

# MORPHOSYNTACTIC COMPLEXITY: A TYPOLOGY OF LEXICAL SPLITS

GREVILLE G. CORBETT

*University of Surrey*

A key notion in understanding language is ‘possible word (lexeme)’. While there are lexemes that are internally homogeneous and externally consistent, we find others with splits in their internal structure (morphological paradigm) and inconsistencies in their external behavior (syntactic requirements). I first explore the characteristics of the most straightforward lexemes, in order to establish a point in the theoretical space from which we can calibrate the real examples we find. I then schematize the interesting phenomena that deviate from this idealization, including suppletion, syncretism, deponency, and defectiveness. Next I analyze the different ways in which lexemes are ‘split’ by such phenomena. I set out a typology of possible splits, along four dimensions: splits that are (i) based on the composition/feature signature of the paradigm versus those based solely on morphological form; (ii) motivated (following a boundary motivated from outside the paradigm) versus purely morphology-internal (‘morphomic’); (iii) regular (extending across the lexicon) versus irregular (lexically specified); (iv) externally relevant versus irrelevant: we expect splits to be internal to the lexeme, but some have external relevance (they require different syntactic behaviors).

I identify instances of these four dimensions separately: they are orthogonal, and therefore not dependent on each other. Their interaction gives a substantial typology, and it proves to be surprisingly complete: the possibilities specified are all attested. The typology also allows for the unexpected patterns of behavior to overlap in particular lexemes, producing some remarkable examples. Such examples show that the notion ‘possible word’ is more challenging than many linguists have realized.\*

*Keywords:* inflectional morphology, possible lexeme, canonical typology, split paradigm, morpheme, allomorphy

**1. INTRODUCTION.** Consider the forms in Table 1 of the verb *vurtinava* ‘turn’ in Sepečides Romani, spoken by a Balkan Romani group, now settled in Izmir, Turkey. The data are cited in Adamou 2012:156 from Cech & Heinschink 1999:47.<sup>1</sup>

\* The support of the ERC (grant ERC-2008-AdG-230268 *MORPHOLOGY*), the AHRC (grant AH/I027193/1 ‘From competing theories to fieldwork: The challenge of an extreme agreement system’, and grant AH/K003194/1 ‘Combining gender and classifiers in natural language’), and the ESRC and AHRC jointly (grant ES/I029621/1 ‘Endangered complexity: Inflectional classes in Oto-Manguean languages’) is gratefully acknowledged. The analyses presented here have been well aired at different meetings, and I thank all those present who helped to clarify the typology: University of Bremen, September 2009; Vrije Universiteit Amsterdam, November 2009; Matica Srpska, Novi Sad, September 2010; 29th International Conference on Lexis and Grammar, Belgrade City Library, September 2010; the Société de Linguistique de Paris, December 2010; Department of Linguistics, UC Santa Barbara, February 2011; Department of Slavic Languages and Literatures, UC Berkeley, February 2011; 6th Meeting of the Slavic Linguistics Society, Université de Provence, Aix-en-Provence, September 2011; Societas Linguistica Europaea, Universidad de la Rioja, September 2011; 8th Mediterranean Morphology Meeting, Cagliari, September 2011; Philipps-Universität Marburg, October 2011; Hokkaido University, Slavic Research Centre, Sapporo, November 2011; University of the West of England, February 2012; HSE Moscow, May 2013; CIESAS Sureste, July 2013; Association for Linguistic Typology, Leipzig, August 2013. For helpful comments on previous versions I am particularly grateful to Stephen Anderson, Matthew Baerman, Olivier Bonami, Dunstan Brown, Wayles Browne, Patricia Cabredo-Hofherr, Marina Chumakina, Scott Collier, Nick Evans, Sebastian Fedden, Tim Feist, Alice Harris, Andrew Hippisley, Alexander Krasovitsky, Enrique Palancar, Gilles Polian, Erich Round, Greg Stump, Anna Thornton, and Claire Turner, to Greg Carlson (editor), Stanley Dubinsky (executive editor), and Adam Albright (associate editor), and to Lisa Mack and Penny Everson for help in preparing the manuscript.

<sup>1</sup> There is a mistake in the second-person plural in the original source; I have corrected it here, thanks to Petra Cech (p.c.).

	SINGULAR	PLURAL
1	vurt-um	vurtin-d-am
2	vurt-un	vurtin-d-en
3	vurt-u	vurtin-d-e

TABLE 1. Sepečides Romani: past tense of *vurtinava* 'turn'.

This is originally a Slavic verb. In Sepečides Romani, in the past tense, its plural inflections are those that would be expected of a Romani verb. In the singular, however, it inflects mainly as a Turkish verb. This verb's behavior is split, between singular and plural, according to the language where the inflections originate. This example is certainly surprising: we do not expect the paradigm of an individual lexeme to take material from different languages. When we compare crosslinguistically, however, we see that it is just one of many split paradigms. Consider this more familiar example in Table 2.

	SINGULAR	PLURAL
1	vais	allons
2	vas	allez
3	va	vont

TABLE 2. French: present tense of *aller* 'go'.

Here we have a lexeme whose parts are phonologically distinct. If such examples of suppletion were not familiar, we would be shocked by them. As John Lyons rightly said, 'All languages are exotic languages'.

Now look at the typical Russian verb in Table 3.

		SINGULAR	PLURAL	
PRESENT	1	govorj-u	govor-im	< person and number
	2	govor-iš'	govor-ite	
	3	govor-it	govorj-at	
PAST	M	govoril		< gender and number
	F	govoril-a	govoril-i	
	N	govoril-o		

TABLE 3. Russian *govorit'* 'speak'.

When we look carefully, Russian verbs as illustrated in Table 3 are in some ways 'worse' than the two previous examples. There is a deep split in the paradigm—not just in the forms, but rather in the features to which the two segments are sensitive. In the present tense (more generally the nonpast) the verb marks person and number, while in the past tense it marks number and gender.

**1.1. OUTLINE.** The variety in the examples we have seen raises the question of just how different lexemes can be: in other words, the question of what is a possible lexeme. We therefore consider below a typology of inflectional phenomena (§2). However, we shall be more interested in the splits themselves than in the phenomena that induce them.<sup>2</sup> Indeed, some current work on allomorphy misses the point that splits induced by allomorphy are matched by splits induced by suppletion, defectiveness, and periphrasis.

<sup>2</sup> At its simplest, a split is simply a division of a paradigm into parts (segments). We are concerned with those splits that are more surprising and challenging. The work relates readily to Stump's analysis of heteroclisism (2006). It also links to Ackerman and colleagues' approach (2009); however, while one of their central claims is that within segments there is interpredictability of forms (they argue against their being derived from a single base form), that issue is orthogonal to the present article. I aim to show how and where paradigms can be split, and the effects of such splits, and so I leave aside the relations of the cells within the segments.

In investigating the wide range of differences across languages, the perspective of CANONICAL TYPOLOGY proves useful, allowing us to establish a typology based on four criteria for analyzing splits. I give clear instances for each of the four criteria, and the two possibilities for each, and this is the core of the article (§3). Furthermore, since these criteria are in principle independent of each other, they imply a typological space with sixteen possibilities (§4). I investigate which of these are actually found, thus adopting what Bond (2013) terms ‘exploratory canonical typology’. The results are given in the appendix. It turns out that each type, even the most unlikely, is attested. The range of possible lexemes proves to be surprisingly large.

**1.2. NOTES FOR THE READER.** Earlier presentations of this research have shown that some linguists find the material disturbing. This highlights the need to be clear about the assumptions brought to the data—by reader and writer. I assume that in many instances we can distinguish lexical meaning, which can be quite idiosyncratic, from grammatical meaning. The latter is more predictable since it is available across a range of lexical items. Grammatical meanings are orthogonal not only to lexical items but also to each other; hence we model them with features. Thus in the French example in Table 2, the distinctions (values) of the person feature are available to verbs with very different lexical semantics, and the person values are also cross-cut by number. Given a featural analysis, we can draw up a grid of the combination of a lexical item and all of its featural possibilities. This notional grid is often called a paradigm.

Now if the mapping from the grid of possible feature descriptions to the set of actual forms is straightforward, we are done. However, a brief glance at the grammar of Archi or Zulu (and many languages in between) will convince us that we are not yet done. The easy picture implied by our featural system is often not what we find. For instance, a reasonable account of the features of person, number, and tense in French would not lead us to expect the difference in the forms of *aller* ‘go’ in Table 2.

Given nontrivial mappings, as in the instances we have seen so far, some extra step is needed. This is equally true whatever model is adopted. Many linguists look first for a phonological solution. Once the synchronic phonological rules have been found insufficient, these linguists may be willing to propose additional phonological rules that recapitulate historical change, to postulate zeros to facilitate derivations, to add new constraints, and so on. A different approach is to consider whether the featural description is inadequate, and to modify it by devices such as impoverishment. A third line is to suggest that the different mappings correspond to slightly different syntactic structures, which would justify different featural specifications. And finally we may look for semantic differences (possibly invoking markedness) to account for the different mappings. Indeed, some would appeal to any of these possibilities, or any combination of them, rather than allow for an intraparadigmatic solution. That is, for some linguists the paradigm is no more than an inventory of the grammatical forms that can be associated with a given lexical item.

Those who wish to explain away the data presented, assuming that any phonological, syntactic, or semantic explanation is inherently preferable to a morphological one, are thereby acknowledging the importance of the data. Even our terminology hints at the interesting issues: we do not have a term for lexical items whose forms are fully predictable, but we do have a term for the surprising behavior of French *aller* ‘go’, as in Table 2, namely ‘suppletion’. If all of the cells in our notional grid are filled with different forms, this situation is not name-worthy; if, however, there are identities between forms, we flag the potential mapping problem by terming it ‘syncretism’ (see Baerman

et al. 2005 for a crosslinguistic analysis, and Opitz et al. 2013 for a neurophysiological investigation).

By contrast, morphologists of the inferential-realizational persuasion (those adopting word-and-paradigm models) are persuaded of the reality of structuring within paradigms by evidence from phenomena such as suppletion (Corbett 2007a, Hinzeln 2011, Aronoff 2012), directional syncretism (Corbett & Fraser 1997, Evans et al. 2001, Stump 2001: 223–30, Bobaljik 2002:80–81, Baerman et al. 2005:130–45, 163–66, 175–77, Müller 2011, Xu & Aronoff 2011, Albright & Fuß 2012:267–71, Brown & Hippisley 2012: 167–80),<sup>3</sup> the patterns of predictability based on principal parts (Stump & Finkel 2013), the existence of canonical inflectional classes with completely distinct sets of affixes (Corbett 2009 and §3.2 below), the fact that the semantic interpretation of an inflected form may be sensitive to the morphosyntactic characterization of other cells in the paradigm (Stump 2009), or from the persistence over centuries of morphomic patterns (Maiden 2005, 2013).<sup>4</sup> For these morphologists, there is still the issue of determining which of the problematic mappings are a matter of morphology and which are not (see Baerman 2014 for an instance of complex interactions of phonological, morphophonological, and morphological conditions). However, given that in this view there are clear instances of intraparadigmatic structuring (and hence paradigms are an essential component of the theoretical apparatus, rather than just a handy means of presenting and discussing data), there is no imperative to explain away difficult mappings at any cost. Rather, these are accepted as the stuff, and indeed the interest, of inflectional morphology.<sup>5</sup>

Whatever our theoretical persuasion, there are interesting phenomena we have to account for (including those that are already name-worthy, such as syncretism). There is legitimate disagreement about how we treat them, which makes it useful to have a coherent way of talking about them. Such a means of identification helps to ensure that the value of analyses is not lost when formalisms change. Thus, providing a metarepresentation of paradigms, where competing analyses can meet, is useful intellectual housekeeping. Its value can be seen simply by recalling what happens when we tackle any interesting inflectional system: part is easy, and then there are particular paradigmatic configurations that require time and effort to give an elegant analysis. If our assumptions remain covert, we can miss the common properties of those configurations that we all work hard over. Simple affixation is not a problem; it is rather phenomena such as difficult syncretisms, apparently unmotivated allomorphy, and paradigmatic gaps that take our attention. As noted above, there are commonalities in the patterns of splits, across very different morphological phenomena, that tend to be missed by those primarily focused on phonological form.

<sup>3</sup> A key example in the discussion is Table 6 below.

<sup>4</sup> Luís (2011) finds evidence for autonomous morphology even in inflectionally impoverished languages, specifically Lusophone Creoles, and suggests (p. 236) that: ‘A consistent and increasing body of evidence thus supports the insight, formulated by Matthews, that some aspects of “the structure of the word form must be supplied by statements of a wholly morphological nature”’, citing Matthews 1972:107. The substantial formal differences in the resources available to morphology in comparison to syntax are taken up by Asudeh and colleagues (2013).

<sup>5</sup> A hallmark of these approaches in their modern variants is that they are implemented, and thus proposed analyses can be verified (that is, the linguistic forms predicted by an analysis can be checked for accuracy). For NETWORK MORPHOLOGY, the language is DATR (Evans & Gazdar 1996; see the implemented fragments in Corbett & Fraser 1993, Brown & Hippisley 2012, Baerman 2014, among others); for PARADIGM FUNCTION MORPHOLOGY (Stump 2001) there is a set of programs called Cat’s Claw ([http://www.rch.uky.edu/project\\_claw.html](http://www.rch.uky.edu/project_claw.html)).

It is important to recognize that we are not dealing with rarities: there are many instances where the mapping from featural specification to form is not straightforward. And any account will require some extra step in the analysis. What that step should be can be decided either by prior conviction, or by surveying the full range of comparative data and attempting analyses that address this rich variety, basing ourselves on ‘observable linguistic behaviour’ (Nikolaeva 2013:104). The debate has become strained, and so in the next section I offer a canonical view of the phenomena being discussed, since ‘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).

**2. A CANONICAL VIEW OF LEXEMES.** To highlight the nature and interest of splits, we begin by establishing the characteristics of the most straightforward lexemes. We define what it means for a lexeme to be canonical in terms of its inflectional morphology. This allows us to fix a point in the theoretical space, so that we can then calibrate the real examples we find. By defining canonical inflection, we can give a coherent schema for the interesting phenomena that deviate from this idealization. These include suppletion, syncretism, deponency, and defectiveness. We then move on to the different ways in which lexemes are ‘split’ by these phenomena.

**2.1. CANONICAL TYPOLOGY: ESSENTIAL IDEAS.** This approach to typology involves analyzing and defining phenomena that are subject to variability (across and within languages), extracting the various scales along which we characterize variability, and establishing the logical endpoint of these scales. By integrating these scales, we construct theoretical spaces of possibilities, and only then do we investigate how this space is populated with real instances. To be a canonical instance—that is, clear and indisputable—means matching a full set of criteria, and so it follows that such instances are likely to be at best infrequent or even nonexistent (in accord with the ‘Anna Karenina principle’; Diamond 1998:157). This is therefore an axiomatic approach, which aims to ensure that we are aware of the full range of the phenomena we wish to account for, and have a metalanguage to describe them. It is justified entirely by utility and results, which have been encouraging to date. A more extended account of the method and several recent examples can be found in Brown et al. 2013.

This approach has been worked out for syntax, specifically for the tricky area of agreement by Comrie 2003, Corbett 2003, Evans 2003, Polinsky 2003, and is applied to Miraña in Seifart 2005:156–74 and to Gujarati in Suthar 2006:178–98; the analysis of agreement is taken further in Corbett 2006, where twenty criteria define the theoretical space. Work on other areas of syntax includes Nikolaeva & Spencer 2008 and Van de Velde 2013. Work in inflectional morphology is presented in Corbett 1999, 2007a, 2009, Baerman et al. 2005:27–35, Stump 2006, 2007, 2013, Spencer 2007, Thornton 2011a, Paciaroni 2012, and Palancar 2012.<sup>6</sup> For a canonical approach to phonology, specifically to tone, see Hyman 2009, 2011, and for derivational morphology and semantics see Fortin 2011. A working bibliography of research in canonical typology is available on the Surrey Morphology Group website.<sup>7</sup>

**2.2. CANONICAL LEXEMES: REQUIREMENTS OF FORM.** This article represents an application of the canonical approach to inflectional morphology; it goes beyond pure inflection in tackling also the syntactic relevance of the splits investigated. It is important

<sup>6</sup> A computational approach to inflection, within a canonical framework, can be found in Walther 2013.

<sup>7</sup> <http://www.surrey.ac.uk/englishandlanguages/research/smg/canonicaltypology/bibliography/index.htm>

to be explicit about the assumptions we are making, in order to be able to concentrate on the key issues involving splits. We assume that for the systems we are analyzing we have already established the features and their values. This is not a trivial issue; see Zaliznjak 1973, van Helden 1993, Meyer 1994, Corbett 2008, 2012. With this in mind, I have chosen, where possible, examples where the feature inventory is not a contentious issue (for instance, there is little argument about French having three person and two number values, as in Table 2). Where the decisions are difficult, I point this out. Features model regularities, but they may have a ‘penumbra’ of unclarity around the central system (Corbett 2011). Similarly, there are deep questions as to what comprises a lexeme and how we decide whether particular forms constitute a single lexeme. Taking for instance the lexeme *go*, linguists might disagree as to how many of the following are included in it: *going*, *gone*, *went*, *is going*, and *has been going*. Again, I have avoided this question wherever possible by selecting examples that are generally recognized, and rightly in my view, as being single lexemes. Finally, the stem-affix distinction may be obvious, and where possible I take examples where this is the case. In some instances the division is highly problematic in terms of segmentation or of the distribution of lexical and grammatical information over stem and affix (see Baerman & Corbett 2012 and Spencer 2012 for discussion).

If the features and their values are established, these should ‘multiply out’, so that all possible cells in a paradigm exist. For example, if a given language has four cases and three numbers in its nominal system, the paradigm of a noun should have twelve cells. (This is equivalent to Spencer’s notion of ‘exhaustivity’ (2003:252).) Furthermore, to be fully canonical, a paradigm should be ‘consistent’, according to the requirements in Table 4 (the levels of comparison are described below).

	COMPARISON ACROSS CELLS OF A LEXEME (level one comparison)	COMPARISON ACROSS LEXEMES (level two comparison)
1. LEXICAL MATERIAL (≈ shape of STEM)	same	different
2. INFLECTIONAL MATERIAL (≈ shape of AFFIX)	different	same
OUTCOME (≈ shape of inflected word)	different	different

TABLE 4. Canonical inflection: requirements of form.

In the canonical situation there is a transparent mapping between meaning and form. This mapping is consistent in the expression of lexical meaning through lexical means, and grammatical meaning through inflectional means.

The schema in Table 4 offers two levels of comparison.

- **LEVEL ONE:** We start from the abstract paradigm that results from taking the features and values and multiplying them out. We then examine the inflected forms of any one lexeme in terms of this paradigm. Thus the center column of Table 4 compares cell with cell, within the paradigm of a single lexeme. We take the criteria in the left column in turn:
  - (i) in terms of the LEXICAL MATERIAL in the cell, we require identity (the stem should remain the same);
  - (ii) conversely, the INFLECTIONAL MATERIAL ‘should’ be different in every cell.

The outcome for such a canonical lexeme is given in the bottom row: every cell in the paradigm of a canonical lexeme will realize the morphosyntactic specification in a way distinct from that of every other cell.

- LEVEL TWO: this involves comparing lexemes with lexemes within the given language (right column). We use the same criteria as before:
  - (i) we require that the LEXICAL INFORMATION be realized differently (since these are different lexemes);
  - (ii) in the canonical situation, the INFLECTIONAL MATERIAL is identical. That is, if the genitive plural is realized by *-g-a* on one lexeme, every other lexeme realizes this feature specification in the same way.

The outcome is that every form of every lexeme is distinct. Such a system of canonical inflection would seem to make sense in functional terms. There is absolute differentiation within the morphology, while using minimal phonological material. The system would be consistent with the simplest item-and-arrangement model of morphology, a system in which morphology looks like syntax. I have characterized canonical inflection by considering cells and their relations, within and across lexemes. The same result can be achieved by recasting the requirements as properties of the system; this is the approach adopted by Thornton (2011b) and Stump (2012).

Of course, we find substantial divergences from the idealization presented in Table 4. Its value is as a standard from which we can calibrate the phenomena we actually find. This schema allows for various possible deviations from canonicity, and they are in fact all attested. We shall consider briefly only some of these types of possible lexeme (for a fuller account and further illustration see Corbett 2007b); this is because we are concerned with the ‘results’ of these noncanonical phenomena, the splits within paradigms. We look at the deviations from canonicity again at two levels: first internally, comparing the cells of a single lexeme (Table 5), then externally, comparing across lexemes (Table 7). A general pattern is that where we actually find ‘same’ in place of canonical ‘different’, this will give a nonfunctional outcome, since the function–form mapping will be less transparent. If we find ‘different’ in place of canonical ‘same’, this will lead to increased complexity and/or redundancy.

	CANONICAL BEHAVIOR	DEVIANT BEHAVIOR	TYPES OF DEVIATION
1. LEXICAL MATERIAL (≈ shape of STEM)	same	different	alternations suppletion
2. INFLECTIONAL MATERIAL (≈ shape of AFFIX)	different	same	syncretism uninflectability

TABLE 5. Deviations in form established by comparison across the CELLS of a lexeme.

We start with LEXICAL MATERIAL. In canonical situations, we find identity of lexical material across the cells of a given lexeme. However, we may find various types of alternations deviating from this ideal. The least canonical situation is that in which the lexical material is completely different, which is what we find in suppletion. We saw a French instance in Table 2, and another example of suppletion is given in Table 6 from the South Slavic language Slovene.

	SINGULAR	DUAL	PLURAL
NOMINATIVE	človek	človeka	ljudje
ACCUSATIVE	človeka	človeka	ljudi
GENITIVE	človeka	ljudi	ljudi
DATIVE	človeku	človekoma	ljudem
INSTRUMENTAL	človekom	človekoma	ljudmi
LOCATIVE	človeku	ljudeh	ljudeh

TABLE 6. Slovene: suppletion in the noun *človek* ‘man, person’ (Priestly 1993:401).

For discussion of this interesting pattern of stems, see Corbett 2007a:30 and references there.<sup>8</sup>

Within the comparison across the cells of a single lexeme, the other item to examine is INFLECTIONAL MATERIAL. In the canonical situation, the inflectional material is different in every cell of the lexeme. The major deviation here is syncretism; we have an expectation of a given number of inflectional forms, while with syncretism two or more of them are identical (two or more morphosyntactic specifications share a single realization). There are several instances in Table 6 (and while they are more general syncretisms, the relevant point here is that they are identifiable by examining a single lexeme).

I now move on to the second level of comparison, namely deviations from canonicity that emerge when we compare lexemes with each other cell by cell. The canonical situation and types of deviation are as in Table 7.

	CANONICAL BEHAVIOR	DEVIANT BEHAVIOR	TYPES OF DEVIATION
1. LEXICAL MATERIAL ( $\approx$ shape of STEM)	different	same	homonymy
2. INFLECTIONAL MATERIAL ( $\approx$ shape of AFFIX)	same	different	inflectional classes heteroclisis deponency

TABLE 7. Deviations in form established by comparison across LEXEMES.

The question of LEXICAL MATERIAL is relatively straightforward. Naturally, we expect lexemes to be phonologically distinct. However, we find noncanonical instances like English *bank* ('land adjoining a river' or 'financial institution'). These are sufficiently interesting to have their own term (homonyms), and to engage those working on parsing strategies. Finally, in this section we turn to deviations in terms of inflectional material. In the canonical situation, if we compare across lexemes, cell by cell, we find the same inflectional material. Here there is a common type of noncanonical behavior (recall that canonical is not the same as frequent or unmarked): we find inflectional classes, where different sets of formatives realize identical morphosyntactic specifications.<sup>9</sup> For instance, some may mark the genitive plural in *-g-a* and others in *-t-u*.

I now present the different types of deviation together for comparison (Table 8).

	CANONICAL SITUATION		TYPES OF DEVIATION	
	CELLS (level one)	LEXEMES (level two)	CELLS (level one)	LEXEMES (level two)
1. LEXICAL MATERIAL ( $\approx$ shape of STEM)	same	different	alternations suppletion	homonymy
2. INFLECTIONAL MATERIAL ( $\approx$ shape of AFFIX)	different	same	syncretism uninflectability	inflectional classes heteroclisis deponency

TABLE 8. Both types of deviation in form (levels one and two comparison).

It is worth pausing to ask why the particular criteria were chosen and, secondarily, why the directionality was given as it was. The basis for inflectional morphology is that we

<sup>8</sup> Against all of the evidence, it has been suggested that suppletion is restricted to functional items. For several counterexamples see Corbett 2007a, and for discussion see Aronoff 2012:41–43, Bonet & Harbour 2012:218, and Haugen & Siddiqi 2013.

<sup>9</sup> For languages where lexemes may belong simultaneously to inflectional classes of different types (for example, based on affixal and prosodic information independently), and for an initial typology of the possible interactions between the classes, see Baerman 2013.



recognize a distinction between lexical meaning, differentiating for instance *crocodile* and *alligator*, and grammatical meaning, differentiating SINGULAR and PLURAL, among others. While lexical meaning can be particular and idiosyncratic, grammatical meaning is more abstract and general. We distinguish them because the two are orthogonal to each other: the lexical meaning difference between *crocodile* and *alligator* is preserved between SINGULAR and PLURAL. Similarly, the difference in meaning between SINGULAR and PLURAL is equivalent for *crocodile* and *alligator*. Moreover, the same grammatical meaning can be realized by forms that are quite different phonologically; hence we use features to model grammatical meaning. The canonical account above is an idealization of these analytical steps. There must be instances like *crocodile* ~ *crocodiles* for us to posit a number distinction in this paradigm. Therefore we take the logical extreme point that all of the cells in a paradigm should be realized differently. For lexical meaning we have the opposite situation; in the ideal case, lexical meaning is constant through the paradigm, and so the extreme is that the lexical material is unchanging. The directionality of these two criteria is dictated by logic: if we were to suggest that the lexical material changes cell by cell within the paradigm but the inflectional material remains the same, we would simply have analyzed the two items incorrectly.<sup>10</sup> When we widen the analysis to cover the full range of lexemes, the two possible extremes for lexical material are that all lexemes would be different or all would be the same. We choose the first option, which accords with a general intuition since exceptions to this have a name (homonyms). Conversely, inflectional material can be the same, cell by cell, across lexemes, and this option gives a simple metric (it accords with perfect agglutination). The scales, then, are already in use (somewhat like the temperature scale); we have pushed them to the logical extreme (rather like fixing the temperature scale at absolute zero). As already noted, this does not imply that canonical instances, those at the extreme, are frequent or normal—in fact, quite the contrary. But it does give us a clear anchor for the theoretical space in which we can locate the real examples we find, including the many intermediary instances. Moreover, Table 8 gives us a schema integrating some less well-defined phenomena of inflectional morphology.

While the different types of noncanonical behavior in terms of form are of considerable interest, in what follows I treat them together, contrasting them with the structural problems, to which I now turn.

**2.3. CANONICAL LEXEMES: REQUIREMENTS OF STRUCTURE.** The phenomena discussed in the last section are becoming better known, and they have accepted terms to describe them. But there are more serious discrepancies in lexical paradigms, which are less frequently discussed, and in some instances lack even an established terminology. From the canonical perspective, we expect inflected forms: (i) to be built in the same way, and (ii) to follow comparable morphosyntactic requirements. That is, exponence (within a lexeme, and across lexemes) ‘should’ be according to the same structural pattern; for instance, if the first-person singular is constructed from a stem and a suffix, it would not be canonical for the second-person singular to consist of prefix plus stem. I call this type of requirement COMPOSITION. And second, forms in the paradigm ‘should’ realize comparable morphosyntactic requirements. If the present tense realizes person and number, in the canonical world the past tense will do so too. I call this type of requirement FEATURE SIGNATURE. The feature signature can be thought of as an abstract feature specification; we may say, for instance, that cells in a given paradigm realize a selection from a fixed

<sup>10</sup> Though Baerman and Corbett (2012) discuss a highly noncanonical single lexeme that comes close to this, when seen against the remaining lexemes of the language.

set of values belonging to a fixed set of features. The feature specifications of the individual cells vary, of course, but within the possibilities defined by the features and their values (the feature signature). These two types of requirement (composition and feature signature) are similar in being more substantial than those restricted to form; they ‘should’ be the same, both within and across lexemes (Table 9).

	COMPARISON ACROSS CELLS OF A LEXEME (level one comparison)	COMPARISON ACROSS LEXEMES (level two comparison)
1. COMPOSITION (≈ means of exponence)	same	same
2. FEATURE SIGNATURE (≈ morphosyntactic requirement)	same	same

TABLE 9. Canonical inflection: requirements of structure.

The two types of requirement are sufficiently different to be worth discussion here (and indeed they are orthogonal to each other, as the discussion of Slovak in §3.1 below shows).<sup>11</sup> Since the notion of comparison in two levels is familiar from §2.2, we go straight to the combined schema in Table 10, where any deviation from Table 9 above naturally occurs when we find different behavior instead of the (canonical) same behavior.

	CANONICAL SITUATION		TYPES OF DEVIATION	
	CELLS (level one)	LEXEMES (level two)	CELLS (level one)	LEXEMES (level two)
1. COMPOSITION (≈ means of exponence)	same	same	affixal inconsistency: e.g. fused exponence periphrasis	affixal inconsistency  antiperiphrasis
2. FEATURE SIGNATURE (≈ morphosyntactic requirement)	same	same	featural inconsistency	featural inconsistency: e.g. defectiveness, overdifferentiation

TABLE 10. Both types of deviation in structure (levels one and two comparison).

We look first at the (morphological) COMPOSITION of the cells. Canonical instances show identity here. For instance, if we find a structure composed of stem plus inflectional affix in one cell of a paradigm, canonicity would require this same composition in all cells of the paradigm, and across lexemes. As before, we compare first across the cells of a lexeme, and then across lexemes.

LEVEL 1. Comparing just within a lexeme, the canonical situation requires what I term AFFIXAL CONSISTENCY: the number and position of the affixes that we find in one cell should be matched in all other cells of the lexeme’s paradigm. However, we may find a single affix in one cell and more than that in another (a type of multiple exponence). Or we might find prefixes, infixes, and suffixes in different cells where the canonical situation would require consistency of position. Consider the number forms

<sup>11</sup> In earlier papers on canonical inflection (such as Corbett 2007b), I treated them as one criterion, since it seemed evident that canonical paradigms should have the same basic structure; after well-founded challenges I realize that there is a good deal to be explored here, and that composition and feature signature represent different aspects of the structural part of canonical paradigms.

of a particular noun in Krongo (a Nilo-Saharan language spoken in the Krongo hills to the south-west of Kartoum, Sudan).<sup>12</sup>

(1) Krongo: the noun *mùsí* ‘sorcerer’ (Reh 1985:117)

SINGULAR	PLURAL
m-ùsí	nú-kù-kk-ùs-óoní
SG-sorcerer	PL-PL-PL-sorcerer-PL
‘sorcerer’	‘sorcerers’

This remarkable noun has a prefix to mark singular, but no less than three prefixes and a suffix marking plural.<sup>13</sup> Each of these affixes is attested elsewhere in the system of number marking; it is the combination that is noteworthy. This noun is splendidly non-canonical in respect to affixal consistency; the number of affixes differs, and the placement too, since its singular is prefixal but its plural involves a suffix as well as prefixes.

One specific phenomenon under the general heading of affixal inconsistency deserves mention; this is FUSED EXPONENCE. Here there is no identifiable affix: the composition found in other cells is not available. The morphosyntactic distinctions are realized but not according to the pattern of the remaining paradigm. For instance, the Russian verb *byt’* ‘be’ has the future *bud-u*, *bud-eš’*, *bud-et* ... ‘I will be, you will be, he/she/it will be ...’; however, the present tense for all persons and both numbers is null. This is the complete form; it is not a zero stem to which affixes can be added. Nor is the verb defective, since it allows the expected configurations with the expected semantic interpretations. (I return to it in the appendix under type 14.)

The means of exponence can be more evidently different in that instead of a single inflected word we may find more than one; this is the phenomenon known as periphrasis. We see several instances in §3.1 below. For now, a simple example would be the Russian future imperfective, which involves the forms just mentioned. While the present is *čitaj-u*, *čitaj-eš’*, *čitaj-et* ‘I read, you read, he/she reads’, the future imperfective is *bud-u čitat’*, *bud-eš’ čitat’*, *bud-et čitat’* ... ‘I will read (will be reading), you will read, he/she will read’. Thus the cells in this segment have two word forms in place of one.<sup>14</sup>

LEVEL 2. We move on to those instances of noncanonicity that can be recognized by comparing across lexemes. There are instances of affixal inconsistency that belong here; that is, the individual lexemes may be internally consistent, but they are not consistent with each other. Consider the verbs from the Daghestanian language Archi given in Table 11 (for simplicity we consider just the perfective forms).

<sup>12</sup> Morphosyntactic glossing follows the Leipzig glossing rules (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>). Note that infixal material is indicated between <>; where there is featural information with no segment corresponding (the information is realized through a form identical to the bare stem), this is indicated between [ ]; inherent, nonovert values (such as gender on nouns) are given between ( ). I have used these abbreviations: 1, 2, 3: first, second, third person; I, II, III, IV: genders I, II, III, IV; ABL: ablative, ABS: absolutive, ACC: accusative, ART: article, AUX: auxiliary, CVB: converb, DAT: dative, DEF: definite, ERG: ergative, EXCL: exclusive, F: feminine, GEN: genitive, IN: ‘in’ localization, INCL: inclusive, INS: instrumental, IPFV: imperfective, M: masculine, MOD: modal (case), N: neuter, NEG: negative, NMLZ: agent nominal, NOM: nominative, OBL: oblique (stem), PFV: perfective, PL: plural, POT: potential, PRS: present, PST: past, REL: relative, SG: singular.

<sup>13</sup> We need look only at the paradigm of this noun to say that it is not canonical. Comparison with other nouns in Krongo would reveal further noncanonicity, since the normal pattern of plural marking is through a single prefix (Reh 1985:97–126). This will be relevant in the appendix, type 11. Numerous instances of affixal inconsistency can be found in verbs in various Muskogean languages: see, for instance, the account of Alabama in Chiu 1987.

<sup>14</sup> The inclusion of particular forms as periphrastic cells within the paradigm can require careful argumentation; for the types of arguments that are relevant, see Brown et al. 2012 and the various papers in Chumakina & Corbett 2013.

GENDER \ NUMBER	<i>aku</i> 'see', perfective stem		<i>aχas</i> 'lie down', perfective stem	
	SINGULAR	PLURAL	SINGULAR	PLURAL
I	w-ak:u	b-ak:u	a<w>χu	a<b>χu
II	d-ak:u		a<r>χu	
III	b-ak:u	ak:u	a<b>χu	aχu
IV	ak:u		aχu	

TABLE 11. Archi: two verbs, perfective (Marina Chumakina, p.c.).

The markers are for agreement with the absolutive argument. Four gender values are distinguished in the singular, while in the plural there is a two-way distinction: genders I/II versus III/IV, which amounts to a distinction between human and nonhuman.<sup>15</sup> Note the interesting syncretisms: genders I and II in the plural have the form of the gender III singular, while genders III and IV in the plural have the form of the gender IV singular. In the part of the paradigm given in Table 11, *aku* 'see' has only prefixal markers, while *aχas* 'lie down' has infixal markers. The correspondence of prefix *d-* with infixal <*r*> is regular. The comparison between these verbs reveals a noncanonical situation, in that the affixal marking across the verbal lexicon is not consistent. Note that the prefixal position is available on the verb *aχas* 'lie down', as we see when the gender-number marker is in competition with the imperfective marker. Thus infixation in Archi is 'frivolous'; that is, it is not phonologically conditioned (Yu 2007:41–42). For a much fuller account, including various additional complexities, see Chumakina & Corbett 2015.<sup>16</sup>

Before leaving the general criterion of composition, we should consider briefly an instance of noncanonical behavior that has been recognized and named, that is, ANTI-PERIPHRAISIS (Haspelmath 2000). This can be recognized only by comparison across lexemes. We saw earlier that the imperfective future is formed in Russian with the auxiliary *budu* (the future of 'be') plus the infinitive.<sup>17</sup> Thus we find (*ja budu sidet'* 'I will be sitting'). This pattern is found for all imperfective verbs except 'be' itself, which has *budu*. The form according to the general pattern would be *\*budu byt'*. The exceptional form, the synthetic form *budu*, is an instance of antiperiphrasis, since the comparison across lexemes suggests we would find periphrasis here, but we instead have an instance (in fact, the only instance in the language) where it does not occur.

<sup>15</sup> I have argued elsewhere (Corbett 2012:239–51) that a person feature is required in the morphosyntax of Archi. That complication need not detain us here, since the person forms are always syncretic with one of the forms analyzed here.

<sup>16</sup> Another interesting example is found in the Eastern Cushitic language Qafar, discussed in Hayward & Orwin 1991. Two of the inflectional classes, which they term 'prefix verbs' and 'suffix verbs', may be contrasted in the perfective, according to the position of the agreement markers. The same is true in the imperfective (and here we find a second difference, in that aspect is marked by ablaut in the first class and by suffixation in the second). Clearly this is not a canonical situation, in that what is prefixal for some verbs is suffixal for others. Two further points should be mentioned. First, the prefixing verbs are not exclusively prefixing; for instance, plurality in the second- and third-persons plural is marked suffixally (Hayward 1998: 633). And second, while the prefixing verbs are in the minority, numbering around 300 (Hayward & Orwin 1991:159, 161), this apparently relict class has been gaining members through borrowing (a key point in Hayward & Orwin 1991). As pointed out by Bliese (1981:156), the class of the verb can be predicted from the basic form: those in an initial vowel other than *a-* are prefixal, while those in an initial consonant or *a-* are suffixal. At this stage of the argument, all that is required is that we have two types of verb, while in the canonical situation we would have one; see further Rucart 2006 and Paster 2009:42, n. 17. For detailed discussion of another language with prefixing and suffixing stems, see Kim's (2010) phonological approach to Huave, and for the rise of different affixation patterns in Papuan languages, see Foley 2000:377.

<sup>17</sup> For an account of how *budu* established itself in this role in the seventeenth and eighteenth centuries, see Swan 2012.

We see further examples involving composition below, but we turn now to the related issue of FEATURE SIGNATURE. In the canonical situation, the featural description of the cells is consistent. That is, if the present segment of the paradigm requires reference to person and number, so does the past segment.

LEVEL 1. Within a single paradigm, we saw a clear instance of noncanonical behavior in terms of feature signature in Table 3; in that example, the Russian verb marks person and number in the present, but gender and number in the past. This is evident from the single verb presented there.<sup>18</sup> This phenomenon has not been discussed as much as some of the less dramatic instances of noncanonical inflectional behavior, but it clearly deserves our attention. We shall see several further instances as we continue our analysis. Of course, one can stipulate that featural requirements are always identical through a paradigm (that is, all forms of the Russian verb have person, number, and gender), and that there is massive systematic syncretism; indeed, in some models one would be forced into that position. However, we are analyzing the direct evidence provided by the morphology. There is no direct evidence for person in the Russian past tense, and this induces a type of split worth including in a typology of possible splits.

LEVEL 2. In the fully canonical situation, the lexemes of a language are the same in terms of their morphosyntactic feature signatures (see Corbett 2013a for discussion of this logical extreme, which proves useful for exploring the differences between the morphosyntactic features). Of course, we find numerous instances where signatures differ, but in many situations this is for the principled reason that parts of speech have different feature signatures. More interesting are the differences that have no such principled reason, that is, where items with the same part-of-speech specification have different feature signatures. We see various examples of this featural inconsistency in what follows. A clear example can be found in Macedonian (a South Slavic language), when we look at two different types of adjective (Table 12).

SINGULAR			PLURAL	GLOSS
MASCULINE	FEMININE	NEUTER		
nov	nova	ново	novi	'new'
kasmetlija			kasmetlii	'lucky'

TABLE 12. Macedonian adjectives (Victor Friedman 1993:266–67 and p.c.).

Adjectives of the majority type, like *nov* 'new', distinguish three genders and two numbers. In contrast, adjectives like *kasmetlija* 'lucky' agree in number but not in gender. Thus the two types have different feature signatures.

While the possibilities here remain to be fully mapped out, there are two parts of the problem of deviant feature signatures that are better known. The first is the problem of defectiveness. We can recognize an item as defective only by comparison: the other lexemes set up the expectation of the cell(s) that should be realized in the system, and the defective item lacks at least one of these.<sup>19</sup> I return to defectiveness in the appendix, type 8. The second is overdifferentiation, the situation where a particular lexeme (or group of lexemes) makes 'too many' distinctions, in comparison with the majority. For our

<sup>18</sup> Recall that we distinguish phenomena that can be demonstrated by reference to a single lexeme from those that require comparison across lexemes. The characteristics of the single verb discussed are in fact general, but the key point here is that for the logic of the argument one lexeme is sufficient to demonstrate the noncanonical behavior. See further Spencer 2013:260–63 on the split in the Russian verb, and Brown & Hippisley 2012:64–68 for a network morphology account.

<sup>19</sup> Hence defectiveness is unusual, and troubling for our notions of productive morphology (Baerman et al. 2010). Not included here are feature specifications that are logically excluded, such as first-person singular inclusive.

example, take the paucal in Bezhta, a Daghestanian language. Nouns in Bezhta have a singular-plural opposition. However, a minority also have a paucal, as in Table 13.

SINGULAR	PAUCAL	PLURAL
sik	sika	sikla

TABLE 13. Bezhta: example of the paucal: *sik* ‘wineskin’ (Xalilov 1985:137).

This is overdifferentiation within the number feature in that, apart from these items, the minority, there would be no justification to posit the value paucal in the number system of Bezhta. According to Madžid Xalilov (p.c.), the use of the paucal, when available, is obligatory for reference to a small number; there is no difference for agreement purposes, however, since the paucal takes the agreements of the plural.

It is worth comparing these two instances of featural inconsistency. In Macedonian (Table 12) a minority of items have one feature value fewer than the majority, while in Bezhta (Table 13) a minority of items have an extra feature value in comparison with the majority; that is, they show overdifferentiation. (Overdifferentiation is thoroughly reviewed in Thornton 2010–2011:438–43.) Neither the Macedonian nor the Bezhta paradigm shows defectiveness, since all lexemes can provide appropriate forms in all circumstances required by the syntax. The difference between Tables 12 and 13 is a statistical one, viewed across the lexicon. In some models the distinction is of little consequence, since all lexical items will in some sense have all the forms, with a small amount of syncretism in the Macedonian type and a great deal in the Bezhta type. In lexical theories, such as LEXICAL-FUNCTIONAL GRAMMAR (LFG), however, lexical items can be somewhat different in structure (as implied in Table 12), provided that unification will go through over the appropriate syntactic domains.

Having looked at examples of both composition and feature signature issues, for the remainder of the article I mainly treat them together, as different requirements of canonical structure.

**3. LEXICAL SPLITS.** What can induce a split in a lexeme’s paradigm? Possible phenomena inducing the split are suppletion, syncretism, and stem alternation. In fact, they include all of the noncanonical elements of inflectional morphology already discussed in §2.2. Indeed, much of the interest of inflectional morphology lies precisely in these phenomena, which produce differences within paradigms. These are splits induced by form. But there are also splits that are induced by differences in structure, that is, in the composition or in the feature signature of cells, as in §2.3. We return to requirements of form versus requirements of structure in §3.1, and then treat all ‘inducers’ together and look at three other cross-cutting criteria for characterizing splits (§§3.2–3.4). In §3.5 we consider briefly an additional type of complexity, relevant just to the fourth criterion (which concerns external relevance of the split).

For each criterion I give examples that are as near the extremes as possible. Since we are dealing with canonical ideal situations, however, the examples may not always reach the ideal, as mentioned above. For instance, the ideal irregular instance would involve a single lexical item; if there are a few examples, these would still be included, since in comparison to a phenomenon that encompasses all possible instances, one that encompasses just a handful is still clearly to be counted as irregular. In this section I present and illustrate the criteria; the underlying logic is discussed in §4.3.

**3.1. FORM VERSUS COMPOSITION/FEATURE SIGNATURE.** Consider again an item whose paradigm is split by a difference in form, which is as ‘bad’ as it can be—that is, an instance of suppletion. Provided the forms fit into the same scheme as those of other

items, we may say that the split involves forms only. We saw instances in French *aller* ‘go’ in Table 2, and Slovene *človek* ‘man, person’ in Table 6. These are the types of non-canonical behavior that conflict with requirements of form (§2.2). More substantial deviations conflict with the structural requirement (§2.3), that of having a consistent composition/feature signature. To review this distinction, consider a typical Slovak verb in Table 14.<sup>20</sup>

		SINGULAR		PLURAL	
PRESENT	1	nesiem		nesieme	
	2	nesieš		nesiete	
	3	nesie		nesú	
		SINGULAR		PLURAL	
PAST	1	M	niesol	som	niesli sme
		F	niesla		
	2	M	niesol	si	niesli ste
		F	niesla		
	3	M	niesol		niesli
		F	niesla		
N		nieslo <sup>a</sup>			

TABLE 14. Slovak *niest* ‘carry’ (Stanislav 1977:114, 199, Silvia Baučeková p.c.).

<sup>a</sup> The neuter is in principle available in all three persons, but naturally occurs most often in the third person.

Slovak verbs have a present tense that is synthetic in composition and that in terms of feature signature marks person and number. There is a major split between present and past tenses. It is not just that the forms are different, but rather that the composition and the feature signatures are different. In the past tense the composition is periphrastic and, in terms of feature signature, gender is marked as well as number and person. That is the main point. Then there is a second split, WITHIN the past: the first and second persons (singular and plural) are periphrastic, taking forms of the auxiliary ‘be’, but not the third person. This is a pattern shared with Czech and Macedonian, and to some extent with Polish. It represents a point in development on the way toward that reached by Russian, shown in Table 3 above.<sup>21</sup> (See Brown et al. 2012 for the more complex situation of Bulgarian.) Interestingly, there are relevant forms potentially available, since ‘be’ when used as a copula has third person *je* singular and *sú* plural, but the auxiliary has different properties. The secondary split, within the past, is caused by the auxiliary’s lack of third-person forms. Thus linguists ready to postulate zeros can include zeros in the third-person cells of the past tense and for this cost can treat the past as consistently periphrastic.

A second, more subtle, example is the Sanskrit future; the data and analysis are entirely from Stump (2013), who gives many examples and a much fuller account. Sanskrit has a future formed from the nominative form of the agent nominal (glossed NMLZ), together with the auxiliary *as-* ‘be’ in the present tense; thus ‘I shall give’ is apparently expressed as ‘I am giver’. As with Russian and Slovak, then, there is a split between tenses. Here is an instance of the future.

<sup>20</sup> My thanks to Paul Kiparsky and David Short for discussion of the Slovak verb.

<sup>21</sup> Indeed, the comparison of Table 14 with Table 3 illustrates the difference between composition and feature signature: in Russian the past-tense forms are synthetic (hence they have the same composition), but they differ from the present in feature signature; in Slovak, in the first and second persons, they differ in both composition and feature signature.

- (2) Sanskrit (*Śatapatha Brāhmaṇa* 1.8.1.2, from Stump 2013:112)  
 tatas tvā pārayitā<sub>̣</sub>asmi<sup>22</sup>  
 from.that thee.SG.ACC rescue.NMLZ.SG.M-be.PRS.1SG  
 ‘I will rescue (lit: I am rescuer) you from that.’

This example shows that the verb takes a direct object in the accusative, despite its unusual periphrastic structure (Stump 2013). The agent nominal has a full paradigm available. In examples like the one we have just seen, however, the form is that of the masculine singular; the same masculine singular form of the agent nominal occurs in 3.

- (3) Sanskrit (*Gopatha Brāhmaṇa* 1.1.28, from Stump 2013:113)  
 mahac <sub>̣</sub>choka-bhayam prāptā<sub>̣</sub>smaḥ<sup>23</sup>  
 great.SG.ACC pain-fear.SG.ACC meet.NMLZ.SG.M-be.PRS.1PL  
 ‘We are going to meet with great pain and dread.’

Despite the plural subject, the nominal element in the periphrastic verb remains masculine singular. What has been said so far is true of the first and second persons. In the third person the picture changes. There is no auxiliary (reminiscent of Slovak), and the agent nominal now takes the appropriate number form (singular, dual, or plural) as in 4.

- (4) Sanskrit (*Bhojaprabandha* 55, from Stump 2013:115)  
 na jāne yātāras tava ripavaḥ kena ca pathā.  
 not know.1SG go.NMLZ.PL thy enemy.PL.NOM which.SG.INS and path.SG.INS  
 ‘I don’t know by which path your enemies will go.’

In 4 *yātāras* (literally ‘goers’) is plural, showing that the number forms are available. Note that the potential gender opposition is not used: the masculine is used even for a feminine singular subject. This is the key point for our distinction: the agent nominal has number forms available; however, they are not used in the (periphrastic) first and second persons, but they are used in the (nonperiphrastic) third person. This is a subtle split in the paradigm: the agent nominal has a full set of gender and number forms; when forming the future, the nominal is sensitive to number in the third person, but not in the first and second persons. And in this periphrastic use it is not sensitive to gender at all.

As a third instance of this interesting type of noncanonical behavior in respect of feature signature, consider this example from Archi.

- (5) Archi (Kibrik 1994:349)  
 buwa-mu b-ez dita<b>u χ<sup>w</sup>alli  
 mother(II)-SG.ERG SG.III-1SG.DAT early<SG.III> bread(III)[SG.ABS]  
 a<b>u  
 made.PFV<SG.III>  
 ‘Mother made bread for me early.’

We see that the dative pronoun, like the adverb and the verb, agrees with the absolutive argument (χ<sup>w</sup>alli ‘bread’). Table 15 gives a partial paradigm for the first-person singular personal pronoun (Chumakina & Corbett 2008, following Kibrik 1977:257–60); some thirty further (local) case values are omitted.

If we consider the cells representing the various case values, it is evident that two cells are very different from the remainder. The cells with multiple entries show the

<sup>22</sup> For clarity, Stump presents the two elements separated by <sub>̣</sub>. Normally the two parts would be pronounced together, and automatic sandhi processes apply.

<sup>23</sup> This the prepausal sandhi variant of the first-person plural form *smas* (Greg Stump, p.c., 5 August 2013).



ABSOLUTIVE	zon		
ERGATIVE	za-ri		
		SG      PL	
GENITIVE	I	w-is	b-is
	II	d-is	
	III	b-is	is
	IV	is	
		SG      PL	
DATIVE	I	w-ez	b-ez
	II	d-ez	
	III	b-ez	ez
	IV	ez	
COMITATIVE	za-ʔu		
SIMILATIVE	za-q'di		
COMPARATIVE	za-χur		
SUBSTITUTIVE	za-kl'ena		
SUPERESSIVE	za-t		
SUPERRELATIVE	za-ti-š		
SUPERLATIVE	za-ti-k		
SUPERTERMINATIVE	za-ti-kəna		
CONTELATIVE	za-ra-š		
CONTLATIVE	za-ra-k		
CONTALLATIVE	za-r-ši		
CONTTERMINATIVE	za-ra-kəna		
.	.		
.	.		
.	.		

TABLE 15. Archi: the first-person singular pronoun (partial paradigm).

gender and number value forms available there (four gender values, two number values, with regular patterns of syncretism, comparable to Table 11 above). Recall that this is the first-person singular pronoun, and these forms are to realize the incoming feature specification of the absolutive argument (as in 5). Their feature signature is different from that of the other cases. (It might appear that the cells involved can be predicted, but we shall see that this is not straightforward when we consider the other pronouns in §3.3 below.)

Though these examples are similar in showing a split that involves the composition/feature signature of the cells, rather than just the forms, they differ substantially in other ways. These other ways are the remaining dimensions of our typology, to which we now turn.

**3.2. JUSTIFICATION: MORPHOMIC VERSUS MOTIVATED.** This is an increasingly familiar distinction. There are patterns that are internal to the morphology, known as morphomic, following Aronoff (1994). The most familiar examples of morphomic patterns are from Romance languages (Maiden 2005, 2013). Here we find, for example, one stem variant in the first singular of the present indicative and throughout the present subjunctive, contrasting with a second variant elsewhere, a pattern that appears to lack motivation. Other morphomic patterns are found in Germanic (Blevins 2003), and another convincing example is found in the Cushitic language Dhaasanac (Tosco 2001: 111–205, discussed in Baerman et al. 2005:105–6, 183–86, 236–41); see also Stump 2001:169–211 on Sanskrit, Bonami & Boyé 2008 on Nepali, Kaye 2013 on Northern Talyshi, and Bond 2016 on Eleme.

For a less familiar example of a morphomic pattern, consider the fragment of data in Table 16 for the verbal forms of Burmeso, a language of the Mamberamo River area of Western New Guinea.<sup>24</sup> The data are entirely due to Donohue (2001:100, 102).

GENDER	ASSIGNMENT	CLASS 1 'see'		CLASS 2 'bite'	
		SINGULAR	PLURAL	SINGULAR	PLURAL
I	male	j-ih-	s-ih-	b-akwa-	t-akwa-
II	female, animate	g-ih-		n-akwa-	
III	miscellaneous	g-ih-	j-ih-	n-akwa-	b-akwa-
IV	mass nouns	j-ih-	j-ih-	b-akwa-	b-akwa-
V	banana, sago tree <sup>a</sup>	j-ih-	g-ih-	b-akwa-	n-akwa-
VI	arrows, coconuts	g-ih-	g-ih-	n-akwa-	n-akwa-

TABLE 16. Burmeso: morphomic patterns of syncretism (from Donohue 2001).

<sup>a</sup> Gender V could be considered 'inquate' (Corbett 2012:84–85, 179), since it contains only two nouns, and the agreements are simply an irregular combination: IV in the singular and VI in the plural.

The prefixal gender-number markers mark agreement with the absolutive argument. There are two inflectional classes, which are remarkably close to being canonical (as specified in Corbett 2009), according to two principles. First, the two classes are fully comparable and are distinguished as clearly as is possible. The sets of gender-number markers are completely distinct, so that each form predicts every other within the paradigm. Second, the distribution of lexical items over the inflectional classes is synchronically unmotivated. Donohue (2001:101) can find no semantic correlations for verbs in the two inflection classes, nor any phonological correlation (p.c., 2 May 2013). Moreover, the two sets of verbs are of approximately the same size, so it is not a matter of specifying a default class, leaving a few items to be lexically specified. The remarkable inflectional classes of Burmeso provide the backdrop for the morphomic pattern that interests us. Observe the patterns of syncretism in class 1. That genders I and II have a syncretic form in the plural makes semantic sense, and this is therefore left out of the account here. But the other syncretisms make no sense in terms of the semantics of the gender values. Furthermore, there is no possible phonological motivation for the pattern, since the stem is identical in the different cells. Hence we have a morphomic pattern. What makes this particularly convincing is that exactly the same pattern, but with different realization, is found in inflectional class 2.<sup>25</sup>

Morphomic patterns are often discussed in terms of stem allomorphy, particularly in Romance,<sup>26</sup> but like other types of split they can be induced by various phenomena. The Burmeso example shows this, since it involves patterns of syncretism rather than stem allomorphy. Morphomic patterns are particularly well founded when the same pattern serves for different phenomena. For instance, the pattern of suppletion in French (Table

<sup>24</sup> Burmeso was previously considered an isolate, but see Ross 2005 for discussion of its possible affiliation.

<sup>25</sup> A baroque alternative would be to propose separate gender values for singular and plural; this approach is appropriate for lexical exceptions (for example, Serbo-Croat *oko* 'eye', analyzed in §3.4) but not for the full lexicon. It would have the effect of hiding the pattern in Table 16 while stipulating a set of unmotivated mappings between gender values in the singular and plural in the lexical entries for nouns; that is, it would treat every noun as though it were exceptional. Donohue's approach states the mappings at the level of the feature values.

<sup>26</sup> Embick and Halle (2005) argue against recognizing stems at all, but their analysis of the Latin data is challenged succinctly by Vincent (2011:417–18, n. 2, 434), fundamentally by Aronoff (2012:39–46), and comprehensively by Taylor (2012:57–59, 68–76). An overview of the issue of stem allomorphy can be found in Bonami 2012a.

2) is found in other verbs for stem allomorphy. Maiden and O'Neill (2010) show defectiveness determined by a morphomic pattern, and an instance of periphrasis (determined indirectly through suppletion) is given in Table A1 in the appendix. Helpful discussion of the morpheme is provided by O'Neill (2011), Taylor (2012), and Round (2015).

We contrast splits that have a morphomic pattern with those that are motivated from outside the morphology.<sup>27</sup> This motivation may be of various types, the most obvious being semantic and phonological. We consider these in turn. SEMANTIC MOTIVATION is most easily seen by reference to natural classes in the feature system. In a reasonable feature system, perfective forms versus imperfective constitute natural classes, as do past versus nonpast, singular versus plural, and so on. Motivated segments of a paradigm are sometimes called 'subparadigms'. By contrast, first-person plural is not a natural class, since it requires reference both to person and to number. Anything beyond natural classes requires an extra step, and so needs additional justification.<sup>28</sup>

A step that is sometimes taken, rather too readily, is to decompose feature values into binary features and seek motivation there. This approach perhaps reflects the powerful influence of phonology on other domains. However, decomposition is appropriate only when the primitives can be explicitly justified. The most famous example of decomposing features is Jakobson's analysis of the Russian case system (1958), which he analyzed with three binary features. For all of its subsequent influence, the analysis is known to be problematic, at the empirical level and at the conceptual level (see Corbett 2012:18–21 for references). As Worth put it (1984:298): 'the cube was an enticing mistake'. For our purposes, however, the following observation is significant: Gerald Gazdar points out (p.c.) that there are 6,720 (8!/3!) possible ways to describe eight values using three binary features. Given this, in the absence of principled reasons for postulating particular binary features from the outset, it should not be taken as significant if there is an analysis using binary features that is partially successful.<sup>29</sup> More basically, decomposition of features needs to be argued for rather than assumed.

Starting now from the other side, so to speak, a morphomic pattern is also one that has no PHONOLOGICAL MOTIVATION. Naturally, the effects of fully regular phonological rules are a matter of phonology and not morphology. But some give phonology a greater role, and indeed occasionally a much greater role. Again, we should recognize that each such step represents a cost in accounting for the mapping from morphosyntax that is our concern here.<sup>30</sup> For instance, we may have rules similar to normal phonological rules but re-

<sup>27</sup> In his discussion of heteroclitism, Stump (2006:309) distinguishes paradigms that split along a morphosyntactic divide as 'cloven', while those that do not he terms 'fractured'. This definition covers half of what is necessary to prove a morphomic pattern; a morphomic pattern must be fractured, in Stump's sense, AND it must not be phonologically predictable.

<sup>28</sup> Of course, a new semantic analysis can provide a semantic justification for an apparently morphomic pattern. In French, there is a shared stem, whether regular or irregular, for future and conditional, while the inflections are shared between conditional and imperfect. This pattern might be considered morphomic, but Bonami and Boyé (2007:309–11), following the neo-Reichenbachian analysis by Verkuyl and colleagues (2004), argue that there is a semantic generalization justifying each of the pairings.

<sup>29</sup> Decomposition into binary features brings with it the need for a means of dealing with the superfluous values that arise even with features with three values (like person), and the more so with features with awkward numbers of values like five or nine. One strategy is to propose a geometry in addition, making one feature subordinate to the other. A proposal for person and number is given in Harley & Ritter 2002; see McGinnis 2005 and Cysouw 2011 for detailed discussion of its shortcomings. More generally, this reinforces the point that decomposition is indeed an additional step in accounting for the mapping from morphosyntactic specifications to forms, and one that would require careful justification.

<sup>30</sup> By way of comparison, de Lacy and Kingston (2013:311–18) argue that a sound change that has been phonologized may retain its phonetic motivation for a considerable time. Establishing the argument, from

stricted to occurring only with particular parts of speech. A further condition may be morphosyntactic (e.g. an otherwise normal phonological rule but applying only to verbs when in the present tense). A further step on the slippery slope is to allow in rules whose conditions are phonological, but whose output is not phonologically natural. This is an important distinction, made at various points in the literature. For instance, Baerman (2014:11) distinguishes what is ‘phonologically motivated’ from what is ‘phonologically predictable’; for discussion see Anderson 2011 and Nevins 2011. Specifically with regard to allomorphy, Bonet and Harbour (2012) discuss the continuum between fully systematic phonological rules and wholly unsystematic allomorphy, a discussion prefigured in Dressler 1985. For the unmotivated instances, it is worth recalling Aronoff’s recent reminder (2012:43) that phonological readjustment rules allow the rewriting of any string as any other string, as established by Johnson (1972).

Having seen how different approaches make the definition of ‘motivated’ more lax or more strict, we can return to the specific case of Burmeso. The split there, which is induced by patterns of syncretism, is clearly morphomic. As noted above, it cannot be motivated by semantics or by phonology. We can contrast this split with the opposite type of pattern that we saw earlier in Table 3. That Russian pattern divided present tense from past tense, which is a clearly motivated split, based on the semantics of tense. The split arose when the auxiliary verb was lost, leaving an old participle as the past tense (see, for instance, L’Hermitte 1978). As noted above, the Slovak situation in Table 14 represents an intermediate stage of development, with the auxiliary only partially lost. These splits have different origins, since motivated splits typically come from semantic or syntactic change (as just discussed in relation to Table 3), while morphomic splits are usually the remnants of phonological change (as in the modern Romance morphomic patterns, which go back to Latin patterns of stress; Maiden 2009). We must, of course, be careful to separate the origins of these patterns from our model of their synchronic situation.

The primary aim here has been to elucidate the distinction between morphomic and motivated splits, taking clear examples, but there are less clear examples, which have been taken to indicate that morphomic is a gradient notion (see O’Neill 2013 and Smith 2013 for recent discussion). It is worth observing too that morphomic and motivated splits can be found together, in quite complex systems. Consider the split verbal paradigms from the Oto-Manguean language Chiquihuitlán Mazatec given in Table 17.<sup>31</sup> All examples shown are of positive polarity.

We limit ourselves to the stems, specifically to the initial stem formatives (for example, *ba* and *kua* in the verb ‘wrap’) and the patterns they induce. But note that the superscript numbers represent tone: there are four level tones, with 1 being the highest, and ten contour tones. Tone can be an exponent of aspect, manifested as a lowering in the inchoative. The first verb given shows a straightforward motivated split, dividing neutral from inchoative aspect. The second verb (‘step on’) exhibits a splendidly morphomic split, contrasting the first-person singular and the undifferentiated third person with all other person-number combinations. The verb ‘eat’ has the motivated split, based on aspect, and the morphomic split (first singular and undifferentiated third ver-

---

tonogenesis in Athabaskan, is not straightforward. My point here is not whether they are right, but that they make the case in some detail. Similarly, we cannot simply assume that ‘old phonology’ motivates current morphological patterns: it is an argument that needs careful demonstration, as for instance in the debate between Anderson (2011, 2013) and Maiden (2011), and in the distinction of morphological versus phonological vowel harmony in Finley 2009.

<sup>31</sup> I am very grateful to Enrique Palancar for his help with these data.

		‘wrap’	‘step on’	‘eat’	‘remember’
NEUTRAL ASPECT	1SG	ba <sup>1</sup> štæ <sup>1</sup>	su <sup>3</sup> nɛ <sup>1</sup>	hi <sup>14</sup> nɛ <sup>3</sup>	ba <sup>3</sup> sæ <sup>1</sup>
	2SG	ba <sup>2</sup> šte <sup>2</sup>	nu <sup>3</sup> nɪ <sup>1</sup>	çi <sup>2</sup> nɪ <sup>2</sup>	ča <sup>2</sup> se <sup>2</sup>
	3	ba <sup>2</sup> štæ <sup>2</sup>	su <sup>3</sup> nɛ <sup>1</sup>	hi <sup>3</sup> nɛ <sup>3</sup>	ba <sup>3</sup> se <sup>2</sup>
	1PL.INCL	ba <sup>2</sup> šte <sup>2</sup>	nu <sup>3</sup> nɛ <sup>31</sup>	çi <sup>2</sup> nɛ <sup>2</sup>	ča <sup>2</sup> sɛ <sup>2</sup>
	1PL.EXCL	ba <sup>2</sup> štɪ <sup>24</sup>	nu <sup>3</sup> nɪ <sup>14</sup>	çi <sup>2</sup> nɪ <sup>24</sup>	ča <sup>2</sup> sɪ <sup>24</sup>
	2PL	ba <sup>2</sup> štu <sup>2</sup>	nu <sup>3</sup> nu <sup>1</sup>	çi <sup>2</sup> nu <sup>2</sup>	ča <sup>2</sup> su <sup>2</sup>
INCOMPLETE ASPECT	1SG	kua <sup>1</sup> štæ <sup>1</sup>	su <sup>3</sup> nɛ <sup>1</sup>	si <sup>14</sup> nɛ <sup>3</sup>	kua <sup>3</sup> sæ <sup>1</sup>
	2SG	kua <sup>2</sup> šte <sup>2</sup>	nu <sup>4</sup> nɪ <sup>1</sup>	ši <sup>2</sup> nɪ <sup>3</sup>	ča <sup>4</sup> se <sup>2</sup>
	3	kua <sup>4</sup> štæ <sup>2</sup>	su <sup>2</sup> nɛ <sup>1</sup>	si <sup>4</sup> nɛ <sup>4</sup>	kua <sup>4</sup> se <sup>2</sup>
	1PL.INCL	kua <sup>2</sup> šte <sup>2</sup>	nu <sup>4</sup> nɛ <sup>41</sup>	ši <sup>4</sup> nɛ <sup>42</sup>	ča <sup>4</sup> sɛ <sup>42</sup>
	1PL.EXCL	kua <sup>4</sup> štɪ <sup>24</sup>	nu <sup>4</sup> nɪ <sup>14</sup>	ši <sup>4</sup> nɪ <sup>34</sup>	ča <sup>4</sup> sɪ <sup>24</sup>
	2PL	kua <sup>2</sup> štu <sup>2</sup>	nu <sup>4</sup> nu <sup>1</sup>	ši <sup>4</sup> nu <sup>3</sup>	ča <sup>4</sup> su <sup>2</sup>

TABLE 17. Chiquihuitlán Mazatec: splits in verbs (Jamieson 1982).

sus the rest), with different forms for this specification in both aspects. The verb ‘remember’, and those like it, distinguishes remaining forms from first singular and third, and this split intersects with the motivated split according to aspect. The distinction between morphomic and motivated splits is discussed in Corbett 2015, but the constraint on nesting proposed there will need to be refined; it is consistent with the first three verb types given, but the remarkable pattern of verbs like ‘remember’ in Chiquihuitlán Mazatec is problematic. There is much more to be said, since the discussion has been exclusively on the patterns of initial stem formatives, leaving aside the other patterns, notably the tone patterns.

**3.3. SPECIFICATION OF PATTERN: LEXICALLY SPECIFIED TO FULLY REGULAR.** Although splits are noncanonical, they may be absolutely regular, in that they may apply to every relevant lexeme, including new borrowings.<sup>32</sup> Thus the split in the Russian verb (Table 3) applies to every verb in the language; it is regular across the lexicon. Now contrast this split with those that apply to smaller inventories of the possible lexemes. Thus the split in the paradigm of the French verb *aller* ‘go’ (Table 2) does not apply to every verb, but there are several where this split does apply. Note that they differ in the phenomenon that induces the split (in the sense of being the justification for positing a split): other French verbs do not have full suppletion as in Table 2, but they have less dramatic differences between the cells, involving differences in the stems (as in *jeter* ‘throw’, for instance). While the forms are different, the particular pattern, the split in the paradigm, is shared.

At the end of the regularity scale there are instances that are LEXICALLY SPECIFIED, where we need to specify, lexical item by lexical item, whether a particular split occurs. A particularly dramatic example is found in Archi, where some items of almost every part of speech are able to agree (with the absolutive argument). It is difficult to specify exactly which ones agree, though there are some regularities here (Chumakina & Corbett 2015). It seems clear that, in some instances at least, the items that agree will have to be specified item by item. For the personal pronouns, one of which was discussed above (§3.1), the issue is even ‘worse’, in that it is not just a matter of specifying which pronouns are sensitive to an incoming feature specification, but also a matter of specifying which of their cells are affected. Consider the pronouns given in Table 18.

<sup>32</sup> For a summary of the neuropsychological evidence for the different treatment of regular versus irregular see Marslen-Wilson 2007. Interesting psycholinguistic evidence suggesting that partially irregular verbs pattern with irregular verbs in German, indicating that regularity is a matter of the paradigm rather than of particular forms, is presented by Trompelt, Bordag, and Pechmann (2013).

	SINGULAR		PLURAL		
	1ST	2ND	1ST	2ND	
			EXCL	INCL	
ABSOLUTIVE	zon			nen-t'-u	
ERGATIVE	za-ri	un	nen	nen-a-w nen-a-r-u nen-a-b-u nen-t'-u <i>etc.</i>	ž'en
GENITIVE	w-is d-is b-is is <i>etc.</i>	wit	ulu d-olo b-olo olo <i>etc.</i>	la-w-u la-r-u la-b-u la-t'-u <i>etc.</i>	wiš
DATIVE	w-ez d-ez b-ez ez <i>etc.</i>	wa-s	w-el d-el b-el el <i>etc.</i>	w-ela-w d-ela-r-u b-ela-b-u el-t'-u <i>etc.</i>	wež
COMITATIVE	za-ɬu	wa-ɬu	la-ɬu		ž'wa-ɬu
SIMILATIVE	za-q'di	wa-q'di	la-q'di		ž'wa-q'di
COMPARATIVE	za-ɣur	wa-ɣur	la-ɣur		ž'wa-ɣur
SUBSTITUTIVE	za-kl'ena	wa-kl'ena	la-kl'ena		ž'wa-kl'ena
SUPERESSIVE	za-t	wa-t	la-t		ž'wa-t
SUPERELATIVE	za-ti-š	wa-ti-š	la-ti-š		ž'wa-ti-š
SUPERLATIVE	za-ti-k	wa-ti-k	la-ti-k		ž'wa-ti-k
SUPERTERMINATIVE	za-ti-kəna	wa-ti-kəna	la-ti-kəna		ž'wa-ti-kəna
CONTELATIVE	za-ra-š	wa-ra-š	la-ra-š		ž'wa-ra-š
CONTLATIVE	za-ra-k	wa-ra-k	la-ra-k		ž'wa-ra-k
CONTALLATIVE	za-r-ši	wa-r-ši	la-r-ši		ž'wa-ra-ši
CONTTERMINATIVE	za-ra-kəna	wa-ra-kəna	la-ra-kəna		ž'wa-ra-kəna
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.

TABLE 18. Archi: personal pronouns, partial paradigm (based on Kibrik 1977:257–60).

The cells with multiple entries show the gender and number forms available. The presentation of the complex cells is reduced from that given for the first-person singular pronoun (Table 15 above). There are four singular forms; according to the regular patterns of syncretism of Archi, the plural forms are: genders I/II plural in *b-/<b>* and genders III/IV plural realized by the bare stem or by *-t'*. These cells mark agreement with the absolutive argument, as in 5 and as we see again in 19 below. Other forms cannot show agreement. Clearly, the cells involved in agreement do not form a natural class.<sup>33</sup> Nor are they defined by an agreement rule. It is rather that certain cells are sensitive to 'incoming' features. The agreement rule specifies the agreement controller (the absolutive argument) and the agreement domain; targets show agreement if their paradigm permits it, but the agreement rule does not refer to those cells. Nor is there a phonological prediction. Being consonant-initial is not sufficient to prevent the additional inflection, as is clear in the first-person plural inclusive pronoun. Equally, being vowel-initial is not sufficient to ensure the additional inflection, as the second-person singular ergative shows. Thus the pronouns that have the split must be lexically specified, and it must be further specified which cells are involved.

**3.4. RELEVANCE: 'INTERNAL' AND 'EXTERNAL' SPLITS.** We would expect such splits to be lexeme-internal, that is, with no effect in syntax. When we see an example like

<sup>33</sup> For comparison with other languages of the family, see Kibrik & Kodzasov 1990:220–23.

French *aller* ‘go’ (as in Table 2), we do not expect the differences in the paradigm to have any impact on syntax: all of its forms take a single argument, for instance. The syntax cannot ‘see’ the differences in form. However, some particularly interesting splits are external—they have syntactic consequences. Given their importance, we consider three contrasting instances.

We look first at the Serbo-Croat<sup>34</sup> word *oko* ‘eye’, whose paradigm is given in Table 19.

	SINGULAR	PLURAL
NOMINATIVE	oko	oči
ACCUSATIVE	oko	oči
GENITIVE	oka	očijū <sup>a</sup>
DATIVE	oku	očima
INSTRUMENTAL	okom	očima
LOCATIVE	oku	očima

TABLE 19. Serbo-Croat: the noun *oko* ‘eye’.

<sup>a</sup> More rarely *očī* (see dictionary referenced as SANU).

There is a split, created by the consonant alternation  $k \sim \check{c}$ , which is no longer a live alternation in Serbo-Croat inflection. While the split is therefore not phonologically motivated, it is motivated in that it follows a semantic and morphosyntactic divide between singular and plural. The split is irregular, in that only this noun and *uxo* ‘ear’ (plural *ušī*) behave in this way. The different stems belong to different inflectional classes (making the lexeme heteroclitic; see Table 7). The particular interest of the split is seen in the following examples.

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

We see that the split brings with it a change of gender: agreement is neuter in the singular (example 6) but feminine in the plural (example 7). These gender values are as would be expected from the usual gender assignment rules of Serbo-Croat, the rules assigning gender according to lexical semantics and inflectional class. Nouns whose (complete) paradigm matches *oko* ‘eye’ in the singular are neuter, while those whose (complete) paradigm matches *oči* ‘eyes’ in the plural are feminine. Thus the split, limited to just two nouns, is externally relevant in that the two segments control different agreements.

Contrast that with Georgian verbs where, as part of a complex system, verbs may govern different cases according to different combinations of tense, aspect, and mood. We consider just transitive and ditransitive verbs here.

- (8) Georgian (Alice Harris 1981:1, and p.c.)  
 glex-i      tes-av-s      simind-s  
 peasant-NOM sow-SERIES.I/III-3SG corn-DAT  
 ‘The peasant is sowing corn.’ (present)

<sup>34</sup> I use ‘Serbo-Croat’ here as a cover term for Bosnian, Croatian, Montenegrin, and Serbian (Corbett & Browne 2009). The differences between these do not affect the examples given.

- (9) *glex-ma da-tes-a simind-i*  
 peasant-ERG PFV-SOW-3SG corn-NOM  
 ‘The peasant sowed corn.’ (aorist)
- (10) *glex-s da-u-tes-av-s simind-i*  
 peasant-DAT PFV-3-SOW-SERIES.I/III-3SG corn-NOM  
 ‘The peasant has sown corn.’ (perfect)

Note that ‘dative’ is the traditional name for the third core case in the system (in addition to nominative and ergative); it is used with direct as well as indirect objects, and with some noun phrases with certain subject properties (as in 10); these subject properties are discussed in detail in Harris 1981:116–43 and Anderson 1984.

A good way to come to grips with the complexity of the Georgian splits is to ask how we would predict the cases required in examples 8–10. We first need to know the ‘screeve’ and series of the verb (for which see Harris 1981:39–47); a screeve is simply a combination of tense, aspect, and mood: thus the aorist (past, complete aspect, indicative mood) is one screeve, out of ten or eleven, distinguished by a variety of morphological means. These screeves are organized into three series, which are key to understanding case requirements. The present tense as found in 8 belongs in series I, the aorist in 9 belongs in series II, and the perfect in 10 belongs in series III (though the markers just in these examples are not sufficient to prove this, particularly for 9).

While knowing the series of the verb is a good start, it is insufficient for predicting the required case frame. There are three patterns for the core arguments, shown in Table 20 (I add in the indirect object for completeness).

	subject	direct object	indirect object
pattern A	ERGATIVE	NOMINATIVE	DATIVE
pattern B	NOMINATIVE	DATIVE	DATIVE
pattern C	DATIVE	NOMINATIVE	<i>tvís</i> -nominal

TABLE 20. Georgian: patterns of case frames (Harris 1981:1).

In order to know which pattern to use, we also need to know the class of the verb (Harris 1981:259–67). The four classes are defined by Harris through a combination of morphological, syntactic, and semantic factors; there are regular derivational processes that lead to a verb being in a particular class. In principle, a particular verb belongs to one of the four classes; the verb in our examples above is a class 1 verb. Given both the class of the verb, and the series it is in, the pattern of case values can be predicted according to this key (Harris 1981:1–2); see Table 21.<sup>35</sup>

	series I	series II	series III
class 1: transitive	B	A	C
class 2: unaccusative	B	B	B
class 3: unergative	B	A	C
class 4: inversion (psych verbs)	C	C	C

TABLE 21. Georgian: calculation of appropriate case frame (Harris 1981:2).

Given this link, we can establish the case frame for each combination. In the ones we have, 8 has a class 1 verb, in series I (which includes the present screeve), and so from Table 21 we see that we should have pattern B. That is: nominative (subject) and dative (direct object, and if there were an indirect object that would be dative too). In example

<sup>35</sup> The distinction between class 1 and class 3 is required for other purposes (Harris 1981:259–67), but they have the same effect for calculating the case frame.



9, with the aorist (series II), we should find pattern A, and indeed we have an ergative subject and a nominative direct object. Finally in 10, we expect pattern C, dative and nominative, and that is what we find.

The system is indeed quite complicated (see also Anderson 1992:141–58 for a different approach to the same data, and Aronson 1982, 1991 and Hewitt 1995 for further detail). For our purposes, the key points are these: Georgian verbs distinguish various tense/aspect/mood combinations (screeves) by morphological means. Unusually these have an external effect. Not every screeve is different; rather, they fall into three sets or series. These series are morphological groupings that are not coherent groupings in semantic terms (they are not motivated). This is not sufficient to predict a verb's behavior, however, since verbs fall into four different classes. Given a verb's class, and the series it is used in, the case frame is predictable (according to Table 21 above). Thus the split in the verbal lexeme according to tense/aspect/mood, seen in the attendant morphological distinctions in 8 to 10, is reflected externally in different case frames. The internal morphological split is found externally in the different case requirements.

From the evidence so far, we might already think that Georgian is one of the most complex and interesting examples of split lexemes (and for the rise of this system see Harris 1985, 2008, Anderson 1992:358–59, and the discussion in §4.2 below). But there is an additional element to the split. Georgian verbs have markers to encode the number and person of their core arguments. The markers show different forms; some are prefixal and some are suffixal (hence noncanonical in terms of affixal consistency (§2.3)); they are largely mutually exclusive in particular slots (Anderson 1992:145); there is no unique exponent for particular morphosyntactic descriptions.

Our three examples are sufficient to hint at the interest of the system. There are three sets of markers, which Anderson (1992:145) labels the *v*-set, the *m*-set, and the *h-/u-/e-/a*-set. The suffixes *-s* (in 8) and *-a* (in 9) are alternatives from the *v*-set, and they indicate the subject. The *-s* in 10 is the same marker, but here referencing the object, while in that example the *-u-* is from the *h-/u-/e-/a*-set, referencing the subject (it indicates third person but does not differentiate number). The *m*-set does not occur in our examples. We would need a much more extensive set of examples to demonstrate this scheme fully (see Harris 1981). But the result is that we cannot predict the markers to be used just by looking at the grammatical role of the argument, nor by looking at the argument's case marking. However, we do have a way of predicting the markers already available, namely Table 20 above, which we can extend to include the verbal markers, as in Table 22 (see Anderson 1992:145; pattern labels have been changed to Harris's scheme).

		subject	direct object	indirect object
pattern A	case	ERGATIVE	NOMINATIVE	DATIVE
	verbal marker	<i>v</i> -set	<i>m</i> -set	<i>h-/u-/e-/a</i> -set
pattern B	case	NOMINATIVE	DATIVE	DATIVE
	verbal marker	<i>v</i> -set	<i>m</i> -set	<i>h-/u-/e-/a</i> -set
pattern C	case	DATIVE	NOMINATIVE	<i>tv</i> is-nominal
	verbal marker	<i>h-/u-/e-/a</i> -set	<i>v</i> -set	none

TABLE 22. Georgian: patterns of case frames and verbal markers.

We see that knowing the case of the argument is insufficient, and knowing the grammatical role is equally insufficient: we need both.<sup>36</sup> That is precisely the information

<sup>36</sup> And indeed more besides, such as the distribution of the alternative suffixes *-s* (in example 8) and *-a* (in 9) from the *v*-set, which were mentioned earlier.

given by the three patterns A–C, and the right pattern is determined by the series and the verb class in Table 21. Thus the deep split within the verb is reflected internally in its inflectional morphology (indicators of tense, aspect, and mood) and in its marking of its core arguments, and externally in the cases that it governs. On the one hand, then, the split concerns form and does not follow a motivated pattern.<sup>37</sup> On the other hand, it is regular and has important external relevance.

Our third example shows a particularly rare split, found in Gaelic. There are three case values, of which the dative is reserved for prepositional government (even there, not all prepositions require it). Consider the noun for ‘sea’, given in Table 23.

	INFLECTION	GENDER
NOMINATIVE	muir	masculine
GENITIVE	<b>mara</b>	feminine
DATIVE	muir	masculine

TABLE 23. Gaelic *muir* ‘sea’.

This noun is given in various sources, including Adger 2009 and Konstantopoulos 1998:17. And Lamb (2001:27) states that: ‘There are a small number of nouns with defective gender marking which may be either masculine or feminine (depending on dialect), or treated differently depending on case. An example of the latter is *muir* ‘sea’ which is often masculine when nominative and feminine when genitive (*mara* G.)’. A similar point is made in Lamb 2008:206. The origin of the phenomenon, according to Mackinnon (1910:302), can be traced back to this noun having earlier been neuter. While there are various statements in the literature about this gender switch, there is little hard evidence. William Lamb offers these examples as demonstrating the situation above.

(11) Gaelic (William Lamb, p.c.)

tha am muir thall an sin  
 be.PRS DEF.M.NOM sea.NOM over.there there  
 ‘the sea is over there’

(12) tha fuaim na mara àrd an seo

be.PRS sound DEF.F.GEN sea.GEN high here  
 ‘the sound of the sea is high (loud) here’

This indicates a split in the noun’s paradigm, dividing the genitive from the other two case values;<sup>38</sup> and when there is irregularity in Gaelic nouns it typically affects the genitive. This particular split is externally relevant, however, in that it brings with it a switch in gender agreement in the genitive.<sup>39</sup>

Not all dialects are alike in this (Gillian Ramchand, p.c.). Thus Dwelly’s famous dictionary (1902–1911) states that: ‘In *Lewis* the *nom.* is *f.* and the *gen.* is *m.*’. This means

<sup>37</sup> There are comparable issues in various Indo-Aryan and Iranian languages, where a split based on the different stems of the verb gives rise to different requirements of case government. These stems correspond to a split based on tense and aspect to a greater or lesser degree in the different languages (in other words, the split varies between motivated and morphomic); see Witzlack-Makarevich 2011:144–47 for discussion and references.

<sup>38</sup> To demonstrate the gender of the noun in the dative requires adjectival agreement; William Lamb points out (p.c.) that for younger speakers the gender distinction here tends to be lost; it is only for older speakers that gender in the dative could be established.

<sup>39</sup> Carstairs-McCarthy (1994:771) suggests that ‘mixed’ gender values, those with no target realizations of their own, can be based only on number (as in Romanian), not on other features. If he would count Gaelic as having a mixed gender value (equivalent to masculine in the nominative and feminine in the genitive), this would be a counterexample to his claim. He might reasonably, however, count *muir* ‘sea’ as a lexical exception, not requiring a mixed gender value.

that Dwelly gives the Lewis dialect as having the opposite situation to that described above. And according to Mark (2003:443), though dictionaries usually give *muir* ‘sea’ as masculine in the nominative and feminine in other cases, as in 11 and 12, it is ‘very commonly’ feminine in all cases.

Our three examples contrast in regularity and in the features involved; what they share is that the split in the paradigm has external relevance in syntax.

**3.5. SPLIT IS EXTERNALLY RELEVANT AND IS NOT CONSISTENT.** There is a further dimension to our typology that logically arises only with respect to external relevance. If a split in a paradigm is externally relevant, that external effect may be consistent or inconsistent. Consistent requirements are what we expect—irrespective of splits. Thus when we say that a verb or an adposition governs the instrumental case, we imply that whatever the governee, whether a pronoun, a full phrase, adjacent to the governor or separated from it, the instrumental will be required. Similarly, when we state that a noun is of feminine gender, we imply that whatever agreement target is involved, that target will be feminine. However, there are some interesting exceptions. We take agreement as the example here (see Corbett 2010 for analysis of inconsistent government). While normal consistent controllers require the same agreement on all targets, we also find nouns, known as hybrid nouns, whose gender varies according to the agreement target. Before linking this to splits, let us first consider a familiar instance of a hybrid.

(13) German (Siemund & Dolberg 2011:492)

Das Mädchen dachte an die Blume-n, die sie/es  
 ART.SG.N girl think.PST.3SG of ART.PL flower-PL REL.PL 3SG.F/3SG.N  
 vergessen hatte.  
 forgotten AUX.PST.3SG  
 ‘The girl thought of the flowers that she had forgotten.’

While the noun *Mädchen* ‘girl’ appears to be neuter, according to the form of the article, its gender is not straightforward. The personal pronoun is often feminine, though neuter is also found. The noun is a hybrid, having more than one gender, depending on the target. These agreements, made possible by certain types of lexical semantics, do not vary without limit; the patterns of such agreements are constrained by the AGREEMENT HIERARCHY (Corbett 2006:206–33).

Having clarified the notion of inconsistency in external behavior, we can return to split lexemes. With these, too, we find inconsistency in external behavior, though external consistency is the more common. Thus, as an example of a split with external consistency, we saw that Serbo-Croat *oko* ‘eye’ is neuter when singular, while its plural *oči* is feminine. Both parts of the split are externally consistent, in that whatever agrees with *oko* is neuter (singular), and whatever agrees with *oči* is feminine (plural). Now, for an example of a split with externally inconsistent behavior we turn to the Nordreisa dialect of Norwegian, discussed in Enger & Corbett 2012. Here the split is determined by definiteness. The noun *mamma* ‘mum’ is, unsurprisingly, feminine, except when it is definite. When it is definite, and there is an agreeing possessive within the noun phrase, this target is masculine, as in 14.

(14) Nordreisa dialect of Norwegian (Enger & Corbett 2012:299)

/mama-n di:-n/  
 mum-DEF.SG your-M  
 ‘your mum’

Masculine agreement with *mamma* ‘mum’ is restricted to the noun phrase, in the configuration of 14, when the noun is definite; all other agreements are feminine. Thus this

split triggers inconsistent external behavior, with masculine agreement in 14 and feminine in other circumstances. The pattern of agreement shown is in accord with the agreement hierarchy. For full details, see Enger & Corbett 2012.<sup>40</sup> A comparable example is Serbo-Croat *dete* ‘child’, plural *deca*, described in detail in Corbett 2007b. Since this distinction (between consistent and inconsistent external effects) is applicable only to a small part of the overall typology, namely the externally relevant splits, I set it aside as we tackle the overall pattern in the next section.

**4. COMPLEXITY: HOW THE SPLITS LINE UP.** We have established four criteria according to which splits may be distinguished: form versus composition/feature signature (§3.1), morphomic versus motivated (§3.2), lexically specified versus regular (§3.3), and internally versus externally relevant (§3.4). Evidently, the very existence of a split in a paradigm is noncanonical. But we can iterate the analysis, and consider what type of split would be a canonical split. (Just as, for instance, we say that suppletion is a non-canonical realization of morphosyntactic specification, but can then specify canonical suppletion, as in Corbett 2007a. Similarly, inflection classes are themselves noncanonical, but we can go on to establish criteria for canonical inflection classes, as in Corbett 2009.) Specifically for splits, we may establish the criteria for canonical morphological irregularity, which would be entirely internal to the lexeme (as in the first row of Table 24). Of course, if one were keen to minimize the role of morphology, one could take the opposite position, and this also gives an alignment of criteria, as in the second row. From either perspective, we might have believed that the four criteria would line up neatly, as in Table 24.

	CRITERION 1	CRITERION 2	CRITERION 3	CRITERION 4
Most canonical possibility for morphological irregularity:	form	morphomic	lexically specified	internal
Least canonical possibility for morphological irregularity:	composition/feature signature	motivated	regular	externally relevant

TABLE 24. Hypothetical alignment of criteria.

As we have seen, the real situation is much more complex, but the idealization in Table 24 is a useful start. If we had the combination in the first row, we would have an instance of the most straightforward morphological irregularity. That is to say, an item with a form that does not match expectation (based on the rest of the morphological system), is not externally motivated, is irregular (comparing across the lexicon), and has no external relevance. Such examples are entirely a matter of inflectional morphology. We should take such instances as reflecting canonical morphological splits. At the other extreme we have splits that are totally noncanonical in terms of morphology, affecting the composition/feature signature, being motivated, regular, and externally relevant. These

<sup>40</sup> The situation is more complicated since the feminine has a weak position in this dialect: almost any noun that can take the feminine determiner *ei* can optionally take the masculine determiner *en* (but not vice versa). Still, the important point is that the noun *mamma* triggers only masculine agreement—never feminine—in the definite (within the noun phrase).

The Bugurđzi variant of Romani has a fascinating parallel to the Norwegian dialect data. *Mamos* ‘granny’ is found with a masculine article but feminine adjective, while *sluga* ‘servant’ takes a feminine article and masculine adjective. This is according to Elšik (2000), but see also Boretzky 1993:22, 32, 37.

are intuitively ‘worse’ as lexemes, much further from the profile we might have constructed. Given these two extremes, we need a way of thinking about all of the phenomena that fall between the extremes, in different respects.

A helpful way to grasp the problem is to arrange the possibilities within a Boolean lattice, as in Figure 1. This idea was used by Brown and Chumakina (2013:7), following a suggestion by Roger Evans.

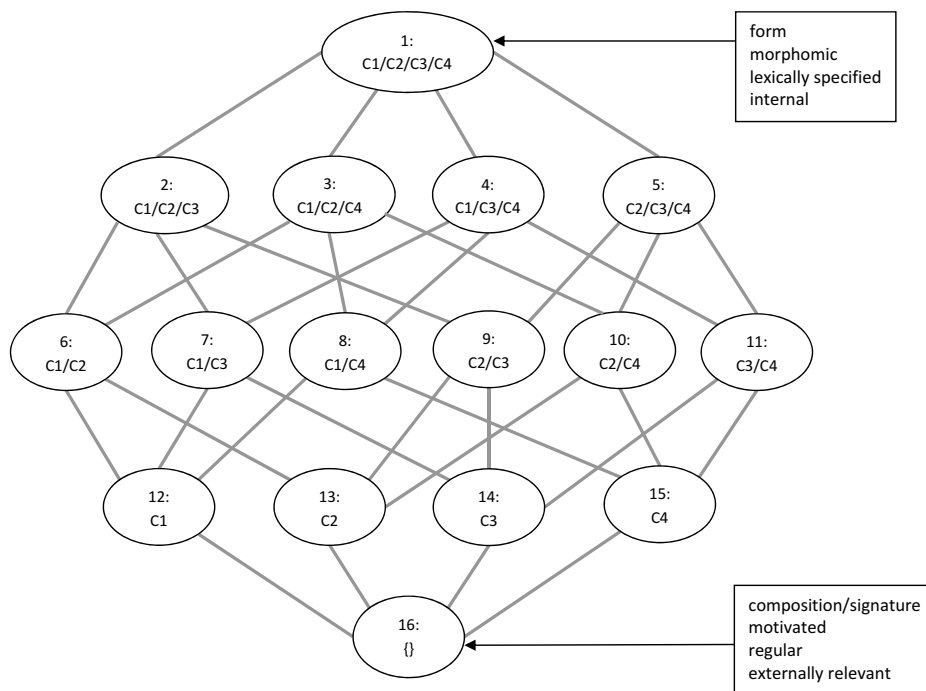


FIGURE 1. Lattice of possibilities.

This lattice gives us the theoretically possible combinations for the four criteria. At the top, we have the situation where the four criteria (C1–C4) all point to straightforward morphological irregularity (which is the situation from the first line of Table 24). In the next row, we have the four possibilities that arise if just one criterion is changed; thus at the first position in row two, three criteria point to straightforward morphological irregularity and one (C4) points in the other direction, and so this position is indicated C1/C2/C3. (A more cumbersome notation would be: +C1, +C2, +C3, –C4.) Then a second criterion does not point to straightforward morphological irregularity (shown in row three), and a third (row four). Finally, we have the situation where all of the criteria indicate that we have the least canonical behavior; this is the situation represented in the second line of Table 24. The possibilities are numbered 1–16 for reference in the appendix.

Let us now consider the real examples we have found. I illustrated each of the four criteria, and noted in passing that our examples varied in terms of other criteria too. Some of our examples are included in the lattice in Figure 2 (through brief indicators), together with others that represent combinations (though not criteria) that did not arise in our previous discussion.

There are interesting conclusions from the picture in Fig. 2 taken as a whole, which are discussed in §4.1; particular attention is paid to one set of specifications in the lattice in §4.2.

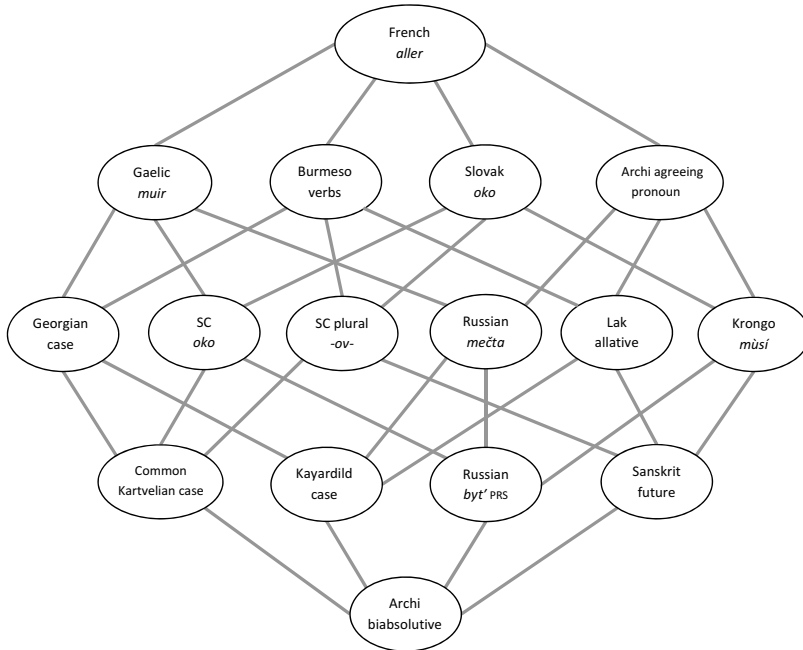


FIGURE 2. Indicative instances of splits.

**4.1. INTERACTIONS OF CRITERIA: THE OVERALL PICTURE.** Given the wholly unexpected nature of some of the splits we have been analyzing, the picture we see in Fig. 2 is striking in its completeness. As mentioned above, not all of the combinations have been discussed so far: of the remaining combinations, some are considered in §4.2, and the basic documentation for all is given in the appendix. Since part of the aim here is to deepen our knowledge of the concept ‘possible lexeme’, this is an important result for our ‘exploratory’ canonical typology: all of the possibilities are found, which means that there are instances of lexemes with remarkable behavior. There exist lexemes that we would not have considered possible. It is true that some combinations are easier to find than others, but the existence of examples, however few, exhibiting the more exotic combinations is unexpected and significant.

The fact that we find all possible combinations of criteria confirms the fact that the criteria are genuinely orthogonal. This justifies our canonical method of laying out and exemplifying the coordinates of a multidimensional space. Within this space we can locate and seek to understand the varied instances of lexical splits. Given that all of the criteria are orthogonal, the underlying logic is declarative. One might consider transposing the account into optimality-style constraints. To the extent that this succeeds, it would be a trivial reimplementation. The machinery of OPTIMALITY THEORY (OT) would not clarify the typology, since we would need to stipulate different rankings for the same pairs of criteria. Of course, there could be interesting implementations of individual phenomena discussed here within OT, where further constraints might be introduced to attempt to motivate the alternative rankings. An obvious starting point would be the papers in Downing et al. 2005. All the same, such an attempt is not promising. Since our phenomena lie not merely at the poles of any given dimension in our space, but rather also at many intermediate positions, each dimension would demand not two opposing OT constraints, but a larger set of ranked and interleavable subconstraints (in

the manner of Prince & Smolensky 1993:144ff.); for a parallel argument regarding classic OT implementations and phonetic intermediacy, see Flemming 2001:32–34. In such a case, the OT reimplementation would become decidedly nontrivial, though without any explanatory gain.

The layout of Fig. 2 also gives an indication of diachrony. The most straightforward instances of morphological noncanonicity arise from phonological causes: regular alternations cease to be current and morphological irregularities remain. These are the instances that retain criterion 1 (closer to the top and left of the lattice). Even the suppletion of French *aller* ‘go’ can be traced back ultimately to a morphomic pattern of phonological origin, which was then the target of invasive suppletion. The least straightforward instances of morphological noncanonicity are those that do not retain criterion 4 (closer to the bottom and left of the lattice). These have their origin in semantic and syntactic changes. Despite the completeness of the picture, it is the latter type, those that are externally relevant, that are—unsurprisingly—the hardest to find. They deserve special discussion, and two types are considered in the next subsection.

**4.2. INTERACTIONS INVOLVING EXTERNAL RELEVANCE.** Instances of lexical splits that are externally relevant are rare. We consider here an instance where the criteria show the least canonical type of morphological irregularity (type 16 in the lattice). Then we discuss type 12, because the evidence here requires explanation.

Consider these data from Archi (thanks to Marina Chumakina; see also Kibrik 2003b:562–64). Archi has periphrastic constructions based on the auxiliary ‘be’ and one of the possible converbs. The syntax is basically ergative. Given that, example 15 with the imperfective simultaneous converb *-ši* is unremarkable.

- (15) Archi (Marina Chumakina, p.c., consulting with Bulbul Musaeva)  
 tu-w-mi      p’aha-r-ši      i      qilin  
 that-I.SG-ERG smoke-IPFV-CVB [IV.SG]be.PRS cigarette(IV)[ABS.SG]  
 ‘He is smoking a cigarette.’

The periphrastic verbal complex consists of a converb and an auxiliary. The subject (gender I singular) stands in the ergative, and so cannot control agreement. The object *qilin* ‘cigarette’ stands in the absolutive, and controls the agreement on the auxiliary.<sup>41</sup> Since *qilin* ‘cigarette’ is singular and gender IV, the agreement on the auxiliary is singular and gender IV. This is indicated by the lack of a prefix (there is a prefix for agreement with each of the other three genders, and a singular-plural opposition, but with interesting patterns of syncretism, comparable to those seen earlier in Table 11).

Example 15 would be appropriate if, for instance, it is indirect evidence that allows the speaker to deduce the smoking. There is an alternative construction, however, which is more likely to be used if the speaker actually witnesses the smoking (example 16).

- (16) tu-w      p’aha-r-ši      w-i      qilin  
 that-I.SG[ABS] smoke-IPFV-CVB I.SG-be.PRS cigarette(IV)[ABS.SG]  
 ‘He is smoking a cigarette.’

Here both subject and object are in the absolutive, instantiating what is known as the BI-ABSOLUTIVE CONSTRUCTION. The subject is now in the absolutive, and this means that it is able to control agreement on the auxiliary. There is agreement in gender and number, specifically gender I and singular. The important contrast is with any synthetic verb form, as in 17.

<sup>41</sup> In principle, it controls agreement on the converb too, but this verb has no agreement slot. Compare example 19 below, where the converb has an initial agreement slot, and the lack of a marker indicates gender IV singular.

- (17) tu-w-mi      qilin                      p'aha-r  
 that-I.SG-ERG cigarette(IV)[ABS.SG] smoke-IPFV  
 'He smokes a cigarette.' (He is a smoker.)

In such examples there is no possibility of the transitive subject being in the absolutive case (and hence being potentially able to control agreement); it must be in the ergative.<sup>42</sup> It is the periphrastic part of the verbal paradigm that makes this possible. That is, the possibility of having the transitive subject in the absolutive, together with a transitive object in the absolutive, is licensed by periphrastic verb forms only.<sup>43</sup> Hence the Archi verb can have different syntactic possibilities when it is periphrastic (16) as compared to when it is synthetic (17).

It is not just any periphrastic form that allows this; the options depend on the particular converb. For the imperfective simultaneous converb ending in *-ši*, the biabsolutive construction is possible, as we have just seen, but never obligatory, whereas for the imperfective converb in *-mat* it is the only option.

- (18) to-r                      č'il                      ɸ'aža-r-mat d-i  
 that-II.SG[ABS] hay(IV)[SG.ABS] cut-IPFV-CVB II.SG-be.PRS  
 'She is still cutting hay.'

This would be appropriate if the woman was expected to finish cutting the hay, but is still busy. The auxiliary *d-i* is gender II singular, agreeing with *tor* 'she'.

In instances where we have two absolutives, we should ask whether they can control agreement on different targets. Given targets with agreement slots, the two absolutive arguments can indeed each be an agreement controller.

- (19) tu-w                      q'onq'                      o<r>k'in-ši                      w-i  
 that-I.SG[ABS] book(IV)[SG.ABS] [IV.SG]read<IPFV>-CVB I.SG-be.PRS  
 ez  
 [IV.SG] I.SG.DAT  
 'He is reading a book for me.'

In this example, again with the converb in *-ši*, the absolutive subject controls agreement on the auxiliary. The absolutive object controls agreement on the lexical verb: this is one that has a prefixal slot, and agreement in gender IV is realized by the bare stem (as on the auxiliary in 15). As we saw in Table 15, the first-person singular pronoun, when in the dative, also has an agreement slot, and this too agrees with the absolutive object. The agreement possibilities of dative pronouns and adverbs, in the biabsolutive construction, are the subject of ongoing research.

<sup>42</sup> The use of the neutral word order in example 17 fits better the present habitual verb form.

<sup>43</sup> Archi is thus important as a 'proof of concept', since it is evidence for the four extremes in our typology and since putative phonological and syntactic accounts can be disposed of convincingly. Phonological accounts do not get off the ground, since the split is between periphrastic and nonperiphrastic verb forms, and the periphrastic forms are built on formants with no phonological commonality. A syntactic account is more plausible: at a quick glance, one might be tempted to propose that the key examples are in fact biclausal and that the specifics of the biabsolutive construction are syntactic in nature. However, we have a project considering and comparing how three syntactic theories (HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR (HPSG), LFG, and MINIMALISM) can account for the agreement phenomena in Archi (full information can be found online: <http://fahs-wiki.soh.surrey.ac.uk/wiki/projects/fromcompetingtheoriestofieldworkarchi/archi.html>). The fourth topic documented on this site is precisely the biabsolutive. Our consultants, Robert Borsley (HPSG), Louisa Sadler (LFG), and Maria Polinsky (minimalism), independently examined the syntactic evidence on the Archi biabsolutive. Each of them rejected a biclausal analysis, within their particular framework (see the wiki for details). In the typology proposed by Gagliardia and colleagues (2013), it is a language of the Lak rather than the Tsez type. See also Forker 2012 for a survey of the biabsolutive in Nakh-Daghestanian languages.



While 19 involves marking by means of bare stems, the following example with gender III gives a clear comparison.

- (20) to-r                      b-ez                      χ<sup>o</sup>ošon                      b-a-r-ši  
           that-II.SG[ABS] III.SG-1SG.DAT dress(III)[SG.ABS] III.SG-make-IPFV-CVB  
           d-i  
           II.SG-be.PRS  
           ‘She is making a dress for me.’

In 20 the converb agrees with the absolutive argument that is the object (in gender III), as does the dative pronoun, while the auxiliary agrees with the absolutive subject (gender II).

Since we are dealing with periphrasis, the split is one based on the composition of cells. The split is motivated (based on the semantics of the converb) and fully regular. The key point is that the particular converb licenses the biabsolutive construction; the possibility of having two absolutive arguments (and so two different agreement controllers) arises because the verb is in a periphrastic form.

Now consider type 12 in the lattice, labeled ‘Common Kartvelian case’. The specification for this is that it concerns form, and it is motivated, regular, and relevant. An example would be a language in which all verbs were split according to form, in a motivated way (say, for tense or aspect), and this split was externally relevant (say, it changed the case requirement). In other words, it is like the Georgian example already discussed (§3.4), except that the distribution of stems is motivated. It seems likely that the original Common Kartvelian situation from which Georgian evolved matched this specification. Indeed, the earliest attested stages of Old Georgian (before the ninth century) were rather like this (Alice Harris 1985, 2008, and p.c.):

Series I opposes Series II aspectually ... All of the forms of Series I are of durative aspect, while those of Series II are punctual, with the exception of the habitual II (Machavariani 1974; Schmidt 1963, 1966). The habitual II expresses durative, continuous (ongoing), or repeated action ... (Harris 1985:96)

Thus, even in Old Georgian the series were aligned fairly closely with semantics: the split is highly motivated, apart from the habitual II. And Harris’s reconstruction for Common Kartvelian, Georgian’s parent language, suggests a picture close to the specification: ‘Series I expressed durative aspect and had accusative case marking while Series II expressed predominantly punctual aspect and had ergative case marking’ (Harris 1985:428). Thus this one part of the lattice is represented by a reconstruction, but one made well before and quite independently of the typology proposed here.

We have now seen evidence for the four criteria, and for several of the combinations of criteria—including the most challenging ones—shown in the lattice. All remaining combinations that have not already been covered in the text are justified in the appendix.

**4.3. THE LOGIC BEHIND THE CRITERIA.** It is reasonable to base a typology on any criteria that yield interesting results, and clearly the four criteria chosen have proved their worth in this respect. However, approaches in canonical typology attempt not just to sample the theoretical space but to cover it. Our four criteria attain this higher requirement. We anchor the space in a single split in a single lexeme. This means that we already have an instance of noncanonical behavior, since that single lexeme is not fully consistent according to the requirements given in Table 4 above. What space is then theoretically available for deviations based on splits? The minimal deviation would be one that is fully internal: limited to the lexeme’s morphological realization of the ‘incoming’ morphosyntactic specification. There are two dimensions for greater deviation: the split might require reference to more than purely morphological information, or it may

require reference external to the given lexeme (or both of these). These two possibilities underlie our four criteria, as Table 25 indicates.

	INTERNAL	EXTERNAL
LEXEME	1. form vs. composition/feature signature	3. irregular vs. regular
COMPONENT	2. morphomic vs. motivated	4. irrelevant vs. externally relevant

TABLE 25. The underlying logic of the four criteria.

In the simplest instance, we have a split that is purely a matter of form in a given lexeme (cell 1 in Table 25). The alternative here is that something more than simple form is involved, as discussed with relation to Table 4 above. This gives us our first criterion, the opposition between splits based solely on morphological form versus those based on the composition/feature signature of the paradigm. Furthermore, a split ‘should’ (in the canonical world) be a matter restricted to the morphological component (cell 2). If there is justification from outside morphology, this is less canonical, and so we have our second criterion, which opposes purely morphology-internal (‘morphomic’) splits to those that are motivated (they follow a boundary motivated from outside the morphological component). In the canonical situation, the split is a matter restricted to the particular lexeme—it is irregular (cell 3). If the split extends externally to other lexemes, that is less canonical: hence our third criterion opposes splits that are irregular (lexically specified) to splits that are regular (extending across the lexicon). Finally, in the canonical case, splits do not extend to affecting other components (cell 4); however, there are interesting instances where splits have syntactic consequences. Thus our fourth criterion contrasts splits that are irrelevant outside the lexeme to those that have external relevance, in that they lead to different syntactic behaviors. Our four criteria, then, do indeed cover the theoretical space. And given that we find examples that are noncanonical in respect to all of the criteria, the range of possible lexemes is considerably broader than was envisaged at the beginning of the investigation.

**5. CONCLUSION.** Lexemes may be split in numerous ways, so that what we understand as ‘possible word’ is broader and harder to pin down than we may have imagined. We first examined the various inflectional means by which paradigms can be split. Then we took a more abstract view, treating these ‘inducers’ of splits together and looking at the nature of the splits. We established four criteria that define the key ways in which lexemes may be split: form versus composition/feature signature, morphomic versus motivated, lexically specified versus regular, and internal versus externally relevant. We found good examples of each of the opposing values for these four types of split individually. Then we considered all of the combinations induced by these four types of split; this produced a substantial typological space, including some apparently bizarre possibilities, particularly those involving splits that are externally relevant. In fact, the typology is surprisingly complete; all of the possibilities are found, as is documented in the appendix. This is not the result I expected, and it is not the result that a traditional typologist would have wished for. By staking out the typological space, more widely and more accurately than is possible in traditional typology, the canonical approach gives us a picture of what could theoretically be, which proves a useful frame for understanding what we actually find.

Looking beyond splits, for some phenomena the typological space is extensive, but only a small part of it is occupied. This is what we find for systems of reported speech, when approached from a canonical typology perspective (Evans 2013). In other cases, it may seem evident that the canonical instance will not occur. This was the situation for

inflection classes, where fully parallel expression of morphosyntactic representations seems highly implausible, yet has been found in Burmeso (Corbett 2009, and §3.2 above, based on Donohue 2001). In the present project too, given the unlikely nature of some of the combinations of criteria, it seemed sure that not all of the theoretically possible lexemes could be found. And yet, remarkably, they have been. Of course, we are looking for individual lexemes rather than whole systems. Nevertheless, the result is surprising. The approach taken demonstrates that there are challenging phenomena that have been slipping through the net. The examples range from those that have some superficial oddity in form, but are clearly single lexemes, through to those whose splits are reflected in their external requirements. The latter show the sort of split that some would wish to treat as derivational, claiming we have two lexemes rather than one, and thus avoiding the embarrassment of lexemes whose external syntactic requirements vary in ways that are not easily predictable from syntactic structure.

Splits may be layered, and this can introduce substantial degrees of complexity (Stump 2006, Corbett 2015). This was the particular interest in the analysis of Slovak (§3.1), where there is a split between present and past tense, and within the past between first and second persons on the one hand and third person on the other. It was similarly striking in Chiquihuitlán Mazatec (§3.2), where there is a motivated split between aspects and a morphomic split within them.

Having established the remarkable completeness of the typology, we can envisage new lines of inquiry. First, we can investigate the frequency of such splits, both as types and as tokens. Second, while we are reasonably well informed about the ways in which such splits arise, it is worth asking how the more radical ones are maintained over time. The stability of the Kartvelian examples over many centuries is particularly noteworthy. And third, there is the difficult and underresearched area of the splits that become total and permanent, so that new lexemes develop and odd remnants are created.<sup>44</sup>

#### APPENDIX

Here I present evidence for each of the combinations of criteria, in summary form. For each combination I first give its number from Fig. 1, the lattice description, and the specification in terms of criteria. These are abbreviated as: criterion 1: form versus composition/feature signature; criterion 2: morphomic versus motivated; criterion 3: irregular versus regular; and criterion 4: internal versus (externally) relevant. The non-canonical values are given in bold. Then, where an example has already been discussed in the text, I refer back to the relevant section. Where it has not, I give the essential points and references to the key data.

TYPE 1, LATTICE DESCRIPTION C1/C2/C3/C4, SPECIFICATION: FORM, MORPHOMIC, IRREGULAR, INTERNAL. This is the most straightforward type of morphological irregularity. We saw a good instance in French *aller* ‘go’ (§1).

TYPE 2, LATTICE DESCRIPTION C1/C2/C3, SPECIFICATION: FORM, MORPHOMIC, IRREGULAR, **RELEVANT**. Being externally relevant while otherwise having the characteristics of normal morphological irregularity is an unusual combination. The example we discussed was Gaelic, where a split in the case paradigm of a noun can involve a different gender requirement (§3.4). A second instance of this unusual combination is found in Marsalese, a dialect of western Sicily (kindly brought to my attention by Anna Thornton). The data are taken from Cardinaletti & Giusti 2001, 2003, considered further in Cruschina 2013, and discussed in Corbett 2015. The construction involves just three verbs, which allow a second inflected verb, but only for some cells of the paradigm. It is this key point that makes the split externally relevant. The distribution of cells forms a morphomic pattern. We then find that we have a construction that is available provided the controlling verb stands

<sup>44</sup> See Corbett 2007a:26–27 for an instance involving suppletion. The Russian suppletive pairing *rebenok* ‘child’ and *deti* ‘children’ has left remnants such as *rebjata* ‘chaps, guys’, where the morphological link to *rebenok* ‘child’ remains, but the semantic relation is no longer close. Arguably, the effect is easier to spot with suppletion than elsewhere, since the form-meaning mismatches are clear; by contrast, the relative infrequency of examples of suppletion makes statistical investigation difficult.

in a form from one segment of the morphomic pattern, but not if the controlling verb is in a form from the other segment.

TYPE 3, LATTICE DESCRIPTION C1/C2/C4, SPECIFICATION: FORM, MORPHOMIC, REGULAR, INTERNAL. This is a specification discussed in §3.2. We saw how Burmese verbs have a pattern of forms (going across the inflectional classes), which is morphomic. The split is regular, applying to all verbs, and it has no external significance.

TYPE 4, LATTICE DESCRIPTION C1/C3/C4, SPECIFICATION: FORM, MOTIVATED, IRREGULAR, INTERNAL. This type can be illustrated by Slovak *oko* ‘eye’, which, like the Serbo-Croat noun illustrated in Table 19 and examples 6–7, has an irregular alternation in the plural (*oči* ‘eyes’). The difference is that the gender agreement forms are more restricted in Slovak: in the plural, for inanimate nouns, there is no distinction between the genders. Hence there can be no external relevance for the split between the singular and plural segments here. A more subtle example is the Russian verb *xotet’* ‘want’, which in the present tense has a split between singular and plural. This sounds unremarkable, but it is highly irregular, since this is not where Russian verbs divide (Corbett & Baerman 2010).

TYPE 5, LATTICE DESCRIPTION C2/C3/C4, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MORPHOMIC, IRREGULAR, INTERNAL. This specification matches the Archi agreeing pronoun, discussed in both §3.2 and §3.3. The split is based on the feature signature; a small number of cells are sensitive to incoming gender and number features, while the large majority are not. The cells involved form a morphomic pattern, and the split is irregular (even within the pronouns) and has no external relevance.

A second example, this one involving composition, comes from a Romanian dialect and is due to Maiden (2004:240–44), who reports on the Transylvanian linguistic atlas (*Noul Atlas Lingistic Român pe Regiuni. Transilvania*, map 1907, unpublished). The map of interest to us gives information on the verb ‘go’. This verb resulted from the alliance of *a merge* ‘to go’ (*a* is the infinitive marker) and *a se duce* ‘to go, betake oneself’, usually used of an animate; this latter is the reflexive form of *a duce* ‘to lead, take’. The present-tense forms in the relevant dialect at Fundătura (point 274 on the atlas) are given in Table A1 (also discussed in Corbett 2007a:30–31).

	SINGULAR	PLURAL
1	mə duk	'merem
2	t'e duc	'merets
3	sə 'duce	sə duk

TABLE A1. Romanian dialect ‘go’, present tense (Maiden 2004:242).

Here one suppletive stem involves periphrastic forms, while the other does not, which means that we are dealing with a split in terms of the composition of the cells; the split follows a well-known morphomic pattern of Romance, and it is irregular and internal.

TYPE 6, LATTICE DESCRIPTION C1/C2, SPECIFICATION: FORM, MORPHOMIC, REGULAR, RELEVANT. This remarkable specification is that of the Georgian case government, as discussed fully in §3.4.

TYPE 7, LATTICE DESCRIPTION C1/C3, SPECIFICATION: FORM, MOTIVATED, IRREGULAR, RELEVANT. The straightforward example we examined was Serbo-Croat *oko* ‘eye’, which has a split according to form, along a number division (hence motivated); it involves a very few nouns, and the split brings with it different gender values (§3.4). The Nordreisa *mamma* ‘mum’ (§3.5) also fits here, though it has the additional complication that the agreements it controls are not consistent. A further example (also with agreements that are not consistent) is Serbo-Croat *deca* ‘children’ (Corbett 2007b). Both of the latter are subject to the requirements of the agreement hierarchy.

TYPE 8, LATTICE DESCRIPTION C1/C4, SPECIFICATION: FORM, MOTIVATED, REGULAR, INTERNAL. This specification is found in Serbo-Croat nouns of the first inflection class (Browne 1993:319–20). The normal nouns of this class, like *prozor* ‘window’, have a set of plural inflections that are comparable to those of the singular (for example, NOM.PL *prozor-i*, ACC.PL *prozor-e*). However, monosyllabic nouns of this type have a split in their paradigm: in the plural they additionally have a stem augment. Thus from *grad* ‘city’ we have NOM.PL *grad-ov-i*, ACC.PL *grad-ov-e*, and so on. This split involves form and it is motivated (it is based on number). It is moderately regular: there are some exceptions in each direction (monosyllabics without *-ov-* and disyllabics with *-ov-*). For a study drawn from a substantial set of texts see Nikolić 2013. The split has no external relevance, since the plural forms take plural agreements, whether or not there is an augment.

This particular split is also of interest in terms of phonological predictability in morphology. First recall that there is a clear morphological requirement: the *-ov-* augment is available within one inflectional class. Then there is a strong but not absolute prediction from phonology: monosyllabic stems (within inflection

class 1) mostly take the augment; longer nouns mostly do not. However, this does not satisfy any general syllabic template: the resulting plural stem has an extra syllable, but the oblique plural inflections have in any case an additional syllable in comparison with any singular inflection. The result is: the nominative singular is monosyllabic; the oblique singular forms are bisyllabic; nominative, accusative, and genitive plural have three syllables; and the remaining plural forms have four.

TYPE 9, LATTICE DESCRIPTION C2/C3, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MORPHOMIC, IRREGULAR, RELEVANT. A more severe type of structural split occurs in lexemes that are defective—that is, they lack particular forms. An example that has been well discussed is the first singular of certain Russian verbs; see Baerman 2008, and more generally on defectiveness see Baerman et al. 2010. Another instance, at least for many speakers, is the Russian noun *mečta* ‘dream’ given in Table A2; see Mel’čuk 1993:360–61, 1996:176–77.

	<i>mečta</i> ‘dream’		<i>mačta</i> ‘mast’	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOMINATIVE	mečta	mečty	mačta	mačty
ACCUSATIVE	mečtu	mečty	mačtu	mačty
GENITIVE	mečty	***	mačty	mačt
DATIVE	mečte	mečtam	mačte	mačtam
INSTRUMENTAL	mečtoj	mečtami	mačtoj	mačtami
LOCATIVE	mečte	mečtax	mačte	mačtax

TABLE A2. Russian: the defective noun *mečta* ‘dream’, with *mačta* ‘mast’ for comparison.

This item is split, with the genitive plural having no form, as compared to the normal forms of the rest of the paradigm (see *mačta* ‘mast’ for comparison). The split is morphomic and irregular. It is also relevant in an obvious way, because any syntactic structure requiring the missing form is simply excluded.

While I have treated the split as morphomic, it might be argued that it is hard to determine whether a single cell can or cannot form a motivated segment. The solution is to look at the complementary set of cells; in this instance, the complementary set is all of the singular cells and the nominative, accusative, dative, instrumental, and locative plural—a segment that clearly has no motivation. Conversely, in a two-member paradigm, with, say, singular and plural cells, each would be a motivated segment. For clarity, wherever possible I give larger paradigms.

A second example is more curious. Consider the paradigm of Russian *tysjača* ‘thousand’ in Table A3.

	SINGULAR	PLURAL
NOMINATIVE	tysjača	tysjači
ACCUSATIVE	tysjaču	tysjači
GENITIVE	tysjači	tysjač
DATIVE	tysjače	tysjačam
INSTRUMENTAL	tysjačej/tysjač’ju	tysjačami
LOCATIVE	tysjače	tysjačax

TABLE A3. Russian *tysjača* ‘thousand’.

This item has all the forms that would be expected of a noun of inflection class II; indeed, apart from predictable adjustments, it is like *mačta* ‘mast’ in Table A2. But in addition it has a second form in the instrumental singular; this cell shows overabundance (Thornton 2010–2011, 2011a,b).<sup>45</sup> The problem thus involves the composition of the cell; the split is morphomic (there is no motivation for singling out the instrumental singular) and it is completely irregular, being unique to this lexical item. The key question is whether there is any external relevance. It has been claimed that the two forms have different government requirements, which would mean that the split has external relevance. The *tysjačej* form is like that of a noun, while the *tysjač’ju* form is comparable to that of several numerals. Russian nouns typically take a second noun in the genitive, while numerals in the oblique cases, including the instrumental, take a noun in the same case (see Corbett

<sup>45</sup> There is even a third CELL MATE (to use Michele Loporcaro’s felicitous term, adopted in Thornton 2011a), since there is a general option for items of this inflection class to have a longer variant *tysjačeju*; this is unusual, but is still found (there are nine instances of the longer variant in the Russian National Corpus, all governing a genitive plural). The Russian National Corpus was accessed on 14 March 2012. Note, however, that Thornton (2011b) does not treat overabundance in this way; she argues that it requires an additional dimension, beyond those postulated in this article.

1993 and references there for the details). It has been claimed that in the more general quantificational meaning the noun-like form is used, together with the genitive plural, while in the exact numerical reading the numeral-like form is used, together with the instrumental plural, as in these examples.

- (A1) Russian (Ivanova 1969)  
 obremennyj tysjač-ej zabor  
 burdened thousand-SG.INS care[PL.GEN]  
 ‘burdened with a thousand cares’
- (A2) s tysjač’-ju rublj-ami  
 with thousand-SG.INS rouble-PL.INS  
 ‘with a thousand roubles’

Ivanova (1969) points out that the normative picture is not reflected in texts. Besides examples like A1 and A2, there are also examples like A3, which is comparable to A1 except for the form of ‘thousand’.

- (A3) obremennyj tysjač’-ju zabor  
 burdened thousand-SG.INS care[PL.GEN]  
 ‘burdened with a thousand cares’

However, if these three variants exist, demonstrating that the neat normative account is unrealistic, we still have an interesting situation. If the fourth variant is excluded (that is, *tysjačej* plus noun in the instrumental), we have evidence for syntactic relevance: the two forms would have different government possibilities. Examples are not frequent, but all of the evidence I have pieced together supports this conclusion. First, Ivanova (1969) reports on various sources, and finds no evidence for the fourth, apparently excluded, combination. Next, there is the corpus study of the language of the press, by Graudina and colleagues (1976:264–66). In their corpus of approximately two million running words, there were just nine relevant examples; the reporting is not fully clear, but they give none of the missing type. I therefore turned to the Russian National Corpus, selecting material just from the twentieth century (searched 14 March 2012). There were numerous examples with a noun in the genitive plural, that is, examples both like A1 and like A3. I found just four instances with a noun in the instrumental plural (examples like A2). I found no instances of the apparently excluded fourth possibility (*tysjačej* plus noun in the instrumental plural). A limited amount of work with consultants also suggested that this fourth possibility is not acceptable. Hence the two forms *tysjačej* and *tysjač’ju* do indeed have different syntactic requirements. The first takes the genitive, and the second takes the genitive or (rarely) the instrumental.

TYPE 10, LATTICE DESCRIPTION C2/C4, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MORPHOMIC, REGULAR, INTERNAL. The example noted in the lattice for this specification was the allative in Lak. Like other Daghestanian languages, Lak has a complex system of spatial case values; they are built out of a set of markers for localization, to which a marker for orientation can be added (see Daniel & Ganenkov 2009 for an overview). One of the orientation markers is the allative, which is built on the lative and indicates motion toward. This particular case value, and no other, brings with it an agreement marker (for agreement with the absolutive argument in the clause).

- (A4) Lak (Kibrik 2003a:54)  
 qāī-lu-wu-n-m-aj  
 house-OBL-IN-LATIVE-SG.III-ALLATIVE  
 ‘into the house’

Here the *-m-* is appropriate for agreement with an absolutive argument that is gender III singular. Thus, within the large paradigm of the Lak noun, just the allative case shows agreement in gender and number. For the full system of spatial case values in Lak, see Žirkov 1955:36–38. The split is regular, since it occurs generally for nouns.

A comparable specification, arguably, applies to the unusual reflection of person in the verb in Archi (for which see Corbett 2013b). The verb’s paradigm is split so that there is a cell for the first and second persons in the plural (signaled by an unusual syncretism), while elsewhere there is agreement according to gender and number. This split, then, involves the feature signature, it is morphomic, it is nevertheless regular, and it has no external relevance.

TYPE 11, LATTICE DESCRIPTION C3/C4, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MOTIVATED, IRREGULAR, INTERNAL. We have already met two items that meet this specification. The first is the Krongo noun *mùsì* ‘sorcerer’, presented in 1 above, from Reh 1985:117. The composition of the cells, in terms of the number of affixes and their order, is dramatically different in the singular and the plural. This is a motivated split (singular versus plural), it is highly irregular, and there is no indication of any external relevance. Similarly, the example of antiperiphrasis cited in §2.3 fits here. The Russian verb *byť* ‘be’ has a synthetic imperfective future (*budu* ‘I will be’, and so on), while all other verbs with an imperfective future have a periphrastic form. Thus

the composition of the imperfective future is out of step with all others in terms of composition; this is a motivated split, irregular, and internal.

TYPE 12, LATTICE DESCRIPTION C1, SPECIFICATION: FORM, MOTIVATED, REGULAR, RELEVANT. This is the specification of Common Kartvelian, as discussed in detail in §4.2.

TYPE 13, LATTICE DESCRIPTION C2, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MORPHOMIC, REGULAR, RELEVANT. Kayardild has a large inventory of case values; there is a split between ‘normal’ cases and ‘verbalizing’ cases (earlier termed ‘verbal cases’; Evans 2003:214, n. 13), which are the last seven in Table A4.

NOMINATIVE	mala-a
LOCATIVE	mala-ya
ABLATIVE	mala-na
PROPRIETIVE	mala-wuru
OBLIQUE	mala-ntha
ALLATIVE	mala-r
GENITIVE	mala-karra
ASSOCIATIVE	mala-nurru
ORIGIN	mala-wan- (always with further inflection)
PRIVATIVE	mala-warri
CONSEQUENTIAL	mala-ngarrba
INSTRUMENTAL	mala-nguni
UTILATIVE	mala-marra
VERBALIZING ALLATIVE	mala-yiiwa-tha
VERBALIZING DATIVE	mala-maru-tha
VERBALIZING TRANSLATIVE	mala-marii-ja
VERBALIZING ABLATIVE	TRANSITIVE mala-wula-tha
	INTRANSITIVE mala-wula-a-ja
VERBALIZING EVITATIVE	TRANSITIVE mala-waalu-tha
	INTRANSITIVE mala-waal-i-ja
VERBALIZING DONATIVE	mala-wu-ja
VERBALIZING PURPOSIVE	mala-jani(i)-ja

TABLE A4. Kayardild *mala-* ‘sea’ (Evans 1995:125, 165, and p.c.; see also Baerman 2006, Erich Round, p.c.).

There is a clear difference in feature signature; the verbalizing cases mark tense, aspect, mood, and polarity, as in this example.

- (A5) Kayardild (Evans 2003:215)  
 ngada waa-jarra wangarr-ina ngijin-maru-tharra thabuju-maru-tharra  
 1SG.NOM sing-PST song-MOD\_ABL my-V\_DAT-PST brother-V\_DAT-PST  
 ‘I sang a song for my brother.’
- (A6) ngada waa-nangku wangarr-u ngijin-maru-nangku thabuju-maru-nangku.  
 1SG.NOM sing-NEG.POT song-MOD\_PROPRIETIVE my-V\_DAT-NEG.POT brother-V\_DAT-NEG.POT  
 ‘I won’t sing a song for my brother.’

Here the items marked with the verbalizing dative case also have a tense marker, determined by the tense of the verb in their clause. See Evans 1995, 2003 and Round 2009:433–73, 2013:65–86, 177–201 for more information on this remarkable phenomenon. The normal cases behave in a more familiar way. They allow case stacking, as in A7, but they do not take tense, aspect, mood, or polarity.<sup>46</sup>

- (A7) Kayardild (Evans 2003:207)  
 dan-kinaba-nguni dangka-naba-nguni mirra-nguni wanggal-nguni  
 this-ABL-INS man-ABL-INS good-INS boomerang-INS  
 ‘with this man’s good boomerang’

The split in feature signature between the normal and the verbalizing case values in the paradigm is morphomic: there is no motivation for the split. However, it extends right across the lexicon: it is fully regular.<sup>47</sup>

<sup>46</sup> The possessor is marked with the ablative in this example. The marker is *-naba-* here, as opposed to *-na* in example A5, because the *-ba* can be lost in word-final position (Evans 1995:125).

<sup>47</sup> For the similarly interesting phenomenon of the split in tense-aspect-mood marking, which is associated with both the specification of the matrix clause and that of the embedded VP, see Round 2013:99–104.

The key remaining question is whether this split is externally relevant. It turns out that it is. There is an effect of the verbal origin of the verbalizing case values, which does not apply to the normal ones. Phrases marked with a verbalizing case value have to show consistent transitivity throughout the clause: for example, a transitive verb is consistent with the verbalizing dative, verbalizing transitive ablative, and so on, whereas with an intransitive verb we find the verbalizing intransitive allative and verbalizing intransitive ablative, and so on (Evans 1995:178–80, and p.c.). While the verbalizing dative is only transitive, and the verbalizing allative is only intransitive, the verbalizing ablative and evitative have forms for both. There is no such constraint of consistent transitivity with the normal case values; hence the split is indeed externally relevant.

TYPE 14, LATTICE DESCRIPTION C3, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MOTIVATED, IRREGULAR, RELEVANT. This seems an unlikely combination; I have one (arguable) example. The Russian verb *byt'* 'be' in the present tense is null. This is a complete null (not a zero stem to which affixes can be attached); thus it represents fused exponence (Corbett 2007a:16–17). This is clearly a matter of the composition of the cells. It is a motivated split, affecting just the present tense. And it is highly irregular, being the only example in the language. Could such a split be relevant? It turns out that it can. Russian copula verbs take complements in the nominative or in the instrumental, according to several competing conditions (see Krasovitsky et al. 2008 for examples and references). We gave these examples in that article.

- (A8) Russian: case possibilities with the copula (Krasovitsky et al. 2008)
- |                    |                |                    |               |
|--------------------|----------------|--------------------|---------------|
| a. On by-I         | vrač.          | b. On by-I         | vrač-om.      |
| he be-PST[SG.M]    | doctor[SG.NOM] | he be-PST[SG.M]    | doctor-SG.INS |
| 'He was a doctor.' |                | 'He was a doctor.' |               |

With the null form of the present tense, however, only the nominative is found.

- (A9) a. On [ ]
- |                   |                |                   |               |
|-------------------|----------------|-------------------|---------------|
| vrač.             | b. *On [ ]     | vrač-om.          |               |
| he [be.PRS]       | doctor[SG.NOM] | he [be.PRS]       | doctor-SG.INS |
| 'He is a doctor.' |                | 'He is a doctor.' |               |

This is not a condition on case marking determined by tense. We see this if we compare the verb *javljat'sja*, also meaning 'be', and found more in written use, particularly in nonfictional prose. This verb normally takes the instrumental case, including in the present tense. Hence the difference between A8b and A9b is due to the split in the paradigm of Russian *byt'* 'be', and so this split has external relevance.<sup>48</sup>

TYPE 15, LATTICE DESCRIPTION C4, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MOTIVATED, REGULAR, INTERNAL. Examples of this type are not rare. The Sanskrit future, discussed in §3.1, is one.<sup>49</sup> The split runs along the line of tense, and is motivated.

TYPE 16, LATTICE DESCRIPTION {}, SPECIFICATION: COMPOSITION/FEATURE SIGNATURE, MOTIVATED, REGULAR, RELEVANT. We discussed this specification when we examined the Archi biabsolutive construction in §4.2. It is a good point on which to end the summary of the data confirming our typology, since the morphology involved is in many respects as far from canonical morphology as is possible.

## REFERENCES

- ACKERMAN, FARRELL; JAMES P. BLEVINS; and ROBERT MALOUF. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. *Analogy in grammar: Form and acquisition*, ed. by James P. Blevins and Juliette Blevins, 54–82. Oxford: Oxford University Press.

<sup>48</sup> For more subtle differences between the present of *byt'* 'be' and other forms, see Letučij 2013.

<sup>49</sup> The Russian past tense (§1) is another. There is an interesting wrinkle here, however. The inclusion or omission of personal pronouns in Russian is variable and complex: it would be hard to characterize Russian either as a pro-drop language or as one that retains pronouns. We might then expect the Russian tense split to have some impact on the use of pronouns, since different features are overt in the past (number and gender) and nonpast (number and person). According to McShane (2009:126), the proportion of dropped pronouns in the first and second persons is not affected by the difference in feature signature between the past and nonpast tenses. Kibrik (2013:240–42) reviews evidence on all three persons, however, and concludes that tense is one factor among several influencing the use of pronouns (somewhat more are found in the past than in the non-past). This is best analyzed as an indirect influence: there is no structural difference involved, but the (non)availability of person and gender information has some influence on pronominal inclusion or omission. Thus this example could be seen as not absolutely canonical in terms of the relevance criterion.



- ADAMO, EVANGELIA. 2012. Verb morphologies in contact: Evidence from the Balkan area. *Morphologies in contact*, ed. by Martine Vanhove, Thomas Stolz, Aina Urdze, and Hitomi Otsuka, 143–62. Berlin: Akademie.
- ADGER, DAVID. 2009. Gaelic: A language tutorial. Presented at the annual conference of the Linguistics Association of Great Britain, University of Edinburgh, 6–9 September 2009. Online: <http://www.lel.ed.ac.uk/lagb2009/>.
- ALBRIGHT, ADAM, and ERIC FUB. 2012. Syncretism. In Trommer, 236–88.
- ANDERSON, STEPHEN R. 1984. On representations in morphology: Case, agreement and inversion in Georgian. *Natural Language and Linguistic Theory* 2.157–218.
- ANDERSON, STEPHEN R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- ANDERSON, STEPHEN R. 2011. Stress-conditioned allomorphy in Surmiran (Rumantsch). In Maiden et al., 8–23.
- ANDERSON, STEPHEN R. 2013. Stem alternations in Swiss Rumantsch. In Cruschina et al., 262–83.
- ARONOFF, MARK. 1994. *Morphology by itself: Stems and inflectional classes*. (Linguistic Inquiry monographs 22.) Cambridge, MA: MIT Press.
- ARONOFF, MARK. 2012. Morphological stems: What William of Ockham really said. In Bonami 2012b, 28–51.
- ARONSON, HOWARD I. 1982. *Georgian: A reading grammar*. Columbus, OH: Slavica.
- ARONSON, HOWARD I. 1991. Modern Georgian. *The Indigenous languages of the Caucasus, vol. 1: The Kartvelian languages*, ed. by Alice C. Harris, 219–312. Delmar: Caravan Books.
- ASUDEH, ASH; MARY DALRYMPLE; and IDA TOIVONEN. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1.1–54.
- BAERMAN, MATTHEW. 2006. Kayardild. *Typological database on deponency*, by Greville G. Corbett, Matthew Baerman, Dunstan Brown, and Andrew Hippisley. Online: <http://www.smg.surrey.ac.uk/deponency/>. Specific page: <http://www.smg.surrey.ac.uk/Deponency/WALS/Kayardild.htm#>, accessed 25 April 2012.
- BAERMAN, MATTHEW. 2008. Historical observations on defectiveness: The first singular non-past. *Russian Linguistics* 32.81–97.
- BAERMAN, MATTHEW. 2013. Inflection class interactions. *Morphology in Toulouse: Selected proceedings of Décembrettes 7 (Toulouse, 2–3 December 2010)*, ed. by Nabil Hathout, Fabio Montermini, and Jesse Tseng, 11–33. Munich: LINCOM Europa.
- BAERMAN, MATTHEW. 2014. Covert systematicity in a distributionally complex system. *Journal of Linguistics* 50.1–47.
- BAERMAN, MATTHEW; DUNSTAN BROWN; and GREVILLE G. CORBETT. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.
- BAERMAN, MATTHEW; DUNSTAN BROWN; and GREVILLE G. CORBETT (eds.) 2015. *Understanding and measuring morphological complexity*. Oxford: Oxford University Press, to appear.
- BAERMAN, MATTHEW, and GREVILLE G. CORBETT. 2012. Stem alternations and multiple exponence. In Bonami 2012b, 52–68.
- BAERMAN, MATTHEW; GREVILLE G. CORBETT; and DUNSTAN BROWN (eds.) 2010. *Defective paradigms: Missing forms and what they tell us*. (Proceedings of the British Academy 163.) Oxford: British Academy and Oxford University Press.
- BAERMAN, MATTHEW; GREVILLE G. CORBETT; DUNSTAN BROWN; and ANDREW HIPPISEY (eds.) 2007. *Deponency and morphological mismatches*. (Proceedings of the British Academy 145.) Oxford: British Academy and Oxford University Press.
- BLEVINS, JAMES P. 2003. Stems and paradigms. *Language* 79.737–67.
- BLIESE, LOREN F. 1981. *A generative grammar of Afar*. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.
- BOBALJK, JONATHAN. 2002. Syncretism without paradigms: Remarks on Williams 1981, 1994. *Yearbook of Morphology* 2001.53–85.
- BONAMI, OLIVIER. 2012a. Introduction. In Bonami 2012b, 1–6.
- BONAMI, OLIVIER (ed.) 2012b. *Stems*. (Special issue of *Word Structure* 5.1.) Edinburgh: Edinburgh University Press.

- BONAMI, OLIVIER, and GILLES BOYÉ. 2007. French pronominal clitics and the design of paradigm function morphology. *Online proceedings of the Mediterranean Morphology Meeting* 5.291–322. Online: [http://www3.lingue.unibo.it/mmm2/?page\\_id=71](http://www3.lingue.unibo.it/mmm2/?page_id=71).
- BONAMI, OLIVIER, and GILLES BOYÉ. 2008. Paradigm shape is morphomic in Nepali. Paper presented at the 13th International Morphology Meeting, Vienna, February 2008.
- BOND, OLIVER. 2013. A base for canonical negation. In Brown et al. 2013, 20–47.
- BOND, OLIVER. 2016. Negation through reduplication and tone: Implications for the LFG/PFM interface. *Journal of Linguistics*, to appear.
- BONET, EULÀLIA, and DANIEL HARBOUR. 2012. Contextual allomorphy. In Trommer, 195–235.
- BORETZKY, NORBERT. 1993. *Bugurdži: Deskriptiver und historischer Abriss eines Romani-Dialekts*. (Balkanologische Veröffentlichungen 21.) Berlin: Harrassowitz.
- BROWN, DUNSTAN, and MARINA CHUMAKINA. 2013. What there might be and what there is: An introduction to canonical typology. In Brown et al. 2013, 1–19.
- BROWN, DUNSTAN; MARINA CHUMAKINA; and GREVILLE G. CORBETT (eds.) 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- BROWN, DUNSTAN; MARINA CHUMAKINA; GREVILLE CORBETT; GERGANA POPOVA; and ANDREW SPENCER. 2012. Defining ‘periphrasis’: Key notions. *Morphology* 22.233–75.
- BROWN, DUNSTAN; GREVILLE G. CORBETT; and CAROLE TIBERIUS (eds.) 2003. *Agreement: A typological perspective*. (Special issue of *Transactions of the Philological Society* 101.2.) Oxford: Blackwell.
- BROWN, DUNSTAN, and ANDREW HIPPISEY. 2012. *Network morphology*. Cambridge: Cambridge University Press.
- BROWNE, WAYLES. 1993. Serbo-Croat. In Comrie & Corbett, 306–87.
- CARDINALETTI, ANNA, and GIULIANA GIUSTI. 2001. ‘Semi-lexical’ motion verbs in Romance and Germanic. *Semi-lexical categories: The function of content words and the content of function words* (Studies in generative grammar 59), ed. by Norbert Corver and Henk van Riemsdijk, 371–414. Berlin: Mouton de Gruyter.
- CARDINALETTI, ANNA, and GIULIANA GIUSTI. 2003. Motion verbs as functional heads. *The syntax of Italian dialects* (Oxford studies in comparative syntax), ed. by Christina Tortora, 31–49. Oxford: Oxford University Press.
- CARSTAIRS-McCARTHY, ANDREW. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70.737–88.
- CECH, PETRA, and MOZES F. HEINSCHNINK. 1999. *Sepečides-Romani: Grammatik, Texte und Glossar eines türkischen Romani-Dialekts*. (Balkanologische Veröffentlichungen 34.) Wiesbaden: Harrassowitz.
- CHIU, BONNIE HUI-CHUN. 1987. The morphology of the Alabama set-I affixes. *Muskogean linguistics* (UCLA occasional papers in linguistics 6), ed. by Pamela Munro, 21–35. Los Angeles: University of California, Los Angeles.
- CHUMAKINA, MARINA, and GREVILLE G. CORBETT. 2008. Archi: The challenge of an extreme agreement system. *Fonetika i nefonetika: K 70-letiju Sandro V. Kodzasova*, ed. by Aleksandr V. Arxipov, Leonid M. Zaxarov, Andrej A. Kibrik, Aleksandr E. Kibrik, Irina M. Kobozeva, Ol’ga F. Krivnova, Ekaterina A. Ljutikova, and Ol’ga V. Fëdorova, 184–94. Moscow: Jazyki slavjanskix kul’tur.
- CHUMAKINA, MARINA, and GREVILLE G. CORBETT (eds.) 2013. *Periphrasis: The role of syntax and morphology in paradigms*. (Proceedings of the British Academy 180.) Oxford: British Academy and Oxford University Press.
- CHUMAKINA, MARINA, and GREVILLE G. CORBETT. 2015. Gender-number marking in Archi: Small is complex. In Baerman et al. 2015, to appear.
- COMRIE, BERNARD. 2003. When agreement gets trigger-happy. In Brown et al. 2003, 313–37.
- COMRIE, BERNARD, and GREVILLE G. CORBETT (eds.) 1993. *The Slavonic languages*. London: Routledge.
- CORBETT, GREVILLE G. 1993. The head of Russian numeral expressions. *Heads in grammatical theory*, ed. by Greville G. Corbett, Norman M. Fraser, and Scott McGlashan, 11–35. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 1999. Defectiveness, syncretism, suppletion, ‘deponency’: Four dimensions for a typology of inflectional systems. Paper presented at the 2nd Mediterranean Morphology Meeting, Malta, September 1999.

- CORBETT, GREVILLE G. 2003. Agreement: Canonical instances and the extent of the phenomenon. *Topics in morphology: Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20–22, 2001)*, ed. by Geert Booij, Janet DeCesaris, Angela Ralli, and Sergio Scalise, 109–28. Barcelona: Universitat Pompeu Fabra. Online: [http://www3.surrey.ac.uk/LIS/SMG/corbett\\_publications.htm](http://www3.surrey.ac.uk/LIS/SMG/corbett_publications.htm).
- CORBETT, GREVILLE G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 2007a. Canonical typology, suppletion, and possible words. *Language* 83.8–42.
- CORBETT, GREVILLE G. 2007b. Deponency, syncretism and what lies between. In Baerman et al. 2007, 21–43.
- CORBETT, GREVILLE G. 2008. Determining morphosyntactic feature values: The case of case. *Case and grammatical relations: Papers in honor of Bernard Comrie*, ed. by Greville G. Corbett and Michael Noonan, 1–34. Oxford: Oxford University Press.
- CORBETT, GREVILLE G. 2009. Canonical inflectional classes. *Selected proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11. Online: <http://www.lingref.com/cpp/decemb/6/abstract2231.html>.
- CORBETT, GREVILLE G. 2010. Classic problems at the syntax-morphology interface: Whose are they? *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar (HPSG 2010)*, 255–68. Online: <http://csli-publications.stanford.edu/HPSG/2010/>.
- CORBETT, GREVILLE G. 2011. The penumbra of morphosyntactic feature systems. *Morphology* (Special issue: *Markedness and underspecification in the morphology and semantics of agreement*, ed. by Jonathan Bobaljik, Uli Sauerland, and Andrew Nevins) 21.445–80.
- CORBETT, GREVILLE G. 2012. *Features*. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 2013a. Canonical morphosyntactic features. In Brown et al. 2013, 48–65.
- CORBETT, GREVILLE G. 2013b. The unique challenge of the Archi paradigm. *Berkeley Linguistics Society (Special session on languages of the Caucasus)* 37.2.52–67. [Dated 2011 but published April 2013.] Online: <http://elanguage.net/journals/bls/article/view/3626>.
- CORBETT, GREVILLE G. 2015. Morphomic splits. *The morphome debate: Diagnosing and analysing morphomic patterns*, ed. by Ana R. Luis and Ricardo Bermúdez-Otero. Oxford: Oxford University Press, to appear.
- CORBETT, GREVILLE G., and MATTHEW BAERMAN. 2010. ‘Deep irregularity’: When the expected lexical splits follow the wrong pattern. Paper presented at the 14th International Morphology Meeting, Budapest, May 2010.
- CORBETT, GREVILLE G., and WAYLES BROWNE. 2009. Serbo-Croat: Bosnian, Croatian, Montenegrin, Serbian (Ch. 18). *The world’s major languages*, 2nd edn., ed. by Bernard Comrie, 330–46. London: Routledge.
- CORBETT, GREVILLE G., and NORMAN M. FRASER. 1993. Network morphology: A DATR account of Russian inflectional morphology. *Journal of Linguistics* 29.113–42. [Reprinted in *Morphology: Critical concepts in linguistics*, vol. 6. *Morphology: Its place in the wider context*, ed. by Francis X. Katamba, 364–96. London: Routledge, 2003.]
- CORBETT, GREVILLE G., and NORMAN M. FRASER. 1997. Vyčislitel’naja lingvistika i tipologija. *Vestnik MGU: Serija 9, Filologija* 2, 122–40.
- CRUSCHINA, SILVIO. 2013. Beyond the stem and inflectional morphology: An irregular pattern at the level of periphrasis. In Cruschina et al., 262–83.
- CRUSCHINA, SILVIO, MARTIN MAIDEN; and JOHN CHARLES SMITH (eds.) 2013. *The boundaries of pure morphology: Diachronic and synchronic perspectives*. Oxford: Oxford University Press.
- CYSOUW, MICHAEL. 2011. The expression of person and number: A typologist’s perspective. *Morphology* (Special issue: *Markedness and underspecification in the morphology and semantics of agreement*, ed. by Jonathan Bobaljik, Uli Sauerland, and Andrew Nevins) 21.419–43.
- DANIEL, MICHAEL, and DMITRI GANENKOV. 2009. Case marking in Daghestanian: Limits of elaboration. *The Oxford handbook of case*, ed. by Andrej Malchukov and Andrew Spencer, 668–85. Oxford: Oxford University Press.

- DE LACY, PAUL, and JOHN KINGSTON. 2013. Synchronic explanation. *Natural Language and Linguistic Theory* 31.287–355.
- DIAMOND, JARED. 1998. *Guns, germs and steel: A short history of everybody for the last 13,000 years*. London: Vintage.
- DONOHUE, MARK. 2001. Animacy, class and gender in Burmeso. *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific linguistics 514), ed. by Andrew Pawley, Malcolm Ross, and Darrell Tryon, 97–115. Canberra: Australian National University.
- DOWNING, LAURA J.; T. ALAN HALL; and RENATE RAFFELSIEFEN (eds.) 2005. *Paradigms in phonological theory*. Oxford: Oxford University Press.
- DRESSLER, WOLFGANG U. 1985. On the predictiveness of natural morphology. *Journal of Linguistics* 21.321–37.
- DWELLY, EDWARD. 1902–1911. *Faclair Gaidhlig: A Gaelic dictionary: Specially designed for beginners and for use in schools*. Herne Bay: E. MacDonald and Co. at the Gaelic Press. Online: <http://www.dwelly.info>, accessed 25 April 2012.
- ELŠÍK, VIKTOR. 2000. Romani nominal paradigms: Their structure, diversity, and development. *Grammatical relations in Romani: The noun phrase* (Current issues in linguistic theory 211), ed. by Viktor Elšík and Yaron Matras, 9–30. Amsterdam: John Benjamins.
- EMBICK, DAVID, and MORRIS HALLE. 2005. On the status of stems in morphological theory. *Romance languages and linguistic theory 2003: Selected papers from 'Going Romance' 2003, Nijmegen, 20–22 November*, ed. by Twan Geerts, Ivo van Ginneken, and Haike Jacobs, 37–62. Amsterdam: John Benjamins.
- ENGER, HANS-OLAV, and GREVILLE G. CORBETT. 2012. Definiteness, gender and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics* 24.287–324.
- EVANS, NICHOLAS. 1995. *A grammar of Kayardild: With historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.
- EVANS, NICHOLAS. 2003. Typologies of agreement: Some problems from Kayardild. In Brown et al. 2003, 203–34.
- EVANS, NICHOLAS. 2013. Some problems in the typology of quotation: A canonical approach. In Brown et al. 2013, 66–98.
- EVANS, NICHOLAS; DUNSTAN BROWN; and GREVILLE G. CORBETT. 2001. Dalabon pronominal prefixes and the typology of syncretism: A network morphology analysis. *Yearbook of Morphology* 2000.187–231.
- EVANS, ROGER, and GERALD GAZDAR. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22.167–216.
- FINLEY, SARA. 2009. Morphemic harmony as featural correspondence. *Lingua* 119.478–501.
- FLEMMING, EDWARD. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.7–44.
- FOLEY, WILLIAM A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29.357–404.
- FORKER, DIANA. 2012. The bi-absolutive construction in Nakh-Daghestanian. *Folia Linguistica* 46.75–108.
- FORTIN, ANTONIO. 2011. *The morphology and semantics of expressive affixes*. Oxford: University of Oxford dissertation.
- FRIEDMAN, VICTOR A. 1993. Macedonian. In Comrie & Corbett, 249–305.
- GAGLIARDIA, ANNIE; MICHAEL GONCALVES; MARIA POLINSKY; and NINA RADKEVICH. 2013. The Nakh-Daghestian biabsolutive: Unity and diversity. Paper presented at the 10th biennial meeting of the Association for Linguistic Typology, Leipzig, August 2013.
- GRAUDINA, LJUDMILA K.; VIKTOR A. ICKOVIČ; and LIJA P. KATLINSKAJA. 1976. *Grammatičeskaja pravil'nost' russkoj reči: Opyt častotno-stilističeskogo slovarja variantov*. Moscow: Nauka.
- HARLEY, HEIDI, and ELIZABETH RITTER. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78.482–526.
- HARRIS, ALICE C. 1981. *Georgian syntax: A study in relational grammar*. Cambridge: Cambridge University Press.
- HARRIS, ALICE C. 1985. *Syntax and semantics, vol. 18: Diachronic syntax: The Kartvelian case*. New York: Academic Press.

- HARRIS, ALICE C. 2008. On the explanation of typologically unusual structures. *Linguistic universals and language change*, ed. by Jeff Good, 54–78. Oxford: Oxford University Press.
- HASPELMATH, MARTIN. 2000. Periphrasis. *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An international handbook on inflection and word-formation*, vol. 1, ed. by Geert Booij, Christian Lehmann, and Joachim Mugdan, in collaboration with Wolfgang Kesselheim and Stavros Skopeteas, 654–64. Berlin: Mouton de Gruyter.
- HAUGEN, JASON D., and DANIEL SIDDIQI. 2013. Roots and the derivation. *Linguistic Inquiry* 44.493–517.
- HAYWARD, RICHARD J. 1998. Qafar (East Cushitic). *The handbook of morphology*, ed. by Andrew Spencer and Arnold M. Zwicky, 624–47. Oxford: Blackwell.
- HAYWARD, RICHARD J., and MARTIN ORWIN. 1991. The prefix conjugation in Qafar-Saho: The survival and revival of a paradigm, part 1. *African Languages and Cultures* 4.157–76.
- HEWITT, B. GEORGE. 1995. *Georgian: A structural reference grammar*. Amsterdam: John Benjamins.
- HINZELIN, MARC-OLIVIER. 2011. Syncretism and suppletion in Gallo-Romance verb paradigms. In Maiden et al., 287–310.
- HYMAN, LARRY M. 2009. How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* (Special issue: *Data and theory: Papers in phonology in celebration of Charles W. Kisseberth*, ed. by Michael J. Kenstowicz) 3.213–38.
- HYMAN, LARRY M. 2011. In defense of prosodic typology: A response to Beckman and Venditti. *UC Berkeley Phonology Lab Annual Report (2011)*, 200–235.
- IVANOVA, V. F. 1969. Čelovek s tysjač'ju lic ili čelovek s tysjač'ju licami? *Russkij jazyk v škole* 2.66–73.
- JAKOBSON, ROMAN O. 1958. Morfoložičeskie nabljudenija nad slavjanskim sklonenijem (sostav russkix padežnyx form). *American contributions to the fourth International Congress of Slavists, Moscow, September 1958*, 127–56. The Hague: Mouton. [Reprinted in *Selected writings, vol. 2: Word and language*, 154–83. The Hague: Mouton, 1971. Translated as: Morphological observations on Slavic declension (the structure of Russian case forms). *Roman Jakobson: Russian and Slavic grammar: Studies 1931–1981*, ed. by Linda R. Waugh and Morris Halle, 105–33. Berlin: Mouton de Gruyter, 1984.]
- JAMIESON, CAROL. 1982. Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec. *International Journal of American Linguistics* 48.139–67.
- JOHNSON, C. DOUGLAS. 1972. *Formal aspects of phonological description*. The Hague: Mouton.
- KAYE, STEVEN. 2013. Morphomic stems in the Northern Talyshi verb: Diachrony and synchrony. In Cruschina et al., 181–208.
- KIBRIK, ALEKSANDR E. 1977. *Opyt strukturnogo opisanija arčinskogo jazyka, vol. 3: Dinamičeskaja grammatika*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki 13.) Moscow: Izdatel'stvo Moskovskogo universiteta.
- KIBRIK, ALEKSANDR E. 1994. Archi. *Indigenous languages of the Caucasus, vol. 4: North East Caucasian languages. Part 2: Presenting the three Nakh languages and six minor Lezgian languages*, ed. by Rieks Smeets, 297–365. Delmar, NY: Caravan Books.
- KIBRIK, ALEKSANDR E. 2003a. Nominal inflection galore: Daghestanian, with side glances at Europe and the world. *Noun phrase structure in the languages of Europe*, ed. by Frans Plank, 37–112. Berlin: Mouton de Gruyter.
- KIBRIK, ALEKSANDR E. 2003b. *Konstanty i peremennye jazyka*. Saint Petersburg: Aletheia.
- KIBRIK, ALEKSANDR E., and SANDRO V. KODZASOV. 1990. *Sopostavitel'noe izučenie dages-tanskix jazykov: Inja: Fonetika*. Moscow: MGU.
- KIBRIK, ANDREJ A. 2013. Peculiarities and origins of the Russian referential system. *Languages across boundaries: Studies in memory of Anna Siewierska*, ed. by Dik Bakker and Martin Haspelmath, 227–62. Berlin: De Gruyter Mouton.
- KIM, YUNI. 2010. Phonological and morphological conditions on affix order in Huave. *Morphology* 20.133–63.
- KONSTANTOPOULOS, STASINOS. 1998. A morphological analyser for Scottish Gaelic. Edinburgh: Department of Artificial Intelligence, University of Edinburgh master's thesis.

- KRASOVITSKY, ALEXANDER; ALISON LONG; MATTHEW BAERMAN; DUNSTAN BROWN; and GREVILLE G. CORBETT. 2008. Predicate nouns in Russian. *Russian Linguistics* 32.99–113.
- L'HERMITTE, RENÉ. 1978. *La phrase nominale en russe*. Paris: Institut d'études slaves.
- LAMB, WILLIAM. 2001. *Scottish Gaelic*. Munich: LINCOM Europa.
- LAMB, WILLIAM. 2008. *Scottish Gaelic speech and writing: Register variation in an endangered language*. (Belfast studies in language, culture and politics 16.) Belfast: Cló Ollscoil na Banríona.
- LETUČIJ, ALEKSANDR B. 2013. Svojstva nulevoj svjazki v ruskom jazyke v sopostavlenii so svojstvami vyražennogo glagola. Moscow: Vinogradov Russian Language Institute, MS.
- LUÍS, ANA R. 2011. Morphomic structure and loan-verb integration: Evidence from Lusophone Creoles. In Maiden et al., 235–54.
- MAČ'AVARIANI [MACHAVARIANI], GIVI I. 1974. Asp'ekt'is k'at'egoria kartvelur enebši [The category of aspect in the Kartvelian languages]. *Kartvelur Enata St'rukt'uris Sak'itxebi* 4.118–41. [Cited from Harris 1985.]
- MACKINNON, DONALD. 1910. Traces of the neuter gender in modern Scottish Gaelic. *Celtic Review* 6.24.296–304.
- MAIDEN, MARTIN. 2004. When lexemes become allomorphs—On the genesis of suppletion. *Folia Linguistica* 38.227–56.
- MAIDEN, MARTIN. 2005. Morphological autonomy and diachrony. *Yearbook of Morphology* 2004.137–75.
- MAIDEN, MARTIN. 2009. From pure phonology to pure morphology: The reshaping of the Romance verb. *Recherches linguistiques de Vincennes* 38.45–82. Online: <http://rlv.revues.org/1765>.
- MAIDEN, MARTIN. 2011. Morphemes and 'stress-conditioned allomorphy' in Romansh. In Maiden et al., 36–50.
- MAIDEN, MARTIN. 2013. The Latin 'third stem' and its Romance descendants. *Diachronica* 30.492–530.
- MAIDEN, MARTIN, and PAUL O'NEILL. 2010. On morphomic defectiveness: Evidence from the Romance languages of the Iberian Peninsula. In Baerman et al. 2010, 103–24.
- MAIDEN, MARTIN; JOHN CHARLES SMITH; MARIA GOLDBACH; and MARC-OLIVIER HINZELIN (eds.) 2011. *Morphological autonomy: Perspectives from Romance inflectional morphology*. Oxford: Oxford University Press.
- MARK, COLIN. 2003. *The Gaelic-English dictionary*. London: Routledge.
- MARSLÉN-WILSON, WILLIAM D. 2007. Morphological processes in language comprehension. *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell, 175–93. Oxford: Oxford University Press.
- MATTHEWS, P. H. 1972. *Inflectional morphology: A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.
- MCGINNIS, MARTHA. 2005. On markedness asymmetries in person and number. *Language* 81.699–718.
- MC SHANE, MARJORIE. 2009. Subject ellipsis in Polish and Russian. *Studia Linguistica* 63.98–132.
- MEL'ČUK, IGOR. 1993. *Cours de morphologie générale (théorique et descriptive), vol. 1: Introduction et première partie: Le mot*. Montréal: Les Presses de l'Université de Montréal.
- MEL'ČUK, IGOR. 1996. *Cours de morphologie générale (théorique et descriptive), vol. 3: Troisième partie: Moyens morphologiques; Quatrième partie: Syntactiques morphologiques*. Montréal: Les Presses de l'Université de Montréal.
- MEYER, PETER. 1994. Grammatical categories and the methodology of linguistics: Review article on van Helden 1993. *Russian Linguistics* 18.341–77.
- MÜLLER, GEREON. 2011. Syncretism without underspecification: The role of leading forms. *Word Structure* 4.1.53–103.
- NEVINS, ANDREW. 2011. Phonologically conditioned allomorph selection. *The Blackwell companion to phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, and Keren Rice, 2357–82. Oxford: Wiley-Blackwell.
- NIKOLAEVA, IRINA. 2013. Unpacking finiteness. In Brown et al. 2013, 99–122.

- NIKOLAEVA, IRINA, and ANDREW SPENCER. 2008. Nouns as adjectives and adjectives as nouns. London: University of London, SOAS, and Colchester: University of Essex, ms. Online: <http://privatewww.essex.ac.uk/~spena/PossMod/8adjnoun-language-revised-final-non-anon.pdf>.
- NIKOLIĆ, MIROSLAV B. 2013. Imenice koje u srpskom književnom jeziku proširuju osnovu morfemom *-ov-* u množini. *Južnoslovenski filolog* 69.277–318.
- O'NEILL, PAUL. 2011. The notion of the morpheme. In Maiden et al., 70–94.
- O'NEILL, PAUL. 2013. The morpheme and morphosyntactic/semantic features. In Cruschina et al., 221–46.
- OPITZ, ANDREAS; STEFANIE REGEL; GEREON MÜLLER; and ANGELA D. FRIEDERICI. 2013. Neurophysiological evidence for morphological underspecification in German strong adjective inflection. *Language* 89.231–64.
- PACIARONI, TANIA. 2012. Noun inflectional classes in Maceratese. *Inflection and word formation in Romance languages*, ed. by Sascha Gaglia and Marc-Olivier Hinzelin, 231–70. Amsterdam: John Benjamins.
- PALANCAR, ENRIQUE. 2012. The conjugation classes of Tilapa Otomi: An approach from canonical typology. *Linguistics* 50.783–832.
- PASTER, MARY. 2009. Explaining phonological conditions on affixation: Evidence from suppletive allomorphy and affix ordering. *Word Structure* 2.1.18–47.
- POLINSKY, MARIA. 2003. Non-canonical agreement is canonical. In Brown et al. 2003, 279–312.
- PRIESTLY, T. M. S. 1993. Slovene. In Comrie & Corbett, 388–451.
- PRINCE, ALAN, and PAUL SMOLENSKY. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report 2. New Brunswick, NJ: Rutgers University Center for Cognitive Science, and Boulder: University of Colorado at Boulder. [ROA 537, 2002. Revised publication, Oxford: Blackwell, 2004.]
- REH, MECHTHILD. 1985. *Die Krongo-Sprache (Niinò Mó-Dì): Beschreibung, Texte, Wörterverzeichnis*. Berlin: Dietrich Reimer.
- ROSS, MALCOLM. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples* (Pacific linguistics 572), ed. by Andrew Pawley, Robert Attenborough, Robin Hide, and Jack Golson, 15–66. Canberra: Australian National University.
- ROUND, ERICH R. 2009. *Kayardild morphology, phonology and morphosyntax*. New Haven, CT: Yale University dissertation.
- ROUND, ERICH R. 2013. *Kayardild morphology and syntax*. Oxford: Oxford University Press.
- ROUND, ERICH R. 2015. Rhizomorphemes, meromorphemes and metamorphemes. In Baerman et al. 2015, to appear.
- RUCART, PIERRE. 2006. *Morphologie gabaritique et interface phonosyntaxique: Aspects de la morphologie verbale en afar*. Paris: Université Paris Diderot (Paris 7) dissertation.
- SANU (SRPSKA AKADEMIJA NAUKA I UMETNOSTI). 1959. *Rečnik srpskohrvatskog književnog i narodnog jezika*. Belgrade: SANU.
- SCHMIDT, KARL H. 1963. Zu den Aspekten im Georgischen und in indogermanischen Sprachen. *Bedi Kartlisa* 15–16.107–15.
- SCHMIDT, KARL H. 1966. Tempora im Georgischen und in indogermanischen Sprachen. *Studia Caucasica* 2.48–57.
- SEIFART, FRANK. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Nijmegen: Radboud University, Nijmegen dissertation.
- SIEMUND, PETER, and FLORIAN DOLBERG. 2011. From lexical to referential gender: An analysis of gender change in medieval English based on two historical documents. *Folia Linguistica* 45.489–534.
- SMITH, JOHN CHARLES. 2013. The morpheme as a gradient phenomenon: Evidence from Romance. In Cruschina et al., 247–61.
- SPENCER, ANDREW. 2003. Periphrastic paradigms in Bulgarian. *Syntactic structures and morphological information* (Interface explorations 7), ed. by Uwe Junghanns and Luka Szucsich, 249–82. Berlin: Mouton de Gruyter.
- SPENCER, ANDREW. 2007. Extending deponency: Implications for morphological mismatches. In Baerman et al. 2007, 45–70.

- SPENCER, ANDREW. 2012. Identifying stems. In Bonami 2012b, 88–108.
- SPENCER, ANDREW. 2013. *Lexical relatedness*. Oxford: Oxford University Press.
- STANISLAV, JÁN. 1977. *Slowakische Grammatik*. Bratislava: Slovenské pedagogické nakladateľstvo.
- STUMP, GREGORY T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.
- STUMP, GREGORY T. 2006. Heterocclisis and paradigm linkage. *Language* 82.279–322.
- STUMP, GREGORY T. 2007. A non-canonical pattern of deponency and its implications. In Baerman et al. 2007, 71–95.
- STUMP, GREGORY T. 2009. Cells and paradigms in inflectional semantics. *Theory and evidence in semantics*, ed. by Erhard W. Hinrichs and John Nerbonne, 215–33. Stanford, CA: CSLI Publications.
- STUMP, GREGORY T. 2012. The formal and functional architecture of inflectional morphology. *Online proceedings of the Mediterranean Morphology Meeting* 8.255–71. Online: [http://www3.lingue.unibo.it/mmm2/?page\\_id=456](http://www3.lingue.unibo.it/mmm2/?page_id=456).
- STUMP, GREGORY T. 2013. Periphrasis in the Sanskrit verb system. In Chumakina & Corbett 2013, 105–38.
- STUMP, GREGORY T., and RAPHAEL A. FINKEL. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.
- SUTHAR, BABUBHAI KOHYABHAI. 2006. *Agreement in Gujarati*. Philadelphia: University of Pennsylvania dissertation.
- SWAN, OSCAR. 2012. Why *буду*? Почему *буду*? *Russian Linguistics* 36.305–18.
- TAYLOR, CATHERINE. 2012. *Formal relationships in the paradigm: A functional approach with a focus on the Romance verb*. Colchester: University of Essex dissertation.
- THORNTON, ANNA M. 2010–2011. La non canonicità del tipo it. *Braccio/braccia/bracci: Sovrabbondanza, difettività o iperdifferenziazione? Studi di grammatica italiana* 29–30.419–77.
- THORNTON, ANNA M. 2011a. Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. In Maiden et al., 358–81.
- THORNTON, ANNA M. 2011b. Overabundance: A non-canonical phenomenon in morphology. L'Aquila: Università degli Studi dell'Aquila, ms.
- TOSCO, MAURO. 2001. *The Dhaasanac language: Grammar, texts, vocabulary of a Cushitic language of Ethiopia*. (Cushitic language studies 17.) Cologne: Rüdiger Köppe.
- TROMMER, JOCHEN (ed.) 2012. *The morphology and phonology of exponence*. Oxford: Oxford University Press.
- TROMPELT, HELENA; DENISA BORDAG; and THOMAS PECHMANN. 2013. (Ir)regularity of verbs revisited: Evidence for lexical entry complexity. *The Mental Lexicon* 8.26–52.
- VAN DE VELDE, MARK. 2013. The Bantu connective construction. *The genitive (Case and grammatical relations across languages 5)*, ed. by Anne Carlier and Jean-Christophe Verstraete, 217–52. Amsterdam: John Benjamins.
- VAN HELDEN, W. ANDRIES. 1993. *Case and gender: Concept formation between morphology and syntax*. (Studies in Slavic and general linguistics 20.) 2 vols. Amsterdam: Rodopi.
- VERKUYL, HENK; CO VET; ANDRÉE BORILLO; MYRIAM BRAS; ANNE LE DRAOULEC; ARIE MOLENDIJK; HENRIËTTE DE SWART; CARL VETTERS; and LAURE VIEU. 2004. Tense and aspect in sentences. *Handbook of French semantics*, ed. by Francis Corblin and Henriëtte de Swart, 233–70. Stanford, CA: CSLI Publications.
- VINCENT, NIGEL. 2011. Non-finite forms, periphrases, and autonomous morphology in Latin and Romance. In Maiden et al., 417–35.
- WALTHER, GÉRALDINE. 2013. *Sur la canonicité en morphologie: Perspective empirique, formelle et computationnelle*. Paris: Université Paris Diderot (Paris 7) dissertation.
- WITZLACK-MAKAREVICH, ALENA. 2011. *Typological variation in grammatical relations*. Leipzig: University of Leipzig dissertation.
- WORTH, DEAN S. 1984. Russian *gen*<sup>2</sup>, *loc*<sup>2</sup> revisited. *Signs of friendship: To honour A. G. F. van Holk, Slavist, linguist, semiotician*, ed. by Joost J. van Baak, 295–306. Amsterdam: Rodopi.
- XALILOV, MADŽID Š. 1985. Ob ograničenom množestvennom čisle suščestvitel'nyx bežtinskogo jazyka. *Kategorija čisla v dagestanskix jazykax: Sbornik statej*, ed. by K. Š. Mikailov, 136–43. Makhachkala: Dagestanskij filial AN SSSR.



- XU, ZHENG, and MARK ARONOFF. 2011. A realizational optimality-theoretic approach to full and partial identity of forms. In Maiden et al., 257–86.
- YU, ALAN C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.
- ZALIZNJAK, ANDREJ A. 1973. O ponimanii termina ‘padež’ v lingvističeskix opisanijax. *Problemy grammatičeskogo modelirovanija*, ed. by Andrej A. Zaliznjak, 53–87. Moscow: Nauka. [Reprinted as *Russkoe imennoe slovoizmenenie: S priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaniju*, ed. by Andrej A. Zaliznjak, 613–47. Moscow: Jazyki slavjanskoj kul'tury, 2002.]
- ŽIRKOV, L. I. 1955. *Lakskij jazyk: Fonetika i morfologija*. Moscow: AN SSSR.

Surrey Morphology Group  
 School of English and Languages  
 Faculty of Arts and Human Sciences  
 University of Surrey  
 Guildford, Surrey, GU2 7XH, United Kingdom  
 [g.corbett@surrey.ac.uk]

[Received 6 May 2012;  
 revision invited 17 April 2013;  
 revision received 27 September 2013;  
 accepted pending revisions 20 December 2013;  
 revision received 21 January 2014;  
 accepted 28 May 2014]