dutch nouns: A distributed connectionist account. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 427–462). Berlin: de Gruyter.
- de Reuse, W. J. (2000). La partie et le tout: les extensions métonymiques du duel et du pluriel en yupik sibérien central. In N. Tersis & M. Therrien (Eds.), *Les langues eskaléoutes: Sibérie, Alaska, Canada, Groënland* (pp. 269–282). Paris: CNRS.
- Degtjarev, V. I. (1982). Proisxoždenie imen pluralia tantum v slavjanskix jazykax (K opredeleniju semantičeskix mexanizmov leksikalizacii form mn. č.). *Voprosy jazykoznanija*, 1, 65–77.
- Dimmendaal, G. J. (1983). *The Turkana language*. Dordrecht: Foris.
- Diver, W. (1987) The dual. *Columbia University Working Papers in Linguistics*, 8, 100–114.
- Dobrušina, N. R., & Panteleeva, S. A. (2008). Sobiratel’nye čislitel’nye: kollektiv kak individualizacija množestvennosti. In A. Mustajoki, M. V. Kopotev, L. A. Birjulín, & E. Ju. Protasova (Eds.), *Instrumentarij rusistiki: korpusnye podxody* (pp. 107–124). Helsinki: University of Helsinki.

- Doetjes, J. (2012). Count/mass distinctions across languages. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. III, pp. 2559–2580). Berlin: de Gruyter.
- Dryer, M. S., & Brown, L. (2015). *Gender in Walman*. Paper at the workshop “Grammatical Gender and Linguistic Complexity”, Stockholm University, 20–21 November 2015. https://www.ling.su.se/polopoly_fs/1.266857.1453731810/menu/standard/file/DryerBrown_WalmanGenderStockholm.pdf. To appear as: Matthew S. Dryer, Gender in Walman, in the conference volume.
- Enger, H.-O. (2009). The role of core and non-core semantic rules in gender assignment. *Lingua*, 119, 1281–1299.
- Evans, N., Bergqvist, H., & San Roque, L. (2018). The grammar of engagement II: Typology and diachrony. *Language and Cognition*, 10, 141–170. <https://doi.org/10.1017/langcog.2017.22>.
- Farmer, S. (2015). *Establishing reference in Máihkì*. PhD Dissertation, University of California, Berkeley.
- Faßke, H. (1981). *Grammatik der obersorbischen Schriftsprache der Gegenwart: Morphologie*. Bautzen: VEB Domowina Verlag.
- Fedden, S., & Corbett, G. G. (2017). Gender and classifiers in concurrent systems refining the typology of nominal classification. *Glossa: a journal of general linguistics*, 2(1), 34.1–47.
- Fedden, S., & Corbett, G. G. (2018). Extreme classification. *Cognitive Linguistics*, 29(4), 633–675.
- Fedden, S., Audring, J., & Corbett, G. G. (Eds.) (2018). *Non-canonical gender systems*. Oxford: Oxford University Press.
- Forker, D. (2016a). Conceptualization in current approaches of language typology. *Acta Linguistica Hafniensia*, 48, 70–84.
- Forker, D. (2016b). Gender in Hinuq and other Nakh-Daghestanian languages. *International Journal of Language and Culture*, 3, 90–114.
- Gardelle, L. (2016). Lexical plurals for aggregates of discrete entities in English: Why plural, yet non-count, nouns? *Linguistic Investigations*, 39, 355–372.
- Goddard, C. (2009). A piece of cheese, a grain of sand: the semantics of mass nouns and unitizers. In F. J. Pelletier (Ed.), *Kinds, things, and stuff: Mass terms and generics* (pp. 132–165). Oxford: Oxford University Press.
- Greenberg, J. H. (1966). *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.
- Grimm, S. (2018). Grammatical number and the scale of individuation. *Language*, 94, 527–574.
- Hakimov, N. (2016). Explaining variation in plural marking of German noun insertions. In H. Behrens & S. Pfänder (Eds.), *Experience counts: Frequency effects in language* (pp. 21–60). Berlin: de Gruyter.
- Harbour, D. (2007). *Morphosemantic number: From Kiowa noun classes to UG number features*. Dordrecht: Springer.
- Harbour, D. (2011). Valence and atomic number. *Linguistic Inquiry*, 42, 561–594.
- Haskell, T. R., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: some implications of compounds research. *Cognitive Psychology*, 47, 119–163.
- Haspelmath, M. (1993). *A grammar of Lezgian*. Berlin: de Gruyter Mouton.
- Haspelmath, M. (2005). Occurrence of nominal plurality. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 142–145). Oxford: Oxford University Press. [Online edition, edited by Matthew S. Dryer & Martin Haspelmath, at: <http://wals.info>.]
- Haspelmath, M., & Karjus, A. (2017). Explaining asymmetries in number marking: singulatives, pluratives and usage frequency. *Linguistics*, 55, 1213–1235.
- Hayward, R. J. (1979). Bayso revisited: some preliminary linguistic observations—II. *Bulletin of the School of Oriental and African Studies, University of London*, 42, 101–132.
- Hayward, D. [= R. J. Hayward] (1984). *The Arbore language: A first investigation: including a vocabulary*. Hamburg: Buske.
- Isačenko, A. V. (1962). *Die russische Sprache der Gegenwart: I: Formenlehre*. Halle: Max Niemeyer.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41, 9–45.
- Jacobson, S. A. (1995). *A practical grammar of the central Alaskan Yup'ik Eskimo language* (with Yup'ik readings written by Anna W. Jacobson). Fairbanks: Alaska Native Language Center, University of Alaska.
- Janaš, P. (1976/1984). *Niedersorbische grammatik*. Bautzen: Domowina. Page references to the revised 1984 edition.
- Janko, T. E. (2002). Russkie čislitel'nye kak klassifikatory čislitel'nyx. *Russkij jazyk v naučnom osveščeníi*, 1(3), 168–181.
- Jespersen, O. (1954). *A modern English grammar on historical principles: II: Syntax* (Vol. 1). London: George Allen and Unwin.

- Juul, A. (1975). *On Concord of number in modern English*. Copenhagen: Nova.
- Karlssohn, F. (2000). Defectivity. In G. Booij, C. Lehmann, & J. Mugdan (Eds.), *Morphologie (morphology): Vol. 1. Ein Internationales Handbuch Zur Flexion und Wortbildung (an international handbook on inflection and word-formation)* (pp. 647–654). Berlin: Mouton de Gruyter.
- Karlssohn, G. (1968). *Eras suomen inkongruenssi-ilmio. Fenno-Ugrica: Juhlakirja Lauri Postin kuusikymmenvuotispäiväksi* (Vol. 17(3), pp. 117–126). Helsinki: Suomalais-ugrilainen seura.
- Kibrik, A. E. (1977). *Opyt strukturnogo opisanija arčinskogo jazyka: II: Taksonomičeskaja grammatika*. Moscow: Izdatel'stvo Moskovskogo universiteta.
- Kibrik, A. (2003). Nominal inflection galore: Daghestanian, with side glances at Europe and the world. In F. Plank (Ed.), *Noun phrase structure in the languages of Europe* (pp. 37–112). Berlin: Mouton de Gruyter.
- Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. 1, pp. 131–175). Dordrecht: Foris.
- Kiss, T., Pelletier, F. J., Husic, H., Poppek, J. M., & Simunic, R. N. (2016). A sense-based lexicon of count and mass expressions: The Bochum English countability lexicon. In N. Calzolari et al. (Eds.), *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia. Paris: European Language Resources Association (ELRA).
- Klockmann, H. (2017). The design of semi-lexicity: Evidence from case and agreement in the nominal domain. *LOT Dissertation Series*. <http://www.lotpublications.nl/the-design-of-semi-lexicity>.
- Koenig, J.-P. (1999). *Lexical relations*. Stanford: CSLI.
- Koptjevskaja-Tamm, M., & Wälchli, B. (2001). The circum-Baltic languages: An areal-typological approach. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic languages: Vol. II. Grammar and typology* (pp. 615–761). Amsterdam: Benjamins.
- Kostić, D. (1987). *Operativna gramatika srpskohrvatskog jezika*. Beograd: Prosveta.
- Kramer, R. (2016). A split analysis of plurality: number in Amharic. *Linguistic Inquiry*, 47, 527–559.
- Krasil'nikova, E. V. (1990). *Imja suščestvitel'noe v russkoj razgovornoj reči: Funkcional'nyj aspekt*. Moscow: Nauka.
- Kulikov, L. (2004). Review of number. *Southwest Journal of Linguistics*, 23, 124–129.
- Kwon, N. (2017). Total reduplication in Japanese ideophones: An exercise in localized canonical typology. *Glossa: a journal of general linguistics*, 2(1), 40.1–40.31. <https://doi.org/10.5334/gjgl.267>.
- Kwon, N., & Round, E. R. (2015). Phonaesthemes in morphological theory. *Morphology*, 25, 1–27. <https://doi.org/10.1007/s11525-014-9250-z>.
- Lammert, M. (2015). Les pluralia tantum sous l'angle du collectif. *Langue française*, 185, 73–84.
- Lammert, M. (2016). Lexical plurals through meronymy and hyperonymy. *Lingvisticae Investigationes*, 39, 335–354.
- Landau, I. (2016). DP-internal semantic agreement: A configurational analysis. *Natural Language and Linguistic Theory*, 34, 975–1020.
- Landman, F. (2011). Count nouns—mass nouns, neat nouns—mess nouns. *Baltic International Yearbook of Cognition, Logic and Communication*, 6, 1–67. <https://doi.org/10.4148/biyclc.v6i0.1579>.
- Landman, F. (2016). Iceberg semantics for count nouns and mass nouns: classifiers, measures and portions. *Baltic International Yearbook of Cognition, Logic and Communication*, 11 1–48. <https://doi.org/10.4148/1944-3676.1107>.
- Lanman, C. R. (1872). A statistical account of noun-inflection in the Veda. *Journal of the American Oriental Society*, 10, 325–601. <https://doi.org/10.2307/592131>.
- Lasnik, P. (2011). Mass nouns and plurals. In K. von Stechow, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. II, pp. 1131–1153). Berlin: de Gruyter.
- Lauwers, P. (2014). Les pluriels «lexicaux». Typologie quantifiée des déficits de dénombrabilité. *Langue française*, 183, 117–132.
- Lauwers, P. (2016). Les pluriels lexicaux dits «massifs» face au conditionneur universel. *Lingvisticae Investigationes*, 39, 272–288.
- Lauwers, P., & Lammert, M. (2016). New perspectives on lexical plurals. *Lingvisticae Investigationes*, 39, 207–216.
- Leko, N. (2009). *The syntax of numerals in Bosnian*. Munich: Lincom Europa.
- Lichtenberk, F. (1983). *A grammar of Manam*. Honolulu: University of Hawaii Press.
- Lima, S. (2014). *The grammar of individuation and counting*. Ph.D. dissertation, University of Massachusetts, Amherst. https://scholarworks.umass.edu/dissertations_2/109.

- Lin, J., Hachohen, A., & Schaeffer, J. (2018). The interpretation of the mass-count distinction across languages and populations: Introduction. *Glossa: A Journal of General Linguistics*, 3(1), 70. <https://doi.org/10.5334/gjgl.638>.
- Link, G. (1983). The logical analysis of plurals and mass terms: a lattice-theoretic approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (302–323). Berlin: Walter de Gruyter.
- Ljaševskaja, O. N. (2004). *Semantika ruskogo čisla*. Moscow: Jazyki Slavjanskoj Kul'tury.
- Löfstedt, B. (1958). Zum Gebrauch der lateinischen distributiven Zahlwörter. *Eranos*, 56, 71–117, 188–223.
- Loughnane, R. (2009). *A grammar of Oksapmin*. PhD thesis, University of Melbourne.
- Lučić, R. (2015). Observations on collective numerals in standard Croatian. *Journal of Slavic Linguistics*, 23, 3–31.
- MacDonell, A. A. (1927). *A Sanskrit grammar for students* (3rd ed.). London: Oxford University Press.
- Maruyama, Y. (2011). K voprosu o raspredenii form dvojtvennogo čisla v ruskix žitijnyx pamjatnikax načala XV v. (na material “Žitija Stefana Permskogo” i “Žitija Sergeja Radonežskogo”). *Ruskij jazyk v naučnom osveščeenii*, 1(21), 162–188.
- Maruyama, Y. (2014). Upotreblenie dual'nyx form v žitijnyx tekstax staroruskogo perioda: leksičeskij, grammatičeskij i pragmatičeskij faktory. In M. Nomachi, A. Danylenko, & P. Piper (Eds.), *Grammaticalization and lexicalization in the Slavic languages: Proceedings of the 36th meeting of the commission on the grammatical structure of the Slavic languages of the international committee of Slavists* (pp. 51–62). Munich: Otto Sagner.
- Massam, D. (Ed.) (2012). *Count and mass across languages*. Oxford: Oxford University Press.
- Matthews, P. H. (1997). *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.
- McGill, S. (2007). The Cicipu noun class system. *Journal of West African Languages*, 34(2), 51–90.
- McGill, S. J. (2009). *Gender and person agreement in Cicipu discourse*. PhD thesis, SOAS, University of London.
- Mel'čuk, I. A. (1979). “Countability” vs “non-countability” and lexicographic description of nouns in Russian. In P. R. Clyne, William. F. Hanks, & C. L. Hofbauer (Eds.), *CLS 15: Papers from the fifteenth regional meeting Chicago Linguistic Society* (pp. 220–227). Chicago: Chicago Linguistic Society.
- Mel'čuk, I. A. (1985). *Poverxnostnyj sintaksis ruskix čislovyx vyraženij*. Vienna: Institut für Slawistik der Universität Wien.
- Menzel, T. (2018). Kanonische morphologische komplexität im sorbischen. *Lětopis*, 65(2), 81–106.
- Merrifield, W. R. (1959). Classification of Kiowa nouns. *International Journal of American Linguistics*, 25, 269–271.
- Mithun, M. (2010). The search for regularity in irregularity: defectiveness and its implications for our knowledge of words. In M. Baerman, G. G. Corbett, & D. Brown (Eds.), *Defective paradigms: missing forms and what they tell us* (pp. 125–149). Oxford: British Academy and Oxford University Press.
- Miyaoka, O. (2012). *A grammar of Central Alaskan Yupik (CAY)*. Berlin: De Gruyter Mouton.
- Moodie, J. (2016). Number marking in Lopit, an Eastern Nilotic language. In D. L. Payne, S. Pacchiarotti, & M. Bosire (Eds.), *Diversity in African languages: Selected papers from the 46th annual conference on African linguistics* (pp. 397–416). Berlin: Language Science Press.
- Moorhouse, A. C. (1966). F. W. Saas: *Pluralia Tantum*. Bijdrage tot de kennis van het gebruik van de indoeuropese numeri, in het bijzonder in het Grieks. Pp. xii+188. Assen: Van Gorcum, 1965. Paper, fl. 18.50. *The Classical Review*, 16(3), 414–415. <https://doi.org/10.1017/S0009840X00322673>.
- Moravcsik, E. A. (1991). Review article on Anna Wierzbicka ‘the semantics of grammar’. *Studies in Language*, 15, 129–148.
- Moravcsik, E. A. (2017). Number. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *The Cambridge handbook of linguistic typology* (pp. 440–476). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316135716.014>.
- Moszyński, L. (1985). Czy w języku staro-cerkiewno-słowiańskim była kategoria *duale tantum*. *Slavica Hierosolymitana: Slavic Studies of the Hebrew University*, 7, 107–118.
- Nononen, M., & Niemi, J. (2010). Mismatches between grammatical number and conceptual numerosity: a number-decision experiment on collective nouns, number neutralization, pluralia tantum, and idiomatic plurals. *Folia Linguistica*, 44, 103–125.
- Nickels, L., Biedermann, B., Fieder, N., & Schiller, N. O. (2015). The lexical-syntactic representation of number. *Language, Cognition and Neuroscience*, 30(3), 287–304. <https://doi.org/10.1080/23273798.2013.879191>.

- Nicolas, D. (2016). The logic of mass expressions. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (winter 2016 ed.). <https://plato.stanford.edu/archives/win2016/entries/logic-massexpress/>.
- Nikolaeva, I. (2013). Unpacking finiteness. In D. Brown, M. Chumakina, & G. G. Corbett (Eds.), *Canonical morphology and syntax* (pp. 99–122). Oxford: Oxford University Press.
- Nikunlassi, A. O. (2000). The use of collective numerals in contemporary Russian: an empirical approach. *Wiener slawistischer Almanach*, 45, 209–246. http://periodika.digitale-sammlungen.de/wsa/Blatt_bsb00000513.00209.html.
- Nikunlassi, A. O. (2006). On the use of collective numerals with inanimate nouns in Russian (norm and usage). *Scando-Slavica*, 52, 5–17. <https://doi.org/10.1080/00806760600883735>.
- Noyer, R. R. (1992). *Features, Positions and Affixes in Autonomous Morphological Structure*. PhD Dissertation, MIT. [Published, 1997, New York: Garland].
- Nurmio, S. (2017). Collective nouns in Welsh: a noun category or a plural allomorph? *Transactions of the Philological Society*, 115, 58–78.
- Ojeda, A. E. (1997). A semantics for the counting numerals of Latin. *Journal of Semantics*, 14, 143–171.
- Ojeda, A. E. (2001). The plurality of plurals. In K. Vermeulen & A. Copestake (Eds.), *Algebras, diagrams and decisions in language, logic and computation* (pp. 37–65). Stanford: CSLI.
- Ojeda, A. E. (2005). The paradox of mass plurals. In S. S. Mufwene, E. J. Francis, & R. S. Wheeler (Eds.), *Polymorphous linguistics: Jim McCawley's legacy* (pp. 389–410). Cambridge: MIT Press.
- Palmer, F. R. (1990). The semantics of grammar. Review article on Anna Wierzbicka 'the semantics of grammar'. *Journal of Linguistics*, 26, 223–233.
- Panevová, J., & Ševčíková, M. (2011). Jak se počítají substantiva v češtině: poznámky ke kategorii čísla. *Slovo a slovesnost*, 72, 163–176.
- Payne, J., & Huddleston, R. (2002). Nouns and noun phrases. In R. Huddleston & G. K. Pullum (Eds.), *The Cambridge grammar of the English language* (pp. 323–523). Cambridge: Cambridge University Press.
- Pelletier, J. F. (2011). Descriptive metaphysics, natural language metaphysics, Sapir-Whorf, and all that stuff: evidence from the mass-count distinction. *The Baltic international yearbook of cognition, logic and communication: Vol. 6: Formal semantics and pragmatics: discourse, context, and models*, (pp. 1–46). <https://doi.org/10.4148/byiclc.v6i0.1570>.
- Pelletier, F. J. (2012). Lexical nouns are both +mass and +count, but they are neither +mass nor +count. In D. Massam (Ed.), *Count and mass across languages* (pp. 9–26). Oxford: Oxford University Press.
- Pesetsky, D. (2013). *Russian case morphology and the syntactic categories* (Linguistic Inquiry Monograph 66). Cambridge: MIT Press.
- Pesetsky, D., & Torrego, E. (2007). The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian, & W. K. Wilkins (Eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation* (pp. 262–294). Amsterdam: Benjamins.
- Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53, 181–216.
- Plaster, K., Polinsky, M., & Harizanov, B. (2013). Noun classes grow on trees: Noun classification in the North-East Caucasus. In B. Bickel, L. A. Grenoble, D. A. Peterson, & A. Timberlake (Eds.), *Language typology and historical contingency: In honor of Johanna Nichols* (pp. 153–170). Amsterdam: Benjamins.
- Polinsky, M. (2015). *Tsez syntax: A description*. lingbuzz/002315.
- Polinsky, M., & Comrie, B. (1999). Agreement in Tsez. In G. G. Corbett (Ed.), *Agreement (special issue of Folia Linguistica XXXIII/2)* (pp. 109–130). Berlin: Mouton de Gruyter.
- Polivanova, A. K. (1983). Vybór číslonyx form suščestvitel'nogo v russkom jazyke. In V. P. Grigor'ev (Ed.), *Problemy strukturnoj lingvistiki* (pp. 130–145). Moscow: Nauka.
- Popović, L. (1979). Upotreba kardinalnih brojeva u srpskohrvatskom jeziku. *Jugoslavenski seminar za strane slaviste*, 30, 3–25. Belgrade: Filološki fakultet u Beogradu: Međunarodni slavistički centar.
- 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: Crane Publishing Co.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Räsänen, S. (1979). Huomioita suomen sijojen frekvensseistä [Observations on the frequencies of the Finnish cases]. *Sananjalka*, 21, 17–34.

- Rjabuškina, S. V. (2015). Russkie sobirateľ'nye čislitel'nye kak markery oduševlennosti. *Vestnik Nižegorodskogo universiteta im. N. I. Lobačevskogo*, 2(2), 527–531.
- Rombandeeva, E. I. (1973). *Mansijskij (vogul'skij) jazyk*. Moscow: Nauka.
- Rothstein, S. (2017). *Semantics for counting and measuring*. Cambridge: Cambridge University Press.
- Round, E. R., & Corbett, G. G. (2017). The theory of feature systems: One feature versus two for Kayardild tense-aspect-mood. *Morphology*, 27, 21–75.
- Saas, F. W. (1965). *Pluralia tantum: Bijdrage tot de kennis van het gebruik van de indoeuropese numeri, in het bijzonder in het Grieks*. Assen: Van Gorcum.
- Sasse, H.-J. (1993). Syntactic categories and subcategories. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung/An International handbook of contemporary research* (Vol. 1, pp. 646–686). Berlin: de Gruyter.
- Savà, G. (2011). Endangered Bayso (Cushitic): Interesting typological and historical aspects. In L. Busetto, R. Sottile, L. Tonelli, & M. Tosco (Eds.), *He bitaney lagge: Studies on language and African linguistics in honour of Marcello Lamberti* (pp. 163–174). Milan: Qu.A.S.A.R.
- Schmidtko, D., & Kuperman, V. (2017). Mass counts in world Englishes: A corpus linguistic study of noun countability in non-native varieties of English. *Corpus Linguistics and Linguistic Theory*, 13, 135–164.
- Schuh, R. G. (1989). Number and gender in Miya. In Z. Frajzyngier (Ed.), *Current progress in Chadic linguistics: Proceedings of the international symposium on Chadic linguistics*, Boulder, Colorado, 1–2 May 1987 (pp. 171–181). Amsterdam: Benjamins.
- Schuh, R. G. (1998). *A grammar of Miya*. Berkeley: University of California Press.
- Short, D. (1993). Czech. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic languages* (pp. 455–532). London: Routledge.
- Smith, P. W. (2016). Lexical plurals in Telugu: Mass nouns in disguise. *Linguisticae Investigationes*, 39, 234–252.
- Smith, P. W. (2017). The syntax of semantic agreement in English. *Journal of Linguistics*, 53, 823–863.
- Smith-Stark, T. C. (1974). The plurality split. In M. W. La Galy, R. A. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting*, Chicago Linguistic Society, April 19–21, 1974 (pp. 657–671). Chicago: Chicago Linguistic Society.
- Soboleva, P. A. (1984). Číslo suščestvitel'nyx kak grammatiko-slovoobrazovatel'naja kategorija. In N. Ju. Švedova (Ed.), *Russkij jazyk: funkcionirovanie grammatičeskix kategorij: tekst i kontekst: Vinogradovskie čtenija XII–XIII* (pp. 66–75). Moscow: Nauka.
- Spencer, A. (2005). Towards a typology of 'mixed categories'. In C. Orhan Orgun & P. Sells (Eds.), *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe* (pp. 95–138). Stanford: CSLI.
- Spencer, A. (2013). *Lexical relatedness: a paradigm-based model*. Oxford: Oxford University Press.
- Stefanović, A. (2014). Kvantifikacija imenica pluralia tantum. *Sarajevski filološki susreti: zbornik radova*, 2(1), 45–67. <https://www.ceeol.com/search/article-detail?id=594946>.
- Stevanović, M. (1975). *Savremeni srpskohrvatski jezik (gramatički sistemi i književnojezička norma): I: Uvod, fonetika, morfologija* (3rd ed.). Belgrade: Naučna knjiga.
- Stump, G. T. (2001). *Inflectional morphology: a theory of paradigm structure*. Cambridge: Cambridge University Press.
- Stump, G. T. (2007). A non-canonical pattern of deponency and its implications. In M. Baerman, G. G. Corbett, D. Brown, & A. Hippisley (Eds.), *Deponency and morphological mismatches* (pp. 71–95). Oxford: British Academy and Oxford University Press.
- Stump, G. (2015). The interface of semantic interpretation and inflectional realization. In L. Bauer, L. Körtvélyessy, & P. Štekauer (Eds.), *Semantics of complex words* (pp. 27–45). Cham: Springer.
- Stump, G. (2016). *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.
- Stump, G. (2017). The nature and dimensions of complexity in morphology. *Annual Review of Linguistics*, 3, 65–83. <https://doi.org/10.1146/annurev-linguistics-011415-040752>.
- Suprun, A. E. (1959). *O russkix čislitel'nyx*. Frunze: Kirghiz State University.
- Suprun, A. E. (1969). *Slavjanskie čislitel'nye (stanovlenie čislitel'nyx kak oboj časti reči)*. Minsk: Belorussian State University.
- Sutton, L. (2010). Noun class and number in Kiowa-Tanoan: Comparative-historical research and respecting speakers' rights in fieldwork. In A. L. Berez, J. Mulder, & D. Rosenblum (Eds.), *Fieldwork and linguistic analysis in indigenous languages of the Americas*. <http://hdl.handle.net/10125/4451>.
- Sutton, P. R., & Filip, H. (2017). Individuation, reliability, and the mass/count distinction. *Journal of Language Modelling*, 5, 303–356.

- Terrill, A. (2003). *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- Tiersma, P. M. (1982). Local and general markedness. *Language*, 58, 832–849.
- Timberlake, A. (2004). *A reference grammar of Russian*. Cambridge: Cambridge University Press.
- Tommola, H. (2018). Plurativy v jazykax areala Baltijskogo morja. In L. L. Šestakova, N. V. Patroeva, & O. V. Nikitin (Eds.), *Fortunatovskie čtenija v Karelii: Sbornik dokladov meždunarodnoj naučnoj konferencii (10–12 sentjabrja 2018 goda, g. Petrozavodsk)* (Vol. I, pp. 219–222). Petrozavodsk: PetrGU.
- Trask, R. L. (1997). *A student's dictionary of language and linguistics*. London: Arnold.
- Völkel, P. (1981). *Hornjoserbsko-němski słownik: Obersorbisch-deutsches Wörterbuch: prawopisny słownik hornjoserbskeje rěče* (4th ed.). Bautzen: Ludowe nakładnistwo Domowina.
- Vraciu, A. (1976). Considérations sur l'emploi des noms pluralia tantum dans les langues baltiques, slaves et finno-ougriennes. *Acta Baltico-Slavica*, 9, 27–37.
- Wackernagel, J. (1920). *Vorlesungen über Syntax: mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch*. Basel: Emil Birkhäuser.
- Wälchli, B. (2011). The circum-Baltic languages. In B. Kortmann & J. van der Auwera (Eds.), *The languages and linguistics of Europe: A comprehensive guide* (pp. 325–340). Berlin: De Gruyter Mouton.
- Watkins, L. J. (1984). *A grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Watkins, L. J. (1995). *Noun classification in Kiowa-Tanoan*. Paper presented at the meeting of the Society for the Study of the Indigenous Languages of the Americas, Albuquerque, University of New Mexico, July 8–9, 1995. Unpublished.
- 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.
- Wierzbicka, A. (1988). *The semantics of grammar*. Amsterdam: Benjamins.
- Wierzbicka, A. (1991a). The semantics of grammar: a reply to Professor Palmer. *Journal of Linguistics*, 27, 495–498.
- Wierzbicka, A. (1991b). Semantic rules know no exceptions. *Studies in Language*, 15, 371–398.
- Williams, E. (1994). Remarks on lexical knowledge. *Lingua*, 92, 7–34.
- Wisniewski, E. J. (2009). On using count nouns, mass nouns, and pluralia tantum: what counts? In F. J. Pelletier (Ed.), *Kinds, things, and stuff: Mass terms and generics* (pp. 166–190). Oxford: Oxford University Press.
- Wonderly, W., Gibson, L. F., & Kirk, P. L. (1954). Number in Kiowa: nouns, demonstratives, and adjectives. *International Journal of American Linguistics*, 20, 1–7.
- Zaliznjak, A. A. (1967/2002). *Russkoe imennoe slovoizmenenie*. Moscow: Nauka. Reprinted in: Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremenomu russkomu jazyku i obščemu jazykoznaniju*, 1–370. Moscow: Jazyki slavjanskoj kul'tury.] Page references to 2002 edition.
- Zaliznjak, A. A. (1977). *Grammatičeskij slovar' russkogo jazyka: slovoizmenenie*. Moscow: Russkij jazyk.
- Zaliznjak, Anna A. (2009). *Gosti pluralia tantum v sovremennom russkom jazyke*. In T. Berger, M. Giger, S. Kurt, & I. Mendoza (Eds.), *Von grammatischen Kategorien und sprachlichen Weltbildern – Die Slavia von der Sprachgeschichte bis zur Politsprache: Festschrift für Daniel Weiss zum 60. Geburtstag* (pp. 633–642). Munich: Otto Sagner.
- Ziegeler, D. (2010). Count-mass coercion, and the perspective of time and variation. *Constructions and Frames*, 2, 33–73.
- Žolobov, O. F., & Kryš'ko, V. B. (2001). *Dvojtvennoe čislo*. Moscow: Azbukovnik.